

G OPEN ACCESS

Citation: Szalkai B, Varga B, Grolmusz V (2015) Graph Theoretical Analysis Reveals: Women's Brains Are Better Connected than Men's. PLoS ONE 10(7): e0130045. doi:10.1371/journal.pone.0130045

Academic Editor: Yongtang Shi, Nankai University, CHINA

Received: April 3, 2015

Accepted: May 15, 2015

Published: July 1, 2015

Copyright: © 2015 Szalkai et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License,](http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The unprocessed and pre-processed MRI data is available at the Human Connectome Project's website: [http://www.](http://www.humanconnectome.org/documentation/S500) [humanconnectome.org/documentation/S500.](http://www.humanconnectome.org/documentation/S500) The assembled graphs that were analyzed in the present work can be accessed and downloaded at the site [http://braingraph.org/download-pit-group](http://braingraph.org/download-pit-group-connectomes/)[connectomes/.](http://braingraph.org/download-pit-group-connectomes/) The individual graph results are detailed in a large Excel table at the site [http://uratim.](http://uratim.com/bigtable1.zip) [com/bigtable1.zip.](http://uratim.com/bigtable1.zip)

Funding: The authors have no support or funding to report.

RESEARCH ARTICLE

Graph Theoretical Analysis Reveals: Women's Brains Are Better Connected than Men's

Balázs Szalkai¹, Bálint Varga¹, Vince Grolmusz^{1,2}*

1 PIT Bioinformatics Group, Eötvös University, H-1117 Budapest, Hungary, 2 Uratim Ltd., H-1118 Budapest, Hungary

* grolmusz@pitgroup.org

Abstract

Deep graph-theoretic ideas in the context with the graph of the World Wide Web led to the definition of Google's PageRank and the subsequent rise of the most popular search engine to date. Brain graphs, or connectomes, are being widely explored today. We believe that non-trivial graph theoretic concepts, similarly as it happened in the case of the World Wide Web, will lead to discoveries enlightening the structural and also the functional details of the animal and human brains. When scientists examine large networks of tens or hundreds of millions of vertices, only fast algorithms can be applied because of the size constraints. In the case of diffusion MRI-based structural human brain imaging, the effective vertex number of the connectomes, or brain graphs derived from the data is on the scale of several hundred today. That size facilitates applying strict mathematical graph algorithms even for some hard-to-compute (or NP-hard) quantities like vertex cover or balanced minimum cut. In the present work we have examined brain graphs, computed from the data of the Human Connectome Project, recorded from male and female subjects between ages 22 and 35. Significant differences were found between the male and female structural brain graphs: we show that the average female connectome has more edges, is a better expander graph, has larger minimal bisection width, and has more spanning trees than the average male connectome. Since the average female brain weighs less than the brain of males, these properties show that the female brain has better graph theoretical properties, in a sense, than the brain of males. It is known that the female brain has a smaller gray matter/white matter ratio than males, that is, a larger white matter/gray matter ratio than the brain of males; this observation is in line with our findings concerning the number of edges, since the white matter consists of myelinated axons, which, in turn, roughly correspond to the connections in the brain graph. We have also found that the minimum bisection width, normalized with the edge number, is also significantly larger in the right and the left hemispheres in females: therefore, the differing bisection widths are independent from the difference in the number of edges.

Competing Interests: The authors have the following interests: Vince Grolmusz is a PLOS ONE Editorial Board member. Vince Grolmusz is a CEO and shareholder of Uratim Ltd. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

Introduction

In the last several years hundreds of publications appeared describing or analyzing structural or functional networks of the brain, frequently referred to as "connectome" $[1-4]$ $[1-4]$ $[1-4]$. Some of these publications analyzed data from healthy humans $[5-8]$ $[5-8]$ $[5-8]$ $[5-8]$, and some compared the connectome of the healthy brain with diseased one [[9](#page-28-0)–[13](#page-28-0)].

So far, the analyses of the connectomes mostly used tools developed for very large networks, such as the graph of the World Wide Web (with billions of vertices), or protein-protein interaction networks (with tens or hundreds of thousands of vertices), and because of the huge size of original networks, these methods used only very fast algorithms and frequently just primary degree statistics and graph-edge counting between pre-defined regions or lobes of the brain [\[14](#page-28-0)].

In the present work we demonstrate that deep and more intricate graph theoretic parameters could also be computed by using, among other tools, contemporary integer programming approaches for connectomes with several hundred vertices.

With these mathematical tools we show statistically significant differences in some graph properties of the connectomes, computed from MRI imaging data of male and female brains. We will not try to associate behavioral patterns of males and females with the discovered structural differences $[14]$ $[14]$ $[14]$ (see also the debate that article has generated: $[15-17]$ $[15-17]$ $[15-17]$ $[15-17]$ $[15-17]$), because we do not have behavioral data of the subjects of the imaging study, and, additionally, we cannot describe high-level functional properties implied by those structural differences. However, we clearly demonstrate that deep graph-theoretic parameters show "better" connections in a certain sense in female connectomes than in male ones.

The study of [\[14\]](#page-28-0) analyzed the 95-vertex graphs of 949 subjects aged between 8 and 22 years, using basic statistics for the numbers of edges running either between or within different lobes of the brain (the parameters deduced were called hemispheric connectivity ratio, modularity, transitivity and participation coefficients, see $[14]$ for the definitions). It was found that males have significantly more intra-hemispheric edges than females, while females have significantly more inter-hemispheric edges than males.

Results and Discussion

We have analyzed the connectomes of 96 subjects, 52 females and 44 males, each with 83, 129, 234, 463 and 1015 node resolutions, and each graphs with five different weight functions. We considered the connectomes as graphs with weighted edges, and performed graph-theoretic analyses with computing some polynomial-time computable and also some NP-hard graph parameters on the individual graphs, and then compared the results statistically for the male and the female group.

