RESEARCH ARTICLE

The frequent complete subgraphs in the human connectome

Máté Fellner[1], Bálint Varga[1], Vince Grolmusz[1,2]*

1 PIT Bioinformatics Group, Eötvös University, Budapest, Hungary, **2** Uratim Ltd., Budapest, Hungary

* grolmusz@pitgroup.org

Abstract

While it is still not possible to describe the neuronal-level connections of the human brain, we can map the human connectome with several hundred vertices, by the application of diffusion-MRI based techniques. In these graphs, the nodes correspond to anatomically identified gray matter areas of the brain, while the edges correspond to the axonal fibers, connecting these areas. In our previous contributions, we have described numerous graph-theoretical phenomena of the human connectomes. Here we map the frequent complete subgraphs of the human brain networks: in these subgraphs, every pair of vertices is connected by an edge. We also examine sex differences in the results. The mapping of the frequent subgraphs gives robust substructures in the graph: if a subgraph is present in the 80% of the graphs, then, most probably, it could not be an artifact of the measurement or the data processing workflow. We list here the frequent complete subgraphs of the human braingraphs of 413 subjects (238 women and 175 men), each with 463 nodes, with a frequency threshold of 80%, and identify 812 complete subgraphs, which are more frequent in male and 224 complete subgraphs, which are more frequent in female connectomes.

Introduction

Diffusion MRI-based macroscopic mapping of the connections of the human brain is a technology that was developed in the last 15 years [1–4]. Applying the method, we are able to construct braingraphs, or connectomes, from the diffusion MRI images [1, 5, 6]: the vertices of the graph are anatomically labeled areas of the gray matter (called "Regions of Interests", ROIs), and two such ROIs are connected by an edge, if a complex workflow, involving either deterministic or probabilistic tractography, finds axonal fibers between them. Therefore, one can construct graphs, with up to 1015 nodes and several thousand edges, from the MR image of each subject.

The analysis of these graphs is a fast-developing and an important area today: these connections form the "hardware" of all brain functions on a macroscopic level [2–4]. Naturally, it would be exciting to map the neuronal scale human connectome, too: here the nodes were the individual neurons, and two nodes (or neurons), say X and Y, would be connected by a directed edge, say (X, Y), if the axon of X were connected to a dendrite of Y. Unfortunately, to date, the neuronal-level connectome of only one adult organism is described: that of the

nematode *Caenorhabditis elegans* with 302 neurons, in the year 1986 [7]. In larval state, two more neuronal level connectomes are published: the larva of the fruitfly *Drosophila melanogaster* [8], and the tadpole larva of the *Ciona intestinalis* [9]. Despite of some exciting, very recent developments [10], the complete connectome of the adult *Drosophila melanogaster* with 100,000 neurons, is not available yet. Humans have 80 billion neurons in their brains. Therefore, the mapping and the analysis of the neuronal scale human connectome is out of our reach today.

There are numerous results published for the analysis of the diffusion MRI-computed human connectomes, e.g., [1–4]. Our research group has also contributed some more graph-theoretically oriented analytical methods, like the comparison of the deep graph theoretical parameters of male and female connectomes [11–13], the parameterizable human consensus connectome [14, 15], the description of the individual variability in the connections of the major lobes [16], the discovery of the Consensus Connectome Dynamics [17–20] the description of the frequent subgraphs of the human brain [21], and the Frequent Neighborhood Mapping of the human hippocampus [22].

Frequent edges and subgraphs: A robust analysis

The data acquisition and processing workflow, whose results are the braingraphs or structural connectomes, has numerous delicate steps. Naturally, errors may occur in MRI recordings and processing, as well as in segmentation, parcellation, tractography and graph computation steps [1, 23, 24]. When we have hundreds of high-quality MR images, we can analyze the *frequently appearing* graph edges or subgraphs, in order to derive robust, reproducible results, appearing in high fraction of the brains imaged. By analyzing only the frequently appearing structural elements, the great majority of data acquisition and processing errors will be filtered out.

Our first effort for describing frequent edges in human connectome was the construction of the Budapest Reference Connectome Server [14, 15], in which the user can select the frequency threshold $k\%$ of the edges, and the resulting consensus connectome contains only those edges, which are present in at least $k\%$ of the subjects. The generated consensus connectome can be both visualized and downloaded at the site http://pitgroup.org/connectome/.

The frequent, connected subgraphs of at most 6 edges are mapped in the human connectome in [21]. The frequencies were compared between female and male connectomes, and strong sex differences were identified: there are connected subgraphs, which are significantly more frequent in males than in females, and there are a higher number of connected graphs that are more frequent in females than in males.

