

Brain graph networks comparison between different species

Learning from Networks 2022/23 - Final Report

Guglielmin N. 2052422 - Savio A. 2052420 - Zaccarin A. 2052414

January 30, 2022

1 Introduction

Our idea for this project came while looking for graphs to work on. As we checked various animals and also human brain graphs we wondered if it was possible to compare these very different brain structures and verify if the superior intelligence of the human race with respect to every other living being on the planet is at least in part explained by a more interesting brain structure. To perform this comparison, we decided on computing various parameters that we have studied during the course and use them for this task. After a brief discussion on the graphs we have chosen for the project, we will explain in detail what we were able to do and also what went wrong and didn't work.

2 Datasets

The 5 types of brain networks we analyzed, available from public repositories [6], are:

1. Human brain network [1]: unweighted graph network which edges represent fiber tracts that connect one vertex to another (nodes: 177.6K, edges: 15.7M, density: 0.000993727, average degree: 176);
2. Cat brain network [2]: unweighted graph network which edges represent fiber tracts that connect one vertex to another (nodes: 65, edges: 1.1K, density: 0.547596, average degree: 35);
3. Fly-drosophila-medulla brain network [3]: unweighted graph network which edges represent fiber tracts that connect one vertex to another (nodes: 1.8K, edges: 33.5M, density: 0.0211395, average degree: 37);
4. Macaque brain network [4]: unweighted graph network which edges represent fiber tracts that connect one vertex to another (nodes: 242, edges: 4.1K, density: 0.140256, average degree: 33);
5. Mouse brain network [5]: unweighted graph network which edges represent fiber tracts that connect one vertex to another (nodes: 213, edges: 21.7K, density: 0.959075, average degree: 203).

These networks are very different, even more than we expected. Most notably, the brain network regarding the fly is disconnected, with a dense connected component in the middle. The mouse brain network is also very dense, while the cat and the macaque networks are a bit more sparse. Finally, the human brain network is the most complex among these. It is very large in terms of number of nodes and edges, even though it is quite sparse. It is also important to underline that we have chosen only one network for each animal. We do not know how they were obtained and if they were derived from a single animal. If that is the case, we have to assume that every animal of that species has roughly the same brain structure and parameters, which can be a bold assumption: it is known, for example, that every human being has a different brain anatomy, just like fingerprints. Anyhow, as it is obviously impossible to test every brain network of every living being and in order to generalize, this is an assumption we have to make.

3 What went wrong?

As described in the previous reports, we feared that the great dimensions of the chosen brain networks may have posed a problem when computing the various indexes for the comparison. Despite our best efforts, we were not able to compute cliques and motifs (even the basic ones, i.e. triangles) on these networks. In particular, the computation was unbearable in terms of time. We have tried running both locally and on Google Colaboratory algorithms for finding cliques and motifs, even letting the machine run overnight, without success. We have thus decided to not compute these parameters and focus solely on centralities, clustering coefficient, PCA and Z-scores. You will find the code we tried for cliques and graphlets commented in the notebook.

4 Experiments performed

We performed experiments dividing the full analysis in 3 sub-parts:

1. Cat, fly, macaque and mouse brains analysis using exact algorithms (Section 4.1);
2. Human brain analysis using approximate algorithms (Section 4.2);
3. PCA to represent in 2D the data extracted from the brains networks (Section 4.3).

4.1 Computation for cat, fly, macaque and mouse brains

First of all, we loaded the networks and since they are small enough, we drew them and we decided to implement only exact algorithms. Then, for each of these 4 networks, we computed the following indexes:

- Degree centralities;
- Closeness centralities;
- Betweenness centralities.

Since the above are all indexes at node level and we are interested in indexes at graph level, we computed also the global centralities by calculating the corresponding means.

At this point we computed 2 other indexes:

- Clustering coefficients;
- Z-scores.