We have found that female connectomes have more edges, larger (edge-normalized) minimum bisection widths, larger minimum-vertex covers and more spanning trees and are better expanders than the male connectomes.

In order to describe the parameters, which differ significantly among male and female connectomes, we need to place them in the context of their graph theoretical definitions.

Edge number and edge weights

We have found significantly higher number of edges (counted with 5 types of weights and also without any weights) in both hemispheres and also in the whole brain in females, in all resolutions. This finding is surprising, since we used the same parcellation and the same tractography and the same graph-construction methods for female and male brains, and because it is proven that females have, on average, less-weighting brains than males $[18]$. For example, in the

234-vertex resolution, the average number of (unweighted) edges in female connectomes is 1826, in males 174[2,](#page-6-0) with $p = 0.00063$ $p = 0.00063$ (see [Table 1](#page-3-0) with a summary and Tables 2, 3, [4](#page-11-0), [5](#page-14-0) and [6](#page-17-0) with the results). The work of $[14]$ reported similar findings in inter-hemispheric connections only.

It is known that there are statistical differences in the size and the weight of the female and the male cerebra $[18]$ $[18]$. It was also published $[19]$ that female brains statistically have a smaller gray matter/white matter ratio, that is, a higher white matter/gray matter ratio than male brains. We argue that this observation is in line with the quantitative differences in the fibers and edges in the connectomes of the sexes: In a simplified view, the edges of the braingraph correspond to the fibers of the myelinated axons in the white matter, while the nodes of the graph to areas of the gray matter. Therefore, since females have a higher white matter/gray matter ratio than males by $[19]$ that fact implies that the number of detected fibers by the tractography step of the processing is relatively higher in females than in males, and this higher number of fibers imply higher number of edges in female connectomes.

We are carefully dealing with the possibilities of artifacts in the edge number differences in the "Methods" section.

Minimum cut and balanced minimum cut

Suppose the nodes, or the vertices, of a graph are partitioned into two, disjoint, nonempty sets, say X and Y ; their union is the whole vertex-set of the graph. The X , Y cut is the set of all edges connecting vertices of X with the vertices of Y ($Fig 1$ panel A). The size of the cut is the number of edges in the cut. In graph theory, the size of the minimum cut is an interesting quantity. The minimum cut between vertices a and b is the minimum cut, taken for all X and Y, where vertex a is in X and b is in Y . This quantity gives the "bottleneck", in a sense, between those two nodes (c.f., Menger theorems and Ford-Fulkerson's Min-Cut-Max-Flow theorem [\[20](#page-28-0), [21\]](#page-28-0)). The minimum cut in a graph is defined to be the cut with the fewest edges for all non-empty sets X and Y , partitioning the vertices.

Clearly, for non-negative weights, the size of the minimum cut in a non-connected graph is 0. Very frequently, however, in connected graphs, the minimum cut is determined by just the smallest degree node: that node is the only element of set X and all the other vertices of the graph are in Y ([Fig 1](#page-20-0) panel B). Because of this phenomenon, the minimum cut is frequently queried for the "balanced" case, when the size (i.e., the number of vertices) of X and Y needs to be equal (or, more exactly, may differ by at most one if the number of the vertices of the graph is odd), see $Fig 1$ panel C. This problem is referred to as the balanced minimum cut or the minimum bisection problem. If the minimum bisection is small that means that there exist a partition of the vertices into two sets of equal size that are connected with only a few edges. If the minimum bisection is large then the two half-sets in every possible bisections of the graph are connected by many edges.

Therefore, the balanced minimum cut of a graph is independent of the particular labeling of the nodes. The number of all the balanced cuts in a graph with n vertices is greater than

$$
\frac{1}{n+1}2^n,
$$

that is, for $n = 463$, this number is much larger than the number of atoms in the visible universe [\[22](#page-28-0)]. Consequently, one cannot practically compute the minimum bisecton width by reviewing all the bisectons in a graph of that size. Moreover, the complexity of computing this quantity is known to be NP-hard [\[23\]](#page-29-0) in general, but with contemporary integral programming

[Table 1. T](#page-2-0)he results and the statistical analysis of the graph-theoretical evaluation of the sex differences in the 96 diffusion MRI images. The first column gives the resolutions in each hemisphere; the numbers of nodes in the whole graph are 83, 129, 234, 463 and 1015. The second column describes the graph parameter computed: its syntactics is as follows: each parameter-name contains two separating "_" symbols that define three parts of the parameter-name. The first part describe the hemisphere or the whole connectome with the words Left, Right or All. The second part describes the parameter computed, and the third part the weight function used (their definitions are given in section "Materials and methods"). The third column contains the p-values of the first round, the second column the p-values of the second round, and the third column the (very strict) Holm-Bonferroni correction of the p-value. With p = 0.05 all the first 12 rows describe significantly different graph theoretical properties between sexes. One-by-one, each row with italic third column describe significant differences between sexes, with $p = 0.05$. For the details we refer to the section "Statistical analysis".

(Continued)

PLOS

ONE

Table 1. (Continued)

Table 1. (Continued)

Table 1. (Continued)

.Ø

PLOS | ONE

doi:10.1371/journal.pone.0130045.t001

[Table 2. T](#page-2-0)he graph-theoretic parameters computed for the 83-vertex graphs. The table contains their arithmetic means in the male and female groups, and the corresponding p-values in round 1 (see the "Statistical analysis" subsection). The results of the graph-parameters are defined in the caption of [Table 1.](#page-3-0) Significant differences ($p < 0.01$) are denoted with an asterisk in the last column.

Table 2. (Continued)

 \mathcal{D} PLOS \vert one

Table 2. (Continued)

 \mathcal{D} PLOS \vert one

doi:10.1371/journal.pone.0130045.t002

PLOS | ONE .Ø

[Table 3. T](#page-2-0)he graph-theoretic parameters computed for the 129-vertex graphs. The table contains their arithmetic means in the male and female groups, and the corresponding p-values in round 1 (see the "Statistical analysis" subsection). The results of the graph-parameters are defined in the caption of [Table 1.](#page-3-0) Significant differences ($p < 0.01$) are denoted with an asterisk in the last column.