The direct connections of important brain areas are of special interest: correlations between the connections and biological parameters may enlighten the fine structure-function relations of our brain. For error-correction reasons, the frequent neighbors of the relevant brain areas form the robust objects of study: small errors in the data processing workflow will most probably have no effects on the frequent connections. In our work [22], we have introduced the method of the Frequent Neighborhood Mapping, which describes the frequent neighbor sets of the given nodes of the braingraph. In [22], we have demonstrated the method by mapping the frequent neighborhoods of the human hippocampus: one of the most deeply studied part of the brain. We have mapped the frequent neighbor sets of the hippocampus, and we have found sex differences in the frequent neighbor sets: males have much more frequent neighbor sets of the hippocampus than the females; therefore, the neighborhoods of the men's hippocampi are more regular, with less variability than those of women. This observation is in line with the results of [11–13], where we have shown that the female connectomes are better expander graphs than the braingraphs of men.

In the present contribution, we are mapping the frequent complete graphs of the human connectome, based on the large dataset of the Human Connectome Project [25]. Our dataset contains the braingraphs of 413 subjects. A recently appeared work [26] deals with complete subgraphs in braingraphs of 8 subjects, each with 83 nodes. Our results are derived from 413 braingraphs, each of 463 nodes. Therefore, we are able to find frequent structures, i.e., frequent complete subgraphs in our dataset of 413 graphs (while it is not feasible to derive frequent structures from only 8 graphs).

Cliques vs. complete subgraphs

Here we intend to clarify some graph theoretical terms. A complete graph on v vertices contains (undirected) edges, connecting all the $\binom{v}{2} = v(v-1)/2$ vertex-pairs: that is, in a complete graph, each pair of vertices are connected by an edge.

If we have a graph G on n vertices, we can look for the complete subgraphs H of G: all the vertices and the edges of H need to be vertices and edges of G (i.e., H is a subgraph of G), and, moreover, H needs to be a complete graph.

The complete subgraph of the maximum vertex-number of G is called a clique. The clique number of graph G, $\omega(G)$, is the number of vertices in the largest complete subgraph of G. Computing the clique number $\omega(G)$ is a well-known hard problem: it is NP-hard [27], that is, it is not probable that one could find a fast (i.e., polynomial-time) algorithm for computing $\omega(G)$. Moreover, in general, not only the exact value of $\omega(G)$ is hard to compute, but it is also very difficult to approximate, even roughly [28]. In special cases, however, when the number of the vertices is only several hundred, and the graph is not too dense, that is, it has not too many edges, then all the frequently appearing complete subgraphs can be computed relatively quickly by the *apriori* algorithm [29, 30]. The computational details are given in the Materials and Methods section.

Our goal in the present contribution is to map the frequently appearing complete subgraphs in human connectomes. We need to make clear that our analysis is done on 463-vertex braingraphs. Therefore, if a complete subgraph is found, it does not imply the neuronal level existence of complete subgraphs. It implies, however, that the macroscopic ROIs, corresponding to the vertices of the complete graphs discovered, are connected densely to each another, probably even on the neuronal level.

In the literature one may find numerous references to the "rich club property" of some networks, related to the braingraph [31, 32]. Here we prefer using classical graph theoretical terms and definitions instead of this "rich club property", consequently, we intend to map those densely connected subgraphs of the human connectomes, which form complete graphs, and appear in at least the 80% of the all braingraphs considered.

Materials and methods

The data source and the graph computation

The data source of the present study is the website of the Human Connectome Project at the address http://www.humanconnectome.org/documentation/S500 [25]. The dataset contains the high angular resolution diffusion imaging (HARDI) MRI data of 413 healthy human young adults between the ages of 22 and 35 years. We have examined the data of 238 women and 175 men.

The CMTK toolkit [5], together with the FreeSurfer tool and the MRtrix tractography processing tool [33] were applied in the graph generation. In the MRtrix tool, we have applied

randomized seeding and deterministic streamline method, with 1 million streamlines. We have studied here graphs with 463-vertex resolution. The parcellation data is given in the CMTK nypipe GitHub repository https://github.com/LTS5/cmp_nipype/blob/master/cmtklib/data/parcellation/lausanne2008/ParcellationLausanne2008.xls.

Further particularities of the graph processing workflow are described in [6], where the http://braingraph.org repository is also covered. The braingraphs, analyzed here, can be accessed at the http://braingraph.org/cms/download-pit-group-connectomes/ site, choosing the "Full set, 413 brains, 1 million streamlines" option.

The algorithm

We have mentioned that computing the size of the largest complete subgraph, called the *clique-number*, and denoted by $\omega(G)$, is NP-hard [27]. Naturally, finding the largest complete subgraph itself cannot be easier than finding its size $\omega(G)$; therefore, it is also NP-hard.

Finding the largest complete subgraphs in sparse graphs (i.e., graphs with relatively few edges, compared to the number of its vertices) is usually not a very difficult task, since in these graphs, regularly, there are not too many large complete subgraphs. Finding only the frequently appearing complete subgraphs further simplifies the computational tasks, and we can apply an algorithm, which resembles in many points to the apriori algorithm for finding frequent item sets [29, 30], and this algorithm is very fast in the practice.

Now we describe the algorithm: A frequent complete subgraph is characterized by the list of its vertices and the set of its edges. At the beginning, for an (undirected) edge (v_i, v_j), let these two lists be given as $([v_i, v_j], \{(v_i, v_j)\})$, where $i < j$.