Table 3. (Continued)

 \mathcal{D} PLOS \vert one

Table 3. (Continued)

PLOS | ONE

doi:10.1371/journal.pone.0130045.t003

[Table 4. T](#page-2-0)he graph-theoretic parameters computed for the 234-vertex graphs. The table contains their arithmetic means in the male and female groups, and the corresponding p-values in round 1 (see the "Statistical analysis" subsection). The results of the graph-parameters are defined in the caption of [Table 1.](#page-3-0) Significant differences ($p < 0.01$) are denoted with an asterisk in the last column.

Table 4. (Continued)

 \mathcal{D} PLOS \vert one

Table 4. (Continued)

 \mathcal{D} PLOS \vert one

Table 4. (Continued)

 $\cdot \mathbb{Q}$

PLOS | ONE

doi:10.1371/journal.pone.0130045.t004

[Table 5. T](#page-2-0)he graph-theoretic parameters computed for the 463-vertex graphs. The table contains their arithmetic means in the male and female groups, and the corresponding p-values in round 1 (see the "Statistical analysis" subsection). The results of the graph-parameters are defined in the caption of [Table 1.](#page-3-0) Significant differences ($p < 0.01$) are denoted with an asterisk in the last column.

Table 5. (Continued)

 \mathcal{D} PLOS \vert one

Table 5. (Continued)

 \mathcal{D} PLOS \vert one

Table 5. (Continued)

doi:10.1371/journal.pone.0130045.t005

[Table 6. T](#page-2-0)he graph-theoretic parameters computed for the 1015-vertex graphs. The table contains their arithmetic means in the male and female groups, and the corresponding p-values in round 1 (see the "Statistical analysis" subsection). The results of the graph-parameters are defined in the caption of [Table 1.](#page-3-0) Significant differences ($p < 0.01$) are denoted with an asterisk in the last column.

Table 6. (Continued)

 \mathcal{D} PLOS \vert one

Table 6. (Continued)

 \mathcal{D} PLOS \vert one

doi:10.1371/journal.pone.0130045.t006

approaches, and for the graph-sizes we are dealing with, the exact values are computable in reasonable time.

In computer engineering, an important measure of the quality of an interconnection network is its minimum bisection width $[24]$: the higher the width is the better the network. Based on this observation, we can say that the data imply the better quality of female connectome, compared to that of males.

For the whole brain graph, as it is anticipated, we have found that the minimum balanced cut is almost exactly represents the edges crossing the corpus callosum, connecting the two cerebral hemispheres.

We show that within both hemispheres, the minimum bisection size of female connectomes are significantly larger than the minimum bisection size of the males. Much more importantly, we show that this remains true if we normalize with the sum of all edge-weights: that is, this phenomenon cannot be due to the higher number of edges or the greater edge weights in the female brain: it is an intrinsic property of the female brain graph in our data analyzed.

For example, in the 234-vertex resolution, in the left hemisphere, the normalized balanced minimum cut in females, on the average, is 0.09416 , in the males 0.07896 , $p = 0.00153$ (see [Table 1](#page-3-0) with a summary and Tables $2, 3, 4, 5$ $2, 3, 4, 5$ $2, 3, 4, 5$ $2, 3, 4, 5$ $2, 3, 4, 5$ $2, 3, 4, 5$ $2, 3, 4, 5$ and [6](#page-17-0) with the results).

We think that this finding is one of the main results of the present work: even if the significant difference in the weighted edge numbers were due to some artifacts in the data acquisition/processing workflow, the normalized balanced minimum cut size seems to be independent from those processes.

Eigengap and the expander property

Expander graphs and the expander-property of graphs are one of the most interesting area of graph theory: they are closely related to the convergence rate and the ergodicity of Markov

chains, and have applications in the design of communication- and sorting networks and methods for de-randomizing algorithms [\[25\]](#page-29-0). A graph is an ε -expander, if every—not too small and not too large—vertex-set S of the graph has at least ε |S| outgoing edges (see [[25](#page-29-0)] for the exact definition).

Random walks on good expander graphs converge very fast to the limit distribution: this means that good expander graphs, in a certain sense, are "intrinsically better" connected than bad expanders. It is known that large eigengap of the walk transition matrix of the graph implies good expansion property [[25\]](#page-29-0).

We have found that women's connectomes have significantly larger eigengap, and, consequently, they are better expander graphs than the connectomes of men. For example, in the 83-node resolution, in the left hemisphere and in the unweighted graph, the average female connectome's eigengap is 0.306 while in the case of men it is 0.272, with $p = 0.00458$.

The number of spanning forests

A tree in graph theory is a connected, cycle-free graph. Any tree on n vertices has the same number of edges: n−1. Trees, and tree-based structures are common in science: phylogenetic trees, hierarchical clusters, data-storage on hard-disks, or a computational model called decision trees all apply graph-theoretic trees. A spanning tree is a minimal subgraph of a connected graph that is still connected. Some graphs have no spanning trees at all: only connected graphs have spanning trees. A tree has only one spanning tree: itself. Any connected graph on *n* vertices has a minimum of $n-1$ and a maximum of $n(n-1)/2$ edges [\[26\]](#page-29-0). A connected graph with few edges still may have exponentially many different spanning trees: e.g., the n -vertex wheel on [Fig 1](#page-20-0) panel D has at least 2^{n-1} spanning trees (for $n \ge 4$). Cayley's famous theorem, and its celebrated proof with Prüfer codes [[27](#page-29-0)] shows that the number of spanning trees of the complete graph on *n* vertices is n^{n-2} .

If a graph is not connected, then it contains more than one connected components. Each connected component has at least one spanning tree, and the whole graph has at least one spanning forest, comprises the spanning trees of the components. The number of spanning forests is clearly the product of the numbers of the spanning trees of the components.

For graphs in general, one can compute the number of their spanning forests by Kirchoff's matrix tree theorem $[28-31]$ $[28-31]$ $[28-31]$ $[28-31]$ using the eigenvalues of the Laplacian matrix $[29]$ $[29]$ $[29]$ of the graph.

We show that female connectomes have significantly higher number of spanning trees than the connectomes of males. For example, in the 129-vertex resolution, in the left hemisphere, the logarithm of the number of the spanning forests in the unweighted case are 162.01 in females, 158.88 in males with $p = 0.013$.

The workflow of this work is summarized on [Fig 2](#page-22-0).

Figs $\frac{3}{2}$ $\frac{3}{2}$ $\frac{3}{2}$ and $\frac{4}{3}$ $\frac{4}{3}$ $\frac{4}{3}$ visualize the differences of some graph parameters between the connectomes of the sexes.

Methods

Data source and graph computation

The dataset applied is a subset of the Human Connectome Project [\[32\]](#page-29-0) anonymized 500 Subjects Release: (<http://www.humanconnectome.org/documentation/S500>) of healthy subjects between 22 and 35 years of age. Data was downloaded in October, 2014.

The Connectome Mapper Toolkit [[33](#page-29-0)] [\(http://cmtk.org](http://cmtk.org)) was applied for brain tissue segmentation into grey and white matter, partitioning, tractography and the construction of the graphs from the fibers identified in the tractography step. The Connectome Mapper Toolkit [\[33](#page-29-0)] default partitioning was used (computed by the FreeSurfer, and based on the Desikan-

doi:10.1371/journal.pone.0130045.g002

Killiany anatomical atlas) into 83, 129, 234, 463 and 1015 cortical and sub-cortical structures (as the brainstem and deep-grey nuclei), referred to as "Regions of Interest", ROIs, (see [Fig 4](#page-23-0) in [\[33](#page-29-0)]). Tractography was performed by the Connectome Mapper Toolkit [\[33\]](#page-29-0), choosing the deterministic streamline method with the MRtrix processing tool [[34\]](#page-29-0) with randomized seeding.

The graphs were constructed as follows: the nodes correspond to the ROIs in the specific resolution. Two nodes were connected by an edge if there exists at least one fiber (determined by the tractography step) connecting the ROIs, corresponding to the nodes. More than one fibers, connecting the same nodes, may or may not give rise to the weight of that edge, depending on the weighting method. Loops were deleted from the graph.

The weights of the edges are assigned by several methods, taking into account the lengths and the multiplicities of the fibers, connecting the nodes:

• Unweighted: Each edge has weight 1.

[Fig 3. E](#page-21-0)mpirical cumulative distribution function of the Right_MinCutBalDivSum_FAMean graph parameter (that is, edge-number-normed minimum bisection width in the right hemisphere, weighted by the arithmetic mean of the fractional anisotropies [[35](#page-29-0)] of the fibers, belonging to the edge) in the 129-node resolution. For every value x on the horizontal line, the curves demonstrate the male (blue, continuous line) and female (red, dashed line) fraction of subjects with Right_MinCutBalDivSum_FAMean value of at most x. For example, for x = 0.02, 40% of the females have the Right_MinCutBalDivSum_FAMean value less than x, while about 85% of males have that value less than x.

doi:10.1371/journal.pone.0130045.g003

[Fig 4. E](#page-21-0)mpirical cumulative distribution function of the All_PGEigengap_FiberNDivLength graph parameter (that is, the eigengap of the transitionmatrix of the whole brain graph with each edge weighted by the number of fibers belonging to the edge, divided by their average length), in the 129-node resolution. For every value x on the horizontal line, the curves demonstrate the male (blue, continuous line) and female (red, dashed line) fraction of subjects with All PGEigengap FiberNDivLength value of at most x. For example, for $x = 0.025$, about 17% of the females have the All_PGEigengap_FiberNDivLength value less than x, while about 58% of males have that value less than x.

doi:10.1371/journal.pone.0130045.g004

- FiberN: The number of fibers traced along the edge: this number is larger than one if more than one fibers connect two cortical or sub-cortical areas, corresponding to the two endpoints of the edge.
- FAMean: The arithmetic mean of the fractional anisotropies [\[35](#page-29-0)] of the fibers, belonging to the edge.
- FiberLengthMean: The average length of the fibers, connecting the two endpoints of the edge.
- FiberNDivLength: The number of fibers belonging to the edge, divided by their average length. This quantity is related to the simple electrical model of the nerve fibers: by modeling the fibers as electrical resistors with resistances proportional to the average fiber length, this quantity is precisely the conductance between the two regions of interest. Additionally, FiberNDivLength can be observed as a reliability measure of the edge: longer fibers are less reliable than the shorter ones, due to possible error accumulation in the tractography algorithm that constructs the fibers from the anisotropy data. Multiple fibers connecting the same two ROIs, corresponding to the endpoints, add to the reliability of the edge, because of the independently tractographed connections.

By generalized adjacency matrix we mean a matrix of size $n \times n$ where n is the number of nodes (or vertices) in the graph, whose rows and columns correspond to the nodes, and whose each element is either zero if there is no edge between the two nodes, or equals to the weight of the edge connecting the two nodes. By the *generalized degree* of a node we mean the sum of the weights of the edges adjacent to that node. Note that the generalized degree of the node ν is exactly the sum of the elements in the row (or column) of the generalized adjacency matrix corresponding to v. By generalized Laplacian matrix we mean the matrix D−A, where D is a diagonal matrix containing the generalized degrees, and A is the generalized adjacency matrix.