In general, the vertices of the complete subgraphs are listed in the increasing order of their indices, and the vertices of each edge are listed also in the increasing order of its indices.

Now we describe the generating, "apriori" step. Let

$$L_1 = ([v_1, v_2, ..., v_k], \{(v_1, v_2), (v_1, v_3), ...(v_{k-1}, v_k)\})$$

and

$$L_2 = ([u_1, u_2, ..., u_k], \{(u_1, u_2), (u_1, u_3), ...(u_{k-1}, u_k)\})$$

be two frequent complete subgraphs of size k. If the first $k − 1$ vertices of L_1 and L_2 are the same, and the last ones differ, we will consider generating a new, $k + 1$-vertex complete graph, as follows: if $v_1 = u_1, v_2 = u2, \ldots v_{k-1} = u_{k-1}$ and $v_k \neq u_k$, then, by the notation $v_{k+1} = u_k$, we verify the suitable frequency of the complete graph $L = ([v_1, v_2, \ldots v_k, v_{k+1}], \{(v_1, v_2), (v_1, v_3), \ldots (v_k, v_{k+1})\})$.

It is easy to see that in the edge list only the last one, (v_k, v_{k+1}) is new, all the others are already the edges of the frequent subgraphs L_1 or L_2.

In generating L one needs to make sure that the vertices in the vertex-list are ordered by their indices, and that the frequency of L is above the inclusion threshold.

The apriori generating step is correct, since if L is frequent, then both L_1 and L_2 were frequent. Additionally, every $k + 1$-vertex complete graph is generated only once, since the vertices are in increasing order: $(v_1, v_2, \ldots, v_{k+1})$ can be generated only from $(v_1, \ldots v_k)$ and $(v_1, \ldots v_{k-1}, v_{k+1})$.

Statistical notes

The frequent complete subgraphs were chosen in the following way:

First, the subjects were partitioned into two disjoint sets, by the parity of their ID number's second digit from the right. Next, in both sets, the complete graphs with the minimum

frequency of 80% were identified, as it was described in the previous section. Only those complete subgraphs were retained, which have had a minimum frequency of 80% in *both* sets under consideration. Then the frequency of these subgraphs were re-calculated for the whole dataset: these frequencies are given in the supplementary tables.

We used this approach for increasing the robustness of our analysis. The two sets model two complementing random subsets, and if a complete subgraph is frequent in both subsets then it is more uniformly frequent than it would have been counted only in the whole set of braingraphs.

In the computation of sex differences, we have applied χ^2 tests to identify significant differences in the frequencies of the complete subgraphs, similarly as in the work [22]:

For a chosen frequent complete subgraph *F*, we have counted its occurrences in the male dataset by $count_1(F)$ and in the female dataset by $count_2(F)$. The support was calculated as $supp_i(F) = \frac{count_i(F)}{S_i}$, where S_i, for $i = 1, 2$, is the number of male and female braingraphs, respectively. For each graph *F* we need to determine whether $supp_1(F)$ and $supp_2(F)$ significantly differ. For this goal we used the chi-squared test for categorical data:

	contains F	does not contain F	total
1st sample	$count_1(F)$	$S_1 - count_1(F)$	S_1
2nd sample	$count_2(F)$	$S_2 - count_2(F)$	S_2
total	$count_1(F) + count_2(F)$	$S_1 + S_2 - count_1(F) - count_2(F)$	$S_1 + S_2$

Then the test is calculated as

$$\chi^2 = \frac{\left(count_1(F) \cdot (S_2 - count_2(F)) - count_2(F) \cdot (S_1 - count_1(F))\right)^2 \cdot (S_1 + S_2)}{S_1 \cdot S_2 \cdot (count_1(F) + count_2(F)) \cdot (S_1 + S_2 - count_1(F) - count_2(F))}$$

Our null hypothesis was that the frequencies are the same in males and females, and we refute this hypothesis with p = 0.01. The secondary statistical errors were handled by Holm-Bonferroni corrections [34]. The un-corrected and the corrected p values are listed in S2 and S3 Tables.

Since we have more female subjects (238) than male subjects (175), the 80% frequency limit is less restrictive in the case of males. Consequently, our results may be influenced by the difference in the number of males and females. In order to deal with this assumption, we also have chosen randomly 50 female braingraphs for set 1, and 50 female braingraphs for set 2, such that set 1 and set 2 are disjoint sets. Next, in both sets, the complete subgraphs with the minimum frequency of 80% were identified, as it was described in the previous section. Only those complete subgraphs were retained, which have had a minimum frequency of 80% in *both* sets under consideration. Then the frequency of these subgraphs were re-calculated for the whole dataset. We have repeated the random choices and the computations 20 times for females.

The same random choices and computations were performed for male braingraphs (i.e., the random choice of two disjoint 50-member sets from the male subgraphs, the computations of the frequent complete subgraphs in both sets, and the selections of those complete graphs, which were frequent in both sets). The resulting counts of the complete graphs of different cardinalities and the number of complete frequent graphs with significantly (p = 0.01) differing frequencies are given in the S4 Table.