Graph parameters

We calculated various graph parameters for each brain graph and weight function. These parameters included:

- Number of edges (Sum). The weighted version of this quantity is the sum of the weights of the edges.
- Normalized largest eigenvalue ($Ad\overline{j}LMaxDiVD$): The largest eigenvalue of the generalized adjacency matrix, divided by the average degree. Dividing by the average degree of vertices was necessary because the largest eigenvalue is bounded by the average- and maximum degrees, and thus is considered by some a kind of "average degree" itself [\[26\]](#page-29-0). This means that a denser graph may have a bigger λ_{max} largest eigenvalue solely because of a larger average degree. We note that the average degree is already defined by the sum of weights.
- Eigengap of the transition matrix (PGE i gengap): The transition matrix P_G is obtained by dividing all the rows of the generalized adjacency matrix by the generalized degree of the corresponding node. When performing a random walk on the graph, for nodes i and j , the corresponding matrix element describes the probability of transitioning to node j, supposing that we are at node *i*. The eigengap of a matrix is the difference of the largest and the second largest eigenvalue. It is characteristic to the expander properties of the graph: the larger the gap, the better expander is the graph (see $[25]$ $[25]$ for the exact statements and proofs).

• Hoffman's bound (HoffmanBound): The expression

$$
1+\frac{\lambda_{\max}}{|\lambda_{\min}|},
$$

where λ_{max} and λ_{min} denote the largest and smallest eigenvalues of the adjacency matrix. It is a lower bound for the chromatic number of the graph. The chromatic number is generally higher for denser graphs, as the addition of an edge may make a previously valid coloring invalid.

- Logarithm of number of spanning forests (LogAbsSpanningForestN): The number of the spanning trees in a connected graph can be calculated from the spectrum of its Laplacian [\[28,](#page-29-0) [29\]](#page-29-0). Denser graphs tend to have more spanning trees, as the addition of an edge introduces zero or more new spanning trees. If a graph is not connected, then the number of spanning forests is the product of the numbers of the spanning trees of the components. The parameter LogAbsSpanningForestN equals to the logarithm of the number of spanning forests in the unweighted case. In the case of other weight functions, if we define the weight of a tree by the product of the weights of its edges, then this parameter equals to the sum of the logarithms of the weights of the spanning trees in the forests.
- Balanced minimum cut, divided by the number of edges (MinCutBalDivSum): The task is to partition the graph into two sets whose size may differ from each other by at most 1, so that the number of edges crossing the cut is minimal. This is the "balanced minimum cut" problem, or sometimes called the "minimum bisection width" problem. For the whole brain graph, our expectation was that the minimum cut corresponds to the boundary of the two hemispheres, which was indeed proven when we analyzed the results.
- Minimum cost spanning tree (MinSpanningForest), calculated with Kruskal's algorithm.
- Minimum weighted vertex cover (MinVertexCover): Each vertex should have a (possibly fractional) weight assigned such that, for each edge, the sum of the weights of its two end-points is at least 1. This is the fractional relaxation of the NP-hard vertex-cover problem [[36](#page-29-0)]. The minimum of the sum of all vertex-weights is computable by a linear programming approach.
- Minimum vertex cover (MinVertexCoverBinary): Same as above, but each weight must be 0 or 1. In other words, a minimum size set of vertices is selected such that each edge is covered by at least one of the selected vertices. This NP-hard graph-parameter is computed only for the unweighted case. The exact values are computed by an integer programming solver SCIP [\(http://scip.zib.de\)](http://scip.zib.de), [\[37,](#page-29-0) [38\]](#page-29-0).

The 9 parameters above were computed for all five resolutions and for the left and the right hemispheres and also for the whole connectome, with all 5 weight functions (with the following exceptions: MinVertexCoverBinary was computed only for the unweighted case, and the MinSpanningTree was not computed for the unweighted case).

The results are detailed in an Excel table with 480 rows (5 rows of different resolutions for each brain) and 120 columns (7 parameters are computed for all 5 weight-functions for the left- and right hemispheres and the whole brain, one parameter for just one weight function, and one parameter for 4 weight functions only, that is $7 \cdot 5 \cdot 3 + 1 \cdot 3 + 4 \cdot 3 = 120$ at the site <http://uratim.com/bigtable1.zip>

Investigation for possible artifacts

We applied the very same graph construction method for all dMRI data sets, independently of the sex of the subjects. Surprisingly, we have found significant differences in numerous graph parameters between male and female brains, e.g., in the number of the edges in the connectome. We will review here a possible bias in the connectome construction, and we conclude that, by the best of our knowledge, it cannot cause the differences in the graph parameters.

One possible source of error could be the statistically different brain sizes of the sexes [\[18\]](#page-28-0). In the tractography step, when streamlines are progressed by the deterministic method from voxel to voxel, longer fibers may stop prematurely [\[39,](#page-29-0) [40\]](#page-29-0). Therefore, longer fibers may be harder to reconstruct. Since male brains are larger than the brain of the females, they contain longer fiber bundles that could be more difficult to reconstruct.

We have applied five different edge weighting methods. One of these is called Fiber-LengthMean that describes the average lengths of fibers that define the edge in question.

Clearly, the FiberLengthMean weight rewards the longer fibers and penalizes the shorter ones. Consequently, the advantage of the total sum of these weights of the edges in the case of women needs to be smaller or non-existing if the "premature stop" tractography bias were the cause of the edge number difference. The data below show that just the opposite holds true.

More exactly, let us consider [Table 3](#page-9-0), containing data with resolution 129:

- All Sum FiberLengthMean female: 30670.09535 male: 28478.19852 p = 0.03582 f/m ratio: 1.07
- All Sum Unweighted female: 1020.80769 male: 972.86667 p = 0.00026 f/m ratio: 1.049,

Here the unweighted ratio is smaller, meaning that weighting with the fiber lengths increases the advantage of the females!