Discussion and results

First we review the frequent complete subgraphs of the human braingraph, next we analyze the significant differences in their frequencies in males and in females.

Frequent complete subgraphs of the human connectome

S1 Table contains the complete subgraphs of the human connectomes appearing in at least 80% of the graphs of the 413 subjects examined. In each row, the vertices of the complete subgraphs are listed, together with their frequencies of appearance. Note, that the vertices of a complete graph uniquely determine its edges. The list is redundant in the following sense: if a k-vertex complete graph has frequency at least 80%, then all of its complete subgraphs are also listed. We find that this redundancy helps in the analysis of the results, as it will be clear from what follows.

We would like to emphasize the following very simple, but powerful fact: If a given subgraph U has a frequency, say ℓ%, then all subgraphs of U has frequency at least ℓ%. This is the central point in the apriori algorithm [29, 30], and it was noted and applied in [21, 22].

The ROIs in S1 Table carry the names of the resolution-250 parcellation labels (where the number 250 refers to the approximate number of vertices in each hemisphere; the graphs of resolution-250 contain 465 vertices, not just 250), based on the Lausanne 2008 brain atlas [35] and computed by using FreeSurfer [36] and CMTK [5, 37], given at https://github.com/LTS5/cmp_nipype/blob/master/cmtklib/data/parcellation/lausanne2008/ParcellationLausanne2008.xls. The "lh" and the "rh" prefixes abbreviate the "left-hemisphere" and "right-hemisphere" terms of localizations.

Complete subgraphs appearing in each subject

Here we list the maximal complete subgraphs from S1 Table, which are present in all of the braingraphs, and contains at least three nodes. Subgraph R1 is visualized in Fig 1.

L1: (Left-Caudate)(Left-Pallidum)(Left-Putamen)(Left-Thalamus-Proper),

L2: (Left-Hippocampus)(Left-Putamen)(Left-Thalamus-Proper),

L3: (Left-Putamen)(Left-Thalamus-Proper)(lh.insula_1)

R1: (Right-Caudate)(Right-Pallidum)(Right-Putamen)(Right-Thalamus-Proper)

R2: (Right-Hippocampus)(Right-Putamen)(Right-Thalamus-Proper)

R3: (Right-Putamen)(Right-Thalamus-Proper)(rh.insula_2)

R4: (rh.superiorfrontal_7)(rh.superiorfrontal_8)(rh.superiorfrontal_9)

Note that L1 and L2 correspond to R1 and R2, and L3 almost corresponds to R3. Complete graph R4 has no correspondence in the left hemisphere (which are present in each subject), but in the left hemisphere, the superiorfrontal regions are also connected densely, as one can verify easily from S1 Table.

We believe that the connections between the above-listed areas are very strong in each subject: so strong that they are not affected by measurement errors and individual variability.

The largest frequent complete subgraphs

The largest complete subgraphs, which are present in at least the 80% of the subjects, have seven vertices, and they are located in the left hemisphere. The first one connects the left
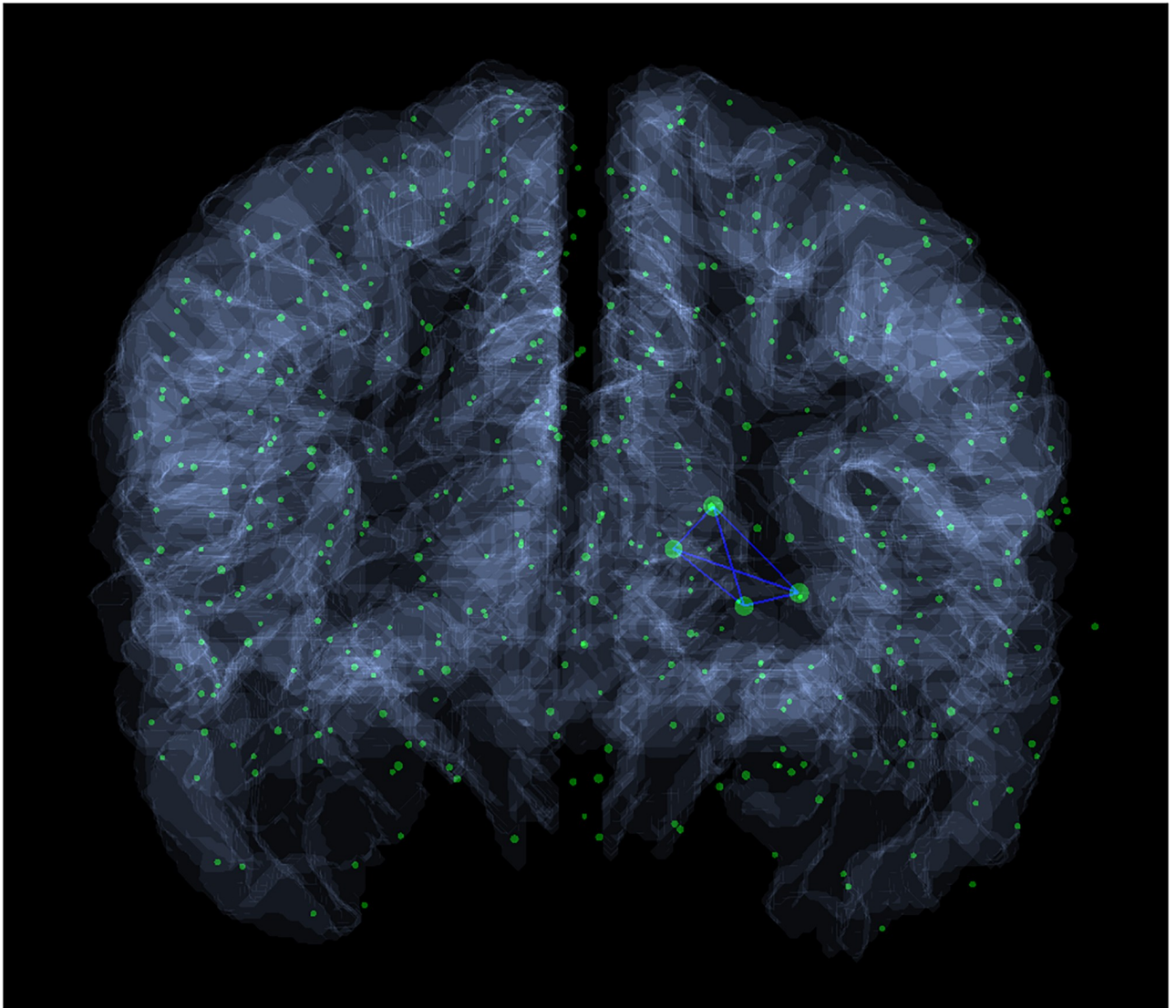
Fig 1. A complete subgraph with four vertices, which is present in all 413 human connectomes examined. The vertices of the depicted subgraph are Right-Caudate, Right-Pallidum, Right-Putamen, Right-Thalamus-Proper. The supporting S1 Table contains all the complete subgraphs with frequency of at least 80%, S2 Table contains the complete subgraphs, where the frequency of their appearance in females is significantly higher (p = 0.01) than in males; S3 Table contains those, where the frequency is significantly higher in males than in females. The supporting tables are available at pdf and Excel formats at http://uratim.com/cliques/tables.zip.

https://doi.org/10.1371/journal.pone.0236883.g001

putamen with six vertices in the left frontal lobe (B1), the second one connects the left caudate and the left putamen ROIs to five left frontal areas (B2):

B1: (Left-Putamen) (lh.lateralorbitofrontal_4) (lh.lateralorbitofrontal_6) (lh.lateralorbitofrontal_7) (lh.parstriangularis_3) (lh.rostralmiddlefrontal_12) (lh.rostralmiddlefrontal_9)

B2: (Left-Caudate) (Left-Putamen) (lh.lateralorbitofrontal_7) (lh.medialorbitofrontal_2) (lh.rostralanteriorcingulate_1) (lh.rostralmiddlefrontal_12) (lh.rostralmiddlefrontal_9)

There are 48 different 6-vertex complete subgraphs, which are present in at least 80% of the connectomes. Only 6 of these are situated in the right hemisphere, the other 42 are in the left hemisphere.

Complete subgraphs across the hemispheres

Since the neural fiber tracts, connecting the two hemispheres of the brain, are very dense in the corpus callosum, their tractography in the diffusion MR images is difficult since the fiber-crossings cannot always be tracked reliably [38, 39].

We have found only relatively few frequent complete subgraphs of the human connectome, which have nodes from both hemispheres. Here we list those, which are present in more than 80% of the braingraphs studied; therefore, they are most probably not false positives. Again, we are listing only the maximal complete subgraphs for clarity. We note that most ROIs in the list are the parts of the striatum: each complete subgraph contains either a caudate nucleus or a nucleus accumbens of either the right- or the left hemisphere:

A1: (Left-Accumbens-area)(Left-Caudate)(Left-Thalamus-Proper)(Right-Caudate)

A2: (Left-Accumbens-area)(Left-Caudate)(Right-Caudate)(lh.rostralanteriorcingulate_1)

A3: (Left-Accumbens-area)(Left-Thalamus-Proper)(Right-Thalamus-Proper)

A4: (Left-Caudate)(Left-Thalamus-Proper)(Right-Caudate)(Right-Thalamus-Proper)

A5: (Left-Caudate)(Right-Caudate)(lh.caudalanteriorcingulate_1)(lh.
   caudalanteriorcingulate_2)

A6: (Left-Caudate)(Right-Caudate)(lh.caudalanteriorcingulate_1)(lh.
   rostralanteriorcingulate_1)

A7: (Left-Caudate)(Right-Caudate)(rh.caudalanteriorcingulate_1)

A8: (Left-Caudate)(Right-Caudate)(rh.rostralanteriorcingulate_2)

A9: (Left-Thalamus-Proper)(Right-Accumbens-area)(Right-Thalamus-Proper)

Counts of the hippocampus, thalamus, putamen, pallidum and the amygdala in the frequent complete subgraphs

In this section we count the appearances of certain ROIs in the frequent complete subgraphs, with a frequency threshold of 80%. Our results show that there are considerable differences between the hemispheres in these numbers: The right hippocampus and the right amygdala are present in much more complete subgraphs than the left ones; the left thalamus-proper, the left putamen and the left pallidum are present in much more complete subgraphs than the right ones (Table 1).