Similarly, in [Table 4](#page-11-0), with resolution 234:

All Sum FiberLengthMean female: 51558.63408 male: 48397.55225 p = 0.05764, f/m ratio: 1,065

All Sum Unweighted female: 1826.03846 male: 1742.66667 p = 0.00063 f/m ratio: 1,048

Here, again, the unweighted ratio is smaller, meaning that weighting with the fiber lengths increases the advantage of the females. We believe that these figures make our results stronger, proving that females have longer and more connections in their connectome than males.

Statistical analysis

Since each connectome was computed in multiple resolutions (in 83, 129, 234, 463 and 1015 nodes), we had five graphs for each brain. In addition, the parameters were calculated separately for the connectome within the left and right hemispheres as well, not only the whole graph, since we intended to examine whether statistically significant differences can be attributed to the left or right hemispheres. Each subjects' brain was corresponded to 15 graphs (5 resolutions, each in the left and the right hemispheres, plus the whole cortex with subcortical areas) and for each graph we calculated 9 parameters, each (with the exceptions noted above) with 5 different edge weights. This means that we assigned $7 \cdot 5 \cdot 3 + 1 \cdot 3 + 4 \cdot 3 = 120$ attributes to each resolution of the 96 brains, that is, 600 attributes to each brain.

The statistical null hypothesis $[41]$ was that the graph parameters do not differ between the male and the female groups. As the first approach, we have used ANOVA (Analysis of

variance) $[42]$ $[42]$ to assign p-values for all parameters in each hemispheres and in each resolutions and in each weight-assignments.

Our very large number of attributes may lead to false negatives, i.e., to "type II" statistical errors: in other words, it may happen that an attribute, with a very small p-value may appear "at random", simply because we tested a lot of attributes. In order to deal with "type II" statistical errors, we followed the route described below.

We divided the population randomly into two sets by the parity of the sum of the digits in their ID. The first set was used for making hypotheses and the second set for testing these hypotheses. This was necessary to avoid type II errors resulting from multiple testing correction. If we made hypotheses for all the numerical parameters, then the Holm-Bonferroni correction [\[43\]](#page-29-0) we used would have unnecessarily increased the p-values. Thus we needed to filter the hypotheses first, and that is why we needed the first set. Testing on the first set allowed us to reduce the number of hypotheses and test only a few of them on the second set.

The hypotheses were filtered by performing ANOVA (Analysis of variance) [[42](#page-29-0)] on the first set. Only those hypotheses were selected to qualify for the second round where the p-value was less than 1%. The selected hypotheses were then tested for the second set as well, and the resulting p-value corrected with the Holm-Bonferroni correction method [\[43\]](#page-29-0) with a significance level of 5%.

In [Table 1](#page-3-0) those hypotheses rejected were highlighted in bold, meaning that all the corresponding graph parameters differ significantly in sex groups at a combined significance level of 5%.

We also highlighted (in italic) those p-values which were individually less than the threshold, meaning that these hypotheses can individually be rejected at a level of 5%, but it is very likely that not all of these graph parameters are significantly different between the sexes.

Conclusions

We have computed 83-, 129-, 234-, 463- and 1015 vertex-graphs from the diffusion MRI images of the 96 subjects of 52 females and 44 males, between the age of 22 and 35. After a careful statistical analysis, we have found significant differences between certain graph parameters of the male and female brain graphs. Our findings show that the female brain graphs have generally more edges (counted with and without weights), have larger normalized minimum bisection widths in its hemispheres, are better expander graphs and have more spanning trees (counted with and without weights) than the connectomes of males [\(Table 1](#page-3-0)). We believe that in the future, due to the relatively small size of the underlying networks, graph theoretical methods could have a wide application spectrum in the analysis of the connectome.

Acknowledgments

Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Author Contributions

Conceived and designed the experiments: VG BV BS. Performed the experiments: BV BS. Analyzed the data: BV BS VG. Contributed reagents/materials/analysis tools: BS BV VG. Wrote the paper: BS VG.