Table 1. The number of appearances of the hippocampus, the amygdala, the thalamus-proper, the putamen and the pallidum in the frequent complete subgraphs, with a frequency threshold of 80%, in each hemisphere. The right hippocampus and the right amygdala are present in much more complete subgraphs than the left ones; the left thalamus-proper, the left putamen and the left pallidum are present in much more complete subgraphs than the right ones.

	Left	Right
Hippocampus	187	247
Amygdala	66	99
Thalamus-Proper	265	175
Putamen	1041	673
Pallidum	149	123

Sex differences

Mapping sex differences in the human connectome is a hot and fast-developing area of research. In our earlier works we have shown—first in the literature—that in numerous well-defined graph theoretical parameters, women have "better connected" braingraphs than men [11–13]. In the work [21] we have mapped the frequent subgraphs of the human brain of at most 6 vertices, and have found sex differences: there are numerous frequent connected subgraphs, which are more frequent in men than in women, and, similarly, which are more frequent in men than in women. In the study of [22], we have mapped the neighbor-sets of the human hippocampus and found also significant sex differences in these sets.

Here we compare the frequencies of the complete subgraphs of the connectomes of men and women. We have found significant differences in the frequencies of some complete subgraphs, with the advantage at men and women, too.

We have found much more complete subgraphs with significantly higher frequency in men than in women. More exactly, S2 Table lists 224 complete subgraphs, with significantly higher frequency in females than in males, while S3 Table lists 812 complete subgraphs, where their frequencies in males were higher than in females (with p = 0.01, and the inclusion threshold was a minimum 80% for the larger frequency).

This observation, in a sense, shows that men's connectomes show less inter-personal variability in complete subgraphs than those of women. This observation is in contrast with our findings in [21], where we have shown that women have much more 6-vertex frequent subgraphs than men: but in [21] we required connectedness, and not completeness.

In S4 Table we present the result of the crosscheck of the sex differences, where the count of the male and female braingraphs were the same (c.f. the last two paragraphs of the Methods section). In S4 Table the results of 20 runs are presented: the female plus and male plus rows show the number of frequent complete subgraphs with significantly higher frequencies (p = 0.01) in females and males, respectively. One can observe that in this smaller control setting males again have much more significant frequent complete subgraphs than females; therefore, the results in S2 and S3 Tables are due to sex differences, and not to cohort-size differences.

Conclusions

By an apriori-like algorithm, we have mapped the frequent (>80%) complete subgraphs of 413 subjects, each with 463 vertices. The largest frequent complete subgraph has 7 vertices. Most of the largest frequent subgraphs are located in the left hemisphere. We have also identified the frequent complete subgraphs, containing vertices from both hemispheres, and identified complete subgraphs with significant frequency-differences between the sexes. We have found that men have much more frequent complete subgraphs than women: this result contrasts our earlier finding [11], where we have shown that women have much better connectivity-related parameters in their connectomes than men in a following sense: while women have better connected braingraphs than men (as it is very precisely described in [11]), the dense subgraphs of men show less inter-individual variability than in women. The complex set of the inter-neuronal connections of the brain cannot be examined today in its entirety. On a much coarser scale, we can study these connections, and we can find frequent subgraphs, which may carry significance in the brain structure and functions. It might be interesting to identify other frequent subgraphs in the future.

Supporting information

S1 Table. Contains the list of all the complete subgraphs of 413 human connectomes with a minimum frequency of 80%.
(PDF)

S2 Table. Contains the complete subgraphs, where the frequency of their appearance in females is significantly higher (p = 0.01) than in males.
(PDF)

S3 Table. Contains those, where the frequency is significantly higher in males than in females. In both S2 and S3 Tables a frequency cut-off 80% is applied to the larger frequency of the appearance in the sexes: only those significant differences are listed, where the larger of the frequencies of males and females are at least 80%.
(PDF)

S4 Table. In S4 Table we present the result of the crosscheck of the sex differences, where the count of the male and female braingraphs were the same (c.f. the last two paragraphs of the Methods section). In S4 Table the results of 20 runs are presented: the female plus and male plus rows show the number of frequent complete subgraphs with significantly higher frequencies (p = 0.01) in females and males, respectively.
(PDF)

Author Contributions

Conceptualization: Bálint Varga, Vince Grolmusz.

Data curation: Bálint Varga.

Formal analysis: Máté Fellner.

Funding acquisition: Vince Grolmusz.

Investigation: Máté Fellner, Vince Grolmusz.

Methodology: Máté Fellner, Bálint Varga.

Project administration: Vince Grolmusz.

Resources: Vince Grolmusz.

Software: Máté Fellner.

Supervision: Vince Grolmusz.

Validation: Máté Fellner, Bálint Varga.

Visualization: Bálint Varga.

Writing – original draft: Vince Grolmusz.

Writing – review & editing: Vince Grolmusz.

References

1. Hagmann P, Grant PE, Fair DA. MR connectomics: a conceptual framework for studying the developing brain. Front Syst Neurosci. 2012; 6:43. Available from: http://dx.doi.org/10.3389/fnsys.2012.00043 PMID: 22707934

2. Seung HS. Reading the book of memory: sparse sampling versus dense mapping of connectomes. Neuron. 2009 Apr; 62(1):17–29. Available from: http://dx.doi.org/10.1016/j.neuron.2009.03.020 PMID: 19376064

3. Sporns O, Tononi G, Kötter R. The human connectome: A structural description of the human brain. PLoS Computational Biology. 2005 Sep; 1(4):e42. Available from: http://dx.doi.org/10.1371/journal.pcbi.0010042 PMID: 16201007

4. Lichtman JW, Livet J, Sanes JR. A technicolour approach to the connectome. Nature Review Neuroscience. 2008 Jun; 9(6):417–422. Available from: http://dx.doi.org/10.1038/nrn2391

5. Daducci A, Gerhard S, Griffa A, Lemkaddem A, Cammoun L, Gigandet X, et al. The connectome mapper: an open-source processing pipeline to map connectomes with MRI. PLoS One. 2012; 7(12): e48121. Available from: http://dx.doi.org/10.1371/journal.pone.0048121 PMID: 23272041

6. Kerepesi C, Szalkai B, Varga B, Grolmusz V. The braingraph. org Database of High Resolution Structural Connectomes and the Brain Graph Tools. Cognitive Neurodynamics. 2017; 11(5):483–486. https://doi.org/10.1007/s11571-017-9445-1

7. White J, Southgate E, Thomson J, Brenner S. The structure of the nervous system of the nematode Caenorhabditis elegans: the mind of a worm. Phil Trans R Soc Lond. 1986; 314:1–340.

8. Ohyama T, Schneider-Mizell CM, Fetter RD, Aleman JV, Franconville R, Rivera-Alba M, et al. A multi-level multimodal circuit enhances action selection in Drosophila. Nature. 2015 Apr; 520:633–639. https://doi.org/10.1038/nature14297 PMID: 25896325

9. Ryan K, Lu Z, Meinertzhagen IA. The CNS connectome of a tadpole larva of Cona intestinalis (L.) highlights sidedness in the brain of a chordate sibling. eLife. 2016 Dec;5.

10. Zheng Z, Lauritzen JS, Perlman E, Robinson CG, Nichols M, Milkie D, et al. A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster. Cell. 2018 Jul; 174:730–743.e22. https://doi.org/10.1016/j.cell.2018.06.019 PMID: 30033368

11. Szalkai B, Varga B, Grolmusz V. Graph Theoretical Analysis Reveals: Women's Brains Are Better Connected than Men's. PLoS One. 2015; 10(7):e0130045. Available from: http://dx.doi.org/10.1371/journal.pone.0130045 PMID: 26132764

12. Szalkai B, Varga B, Grolmusz V. The Graph of Our Mind. arXiv preprint arXiv:160300904. 2016;.

13. Szalkai B, Varga B, Grolmusz V. Brain Size Bias-Compensated Graph-Theoretical Parameters are Also Better in Women's Connectomes. Brain Imaging and Behavior. 2018; 12(3):663–673. Available from: http://dx.doi.org/10.1007/s11682-017-9720-0 PMID: 28447246

14. Szalkai B, Kerepesi C, Varga B, Grolmusz V. The Budapest Reference Connectome Server v2. 0. Neuroscience Letters. 2015; 595:60–62. https://doi.org/10.1016/j.neulet.2015.03.071

15. Szalkai B, Kerepesi C, Varga B, Grolmusz V. Parameterizable Consensus Connectomes from the Human Connectome Project: The Budapest Reference Connectome Server v3.0. Cognitive Neurodynamics. 2017 feb; 11(1):113–116. https://doi.org/10.1007/s11571-016-9407-z

16. Kerepesi C, Szalkai B, Varga B, Grolmusz V. Comparative Connectomics: Mapping the Inter-Individual Variability of Connections within the Regions of the Human Brain. Neuroscience Letters. 2018; 662 (1):17–21.

17. Kerepesi C, Szalkai B, Varga B, Grolmusz V. How to Direct the Edges of the Connectomes: Dynamics of the Consensus Connectomes and the Development of the Connections in the Human Brain. PLOS One. 2016 June; 11(6):e0158680. Available from: http://dx.doi.org/10.1371/journal.pone.0158680 PMID: 27362431

18. Kerepesi C, Varga B, Szalkai B, Grolmusz V. The Dorsal Striatum and the Dynamics of the Consensus Connectomes in the Frontal Lobe of the Human Brain. Neuroscience Letters. 2018 March; 673:51–55. https://doi.org/10.1016/j.neulet.2018.02.052

19. Szalkai B, Kerepesi C, Varga B, Grolmusz V. High-Resolution Directed Human Connectomes and the Consensus Connectome Dynamics. PLoS ONE. 2019 Sep; 14(4). Available from: https://doi.org/10.1371/journal.pone.0215473 PMID: 30990832