References

- [1.](#page-1-0) Hagmann P, Grant PE, Fair DA (2012) MR connectomics: a conceptual framework for studying the developing brain. Front Syst Neurosci 6: 43. doi: [10.3389/fnsys.2012.00043](http://dx.doi.org/10.3389/fnsys.2012.00043) PMID: [22707934](http://www.ncbi.nlm.nih.gov/pubmed/22707934)
- 2. Craddock RC, Milham MP, LaConte SM (2013) Predicting intrinsic brain activity. Neuroimage 82: 127– 136. doi: [10.1016/j.neuroimage.2013.05.072](http://dx.doi.org/10.1016/j.neuroimage.2013.05.072) PMID: [23707580](http://www.ncbi.nlm.nih.gov/pubmed/23707580)
- 3. Ghasemi J, Ghaderi R, Mollaei MK, Hojjatoleslami S (2013) A novel fuzzy dempster-shafer inference system for brain {MRI} segmentation. Information Sciences 223: 205–220. doi: [10.1016/j.ins.2012.08.](http://dx.doi.org/10.1016/j.ins.2012.08.026) [026](http://dx.doi.org/10.1016/j.ins.2012.08.026)
- [4.](#page-1-0) Ortiz A, Gorriz J, Ramirez J, Salas-Gonzalez D (2014) Improving MR brain image segmentation using self-organising maps and entropy-gradient clustering. Information Sciences 262: 117–136. doi: [10.1016/j.ins.2013.10.002](http://dx.doi.org/10.1016/j.ins.2013.10.002)
- [5.](#page-1-0) Ball G, Aljabar P, Zebari S, Tusor N, Arichi T, Merchant N, et al. (2014) Rich-club organization of the newborn human brain. Proc Natl Acad Sci U S A 111: 7456–7461. doi: [10.1073/pnas.1324118111](http://dx.doi.org/10.1073/pnas.1324118111) PMID: [24799693](http://www.ncbi.nlm.nih.gov/pubmed/24799693)
- 6. Bargmann CI (2012) Beyond the connectome: how neuromodulators shape neural circuits. Bioessays 34: 458–465. doi: [10.1002/bies.201100185](http://dx.doi.org/10.1002/bies.201100185) PMID: [22396302](http://www.ncbi.nlm.nih.gov/pubmed/22396302)
- 7. Batalle D, Muñoz-Moreno E, Figueras F, Bargallo N, Eixarch E, Gratacos E (2013) Normalization of similarity-based individual brain networks from gray matter MRI and its association with neurodevelopment in infants with intrauterine growth restriction. Neuroimage 83: 901–911. doi: [10.1016/j.](http://dx.doi.org/10.1016/j.neuroimage.2013.07.045) [neuroimage.2013.07.045](http://dx.doi.org/10.1016/j.neuroimage.2013.07.045) PMID: [23886985](http://www.ncbi.nlm.nih.gov/pubmed/23886985)
- [8.](#page-1-0) Graham DJ (2014) Routing in the brain. Front Comput Neurosci 8: 44. doi: [10.3389/fncom.2014.00044](http://dx.doi.org/10.3389/fncom.2014.00044) PMID: [24782752](http://www.ncbi.nlm.nih.gov/pubmed/24782752)
- [9.](#page-1-0) Agosta F, Galantucci S, Valsasina P, Canu E, Meani A, Marcone A et al. (2014) Disrupted brain con-nectome in semantic variant of primary progressive aphasia. Neurobiol Aging. doi: [10.1016/j.](http://dx.doi.org/10.1016/j.neurobiolaging.2014.05.017) [neurobiolaging.2014.05.017](http://dx.doi.org/10.1016/j.neurobiolaging.2014.05.017)
- 10. Alexander-Bloch AF, Reiss PT, Rapoport J, McAdams H, Giedd JN, Bullmore Ed T. et al. (2014) Abnormal cortical growth in schizophrenia targets normative modules of synchronized development. Biol Psychiatry. doi: [10.1016/j.biopsych.2014.02.010](http://dx.doi.org/10.1016/j.biopsych.2014.02.010) PMID: [24690112](http://www.ncbi.nlm.nih.gov/pubmed/24690112)
- 11. Baker JT, Holmes AJ, Masters GA, Yeo BTT, Krienen F, Buckner, Randy L. et al. (2014) Disruption of cortical association networks in schizophrenia and psychotic bipolar disorder. JAMA Psychiatry 71: 109–118. doi: [10.1001/jamapsychiatry.2013.3469](http://dx.doi.org/10.1001/jamapsychiatry.2013.3469) PMID: [24306091](http://www.ncbi.nlm.nih.gov/pubmed/24306091)
- 12. Besson P, Dinkelacker V, Valabregue R, Thivard L, Leclerc X, Derambure P. et al. (2014) Structural connectivity differences in left and right temporal lobe epilepsy. Neuroimage 100C: 135–144. doi: [10.](http://dx.doi.org/10.1016/j.neuroimage.2014.04.071) [1016/j.neuroimage.2014.04.071](http://dx.doi.org/10.1016/j.neuroimage.2014.04.071)
- [13.](#page-1-0) Bonilha L, Nesland T, Rorden C, Fillmore P, Ratnayake RP, Fridriksson J. (2014) Mapping remote subcortical ramifications of injury after ischemic strokes. Behav Neurol 2014: 215380. doi: [10.1155/2014/](http://dx.doi.org/10.1155/2014/215380) [215380](http://dx.doi.org/10.1155/2014/215380) PMID: [24868120](http://www.ncbi.nlm.nih.gov/pubmed/24868120)
- [14.](#page-1-0) Ingalhalikar M, Smith A, Parker D, Satterthwaite TD, Elliott MA, Ruparel K, et al. (2014) Sex differences in the structural connectome of the human brain. Proc Natl Acad Sci U S A 111: 823–828. doi: [10.](http://dx.doi.org/10.1073/pnas.1316909110) [1073/pnas.1316909110](http://dx.doi.org/10.1073/pnas.1316909110) PMID: [24297904](http://www.ncbi.nlm.nih.gov/pubmed/24297904)
- [15.](#page-1-0) Joel D, Tarrasch R (2014) On the mis-presentation and misinterpretation of gender-related data: the case of Ingalhalikar's human connectome study. Proc Natl Acad Sci U S A 111: E637. doi: [10.1073/](http://dx.doi.org/10.1073/pnas.1323319111) [pnas.1323319111](http://dx.doi.org/10.1073/pnas.1323319111) PMID: [24477693](http://www.ncbi.nlm.nih.gov/pubmed/24477693)
- 16. Ingalhalikar M, Smith A, Parker D, Satterthwaite TD, Elliott MA, Ruparel K, et al. (2014) Reply to Joel and Tarrasch: On misreading and shooting the messenger. Proc Natl Acad Sci U S A 111: E638. doi: [10.1073/pnas.1323601111](http://dx.doi.org/10.1073/pnas.1323601111) PMID: [24672815](http://www.ncbi.nlm.nih.gov/pubmed/24672815)
- [17.](#page-1-0) Fine C (2014) Neuroscience. his brain, her brain? Science 346: 915–916.