20. Szalkai B, Varga B, Grolmusz V. The Robustness and the Doubly-Preferential Attachment Simulation of the Consensus Connectome Dynamics of the Human Brain. Scientific Reports. 2017; 7(16118). https://doi.org/10.1038/s41598-017-16326-0 PMID: 29170405

21. Fellner M, Varga B, Grolmusz V. The Frequent Subgraphs of the Connectome of the Human Brain. Cognitive Neurodynamics. 2019; 13(5):453–460. Available from: https://doi.org/10.1007/s11571-019-09535-y PMID: 31565090

22. Fellner M, Varga B, Grolmusz V. The frequent network neighborhood mapping of the human hippocampus shows much more frequent neighbor sets in males than in females. PLOS ONE. 2020; 15(1): e0227910. Available from: https://doi.org/10.1371/journal.pone.0227910 PMID: 31990956

23. Jbabdi S, Johansen-Berg H. Tractography: where do we go from here? Brain Connectivity. 2011; 1 (3):169–183. Available from: http://dx.doi.org/10.1089/brain.2011.0033 PMID: 22433046

24. Mangin JF, Fillard P, Cointepas Y, Le Bihan D, Frouin V, Poupon C. Toward global tractography. Neuroimage. 2013 Oct; 80:290–296. Available from: http://dx.doi.org/10.1016/j.neuroimage.2013.04.009 PMID: 23587688

25. McNab JA, Edlow BL, Witzel T, Huang SY, Bhat H, Heberlein K, et al. The Human Connectome Project and beyond: initial applications of 300 mT/m gradients. Neuroimage. 2013 Oct; 80:234–245. Available from: http://dx.doi.org/10.1016/j.neuroimage.2013.05.074 PMID: 23711537

26. Sizemore AE, Giusti C, Kahn A, Vettel JM, Betzel RF, Bassett DS. Cliques and cavities in the human connectome. Journal of Computational Neuroscience. 2018; 44(1):115–145. https://doi.org/10.1007/s10827-017-0672-6

27. Garey MR, Johnson DS. Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman; 1979.

28. Håstad J. Clique is Hard to Approximate Within $n^{1-epsilon}$. In: 37th Annual Symposium on Foundations of Computer Science, FOCS'96, Burlington, Vermont, USA, 14-16 October, 1996. IEEE Computer Society; 1996. p. 627–636.

29. Agrawal R, Imielinski T, Swami AN. Mining Association Rules between Sets of Items in Large Databases. In: Buneman P, Jajodia S, editors. Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, D.C., May 26-28, 1993. ACM Press; 1993. p. 207–216.

30. Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. In: Bocca JB, Jarke M, Zaniolo C, editors. Proc. of the 20th International Conference on Very Large Data Bases (VLDB'94),. vol. 1215. Kaufmann Publishers Inc.,; 1994. p. 487–499.

31. Ball G, Aljabar P, Zebari S, Tusor N, Arichi T, Merchant N, et al. Rich-club organization of the newborn human brain. Proc Natl Acad Sci U S A. 2014 May; 111(20):7456–7461. Available from: http://dx.doi.org/10.1073/pnas.1324118111 PMID: 24799693

32. van den Heuvel MP, Sporns O. Rich-club organization of the human connectome. J Neurosci. 2011 Nov; 31(44):15775–15786. Available from: http://dx.doi.org/10.1523/JNEUROSCI.3539-11.2011 PMID: 22049421

33. Tournier J, Calamante F, Connelly A, et al. MRtrix: diffusion tractography in crossing fiber regions. International Journal of Imaging Systems and Technology. 2012; 22(1):53–66. https://doi.org/10.1002/ima.22005

34. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 1979; 6(2):65–70. Available from: https://www.jstor.org/stable/4615733.

35. Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, et al. Mapping the structural core of human cerebral cortex. PLoS Biol. 2008 Jul; 6(7):e159. Available from: http://dx.doi.org/10.1371/journal.pbio.0060159 PMID: 18597554

36. Fischl B. FreeSurfer. Neuroimage. 2012; 62(2):774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

37. Gerhard S, Daducci A, Lemkaddem A, Meuli R, Thiran JP, Hagmann P. The connectome viewer toolkit: an open source framework to manage, analyze, and visualize connectomes. Frontiers in Neuroinformatics. 2011; 5(3):1–15. Available from: http://dx.doi.org/10.3389/fninf.2011.00003.

38. Reginold W, Itorralba J, Luedke AC, Fernandez-Ruiz J, Reginold J, Islam O, et al. Tractography at 3T MRI of Corpus Callosum Tracts Crossing White Matter Hyperintensities. AJNR American journal of neuroradiology. 2016 Sep; 37:1617–1622. https://doi.org/10.3174/ajnr.A4788 PMID: 27127001

39. Hofer S, Frahm J. Topography of the human corpus callosum revisited–comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. NeuroImage. 2006 Sep; 32:989–994. https://doi.org/10.1016/j.neuroimage.2006.05.044