- [18.](#page-1-0) Witelson SF, Beresh H, Kigar DL (2006) Intelligence and brain size in 100 postmortem brains: sex, lateralization and age factors. Brain 129: 386–398. doi: [10.1093/brain/awh696](http://dx.doi.org/10.1093/brain/awh696) PMID: [16339797](http://www.ncbi.nlm.nih.gov/pubmed/16339797)
- [19.](#page-2-0) Taki Y, Thyreau B, Kinomura S, Sato K, Goto R, Kawashima R, et al. (2011) Correlations among brain gray matter volumes, age, gender, and hemisphere in healthy individuals. PLoS One 6: e22734. doi: [10.1371/journal.pone.0022734](http://dx.doi.org/10.1371/journal.pone.0022734) PMID: [21818377](http://www.ncbi.nlm.nih.gov/pubmed/21818377)
- [20.](#page-2-0) Lawler EL (1976) Combinatorial optimization: networks and matroids. Courier Dover Publications.
- [21.](#page-2-0) Ford LR, Fulkerson DR (1956) Maximal flow through a network. Canadian Journal of Mathematics 8: 399–404. doi: [10.4153/CJM-1956-045-5](http://dx.doi.org/10.4153/CJM-1956-045-5)
- [22.](#page-2-0) Ade P, Aghanim N, Armitage-Caplan C, Arnaud M, Ashdown M, Atrio-Barandela, F, et al. (2013) Planck 2013 results. XVI. Cosmological parameters. arXiv preprint arXiv:13035076.
- [23.](#page-2-0) Garey MR, Johnson DS, Stockmeyer L (1976) Some simplified NP-complete graph problems. Theoretical computer science 1: 237–267. doi: [10.1016/0304-3975\(76\)90059-1](http://dx.doi.org/10.1016/0304-3975(76)90059-1)
- [24.](#page-20-0) Tarjan RE (1983) Data structures and network algorithms, volume 44 of CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial Applied Mathematics.
- [25.](#page-21-0) Hoory S, Linial N, Wigderson A (2006) Expander graphs and their applications. Bulletin of the American Mathematical Society 43: 439–561. doi: [10.1090/S0273-0979-06-01126-8](http://dx.doi.org/10.1090/S0273-0979-06-01126-8)
- [26.](#page-21-0) Lovász L (2007) Combinatorial problems and exercises. American Mathematical Society, 2nd edition.
- [27.](#page-21-0) Prüfer H (1918) Neuer Beweis eines Satzes über Permutationen. Arch Math Phys 27: 742–744.
- [28.](#page-21-0) Kirchhoff G (1847) über die Auflösung der Gleichungen, auf welche man bei der untersuchung der linearen verteilung galvanischer Ströme geführt wird. Ann Phys Chem 72.
- [29.](#page-21-0) Chung FR (1997) Spectral graph theory, volume 92. American Mathematical Soc.
- 30. Li R (2013) Lower bounds for the Kirchhoff index. MATCH Commun Math Comput Chem 70: 163–174.
- [31.](#page-21-0) Feng L, Gutman I, Yu G (2013) Degree Kirchhoff index of unicyclic graphs. Communications in Mathematical and in Computer Chemistry/MATCH 69: 629–648.
- [32.](#page-21-0) McNab JA, Edlow BL, Witzel T, Huang SY, Bhat H, Heberlein, Keith, et al. (2013) The Human Connec-tome Project and beyond: initial applications of 300 mT/m gradients. Neuroimage 80: 234-245. doi: [10.](http://dx.doi.org/10.1016/j.neuroimage.2013.05.074) [1016/j.neuroimage.2013.05.074](http://dx.doi.org/10.1016/j.neuroimage.2013.05.074) PMID: [23711537](http://www.ncbi.nlm.nih.gov/pubmed/23711537)
- [33.](#page-21-0) Daducci A, Gerhard S, Griffa A, Lemkaddem A, Cammoun L, Gigandet X, et al. (2012) The connectome mapper: an open-source processing pipeline to map connectomes with MRI. PLoS One 7: e48121. doi: [10.1371/journal.pone.0048121](http://dx.doi.org/10.1371/journal.pone.0048121) PMID: [23272041](http://www.ncbi.nlm.nih.gov/pubmed/23272041)
- [34.](#page-22-0) Tournier J, Calamante F, Connelly A, (2012) Mrtrix: diffusion tractography in crossing fiber regions. International Journal of Imaging Systems and Technology 22: 53–66. doi: [10.1002/ima.22005](http://dx.doi.org/10.1002/ima.22005)
- [35.](#page-23-0) Basser PJ, Pierpaoli C (2011) Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. 1996. J Magn Reson 213: 560–570.
- [36.](#page-25-0) Hochbaum DS (1982) Approximation algorithms for the set covering and vertex cover problems. SIAM Journal on Computing 11: 555–556. doi: [10.1137/0211045](http://dx.doi.org/10.1137/0211045)
- [37.](#page-25-0) Achterberg T, Berthold T, Koch T, Wolter K (2008) Constraint integer programming: A new approach to integrate CP and MIP. In: Integration of AI and OR techniques in constraint programming for combinatorial optimization problems, Springer. pp. 6–20.
- [38.](#page-25-0) Achterberg T (2009) Scip: solving constraint integer programs. Mathematical Programming Computation 1: 1–41. doi: [10.1007/s12532-008-0001-1](http://dx.doi.org/10.1007/s12532-008-0001-1)
- [39.](#page-26-0) Jbabdi S, Johansen-Berg H (2011) Tractography: where do we go from here? Brain Connect 1: 169– 183. doi: [10.1089/brain.2011.0033](http://dx.doi.org/10.1089/brain.2011.0033) PMID: [22433046](http://www.ncbi.nlm.nih.gov/pubmed/22433046)
- [40.](#page-26-0) Girard G, Whittingstall K, Deriche R, Descoteaux M (2014) Towards quantitative connectivity analysis: reducing tractography biases. Neuroimage 98: 266–278. doi: [10.1016/j.neuroimage.2014.04.074](http://dx.doi.org/10.1016/j.neuroimage.2014.04.074) PMID: [24816531](http://www.ncbi.nlm.nih.gov/pubmed/24816531)
- [41.](#page-26-0) Hoel PG (1984) Introduction to mathematical statistics. John Wiley & Sons, Inc., New York, 5fth edition.
- [42.](#page-27-0) Wonnacott TH, Wonnacott RJ (1972) Introductory statistics, volume 19690. Wiley New York.
- [43.](#page-27-0) Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics: 65–70.

Copyright of PLoS ONE is the property of Public Library of Science and its content may not be copied or emailed to multiple sites or posted to ^a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.