The Clustering Coefficients make us understand how a graph can be seen from a different point of view with respect to the previous indexes, while Z-scores have been computed to compare the previous obtained results with the same indexes computed on 10 randomly generated graphs, to highlight which of the obtained features are more interesting. We decided to set as 10 the number of random graphs as a good trade-off between little execution time and a non trivial number of graphs, but enough good for us to get the expected value $E[x_F]$ and the standard deviation $\sigma[x_F]$.

At the end, we saved all indexes relative to each specific brain network in a correspondent dictionary for further steps (Section 4.3).

4.2 Estimation for the human brain

As the human brain network is exceptionally large, we had to resort to utilizing estimation algorithms in order to compute the parameters for the comparison. Because of this, we only loaded the network without drawing it. Then, as for the other networks, we computed the same indexes:

- Average Degree centrality;

- Average Closeness centrality;
- Average Betweenness centrality;
- Clustering coefficient;
- Z-scores.

To be consistent with previous analyses, also these Z-scores have been computed on 10 randomly generated graphs. Since the human brain network is very large, these computations last for a long time before ending.

At the end, we saved all indexes relative to this specific brain network in the correspondent dictionary for further steps (Section 4.3).

4.3 PCA for 2D representation

Now we have all features for each network. A general feature vector has the following form: *[average_degree_centrality, average_closeness_centrality, average_betweenness_centrality, global_clustering_coeficient]*.

To compare the features extracted for each network we saved them in feature vectors and then in a data matrix. Each row of this matrix is a feature vector through which we can represent graphs as points in a \mathbb{R}^4 space, with 4 the number of parameters we take into account. After this, we applied the PCA transform to the data matrix. Therefore, using the fact that two similar networks are close in a \mathbb{R}^n space we can conclude their similarity also in functionalities. For this reason, in Section 5.3 we provide a visual concept of what has been evaluated, that is a 2D representation of the networks in the \mathbb{R}^4 space using PCA.

5 Comparison

Before going in depth discussing the results of the experiments we obtained it is important to keep in mind that we are considering only specific cases since networks are a “photography” of a particular animal, and to be able to generalize our analysis we need to use many more networks per species and this is not feasible for us.

Comparison number of neurons and graphs size			
Species	Average number of neurons	Nodes in the graph	Edges in the graph
Cat	7.6×10^8	65	1.1K
Fly	10^5	1.8K	33.5M
Macaque	10^9	242	4.1K
Mouse	7×10^7	213	21.7
Human	8.6×10^{10}	177.6K	15.7M

Furthermore as highlighted in the table above there seems to be no direct correlation between the average number of neurons composing the brain of one specie and the size of the corresponding graph. As we highlighted in the midterm report too the indices we computed take into account the size of the graph too but this doesn’t stop us from asking some questions about the methods used to extract this data, unfortunately the repository where we found the graphs doesn’t provide any additional information. In the following subsections we will discuss the results we obtained but our analysis only applies to these specific cases.

5.1 Results

Indices obtained on the brain networks				
Network Name	Average Degree Centrality	Average Closeness Centrality	Average Betweenness Centrality	Global Clustering Coefficient
Cat	0.35096	0.59771	0.01110	0.08272
Fly	0.00568	0.34908	0.00106	1.86490e-05
Macaque	0.10473	0.45855	0.00507	0.00554
Mouse	0.71937	0.78343	0.00136	0.39697
Human	18.03056	8.96141e-08	2.05915	0.47008

The table above reports the results we obtained as we illustrated in the Performed Experiments section, before seeing the results we were expecting that the human graph would stand out from the others and that the macaque one would maybe have similar values since it is the second one with more neurons so we can expect a similar brain structure. Looking at the results we see that the human graph definitely stands out from the others, especially in the average degree centrality and average betweenness centrality. It is also possible to notice how even if the graphs were very different every other index is fairly in line with each other. This could be very interesting, if it is not something explained by the natural properties of graphs it could be a property common to brain networks. To verify this hypothesis we decided to compute the z-scores for each parameters.

5.2 Z-scores

As discussed before due to limitations in the computing time we were able to only generate a handful of random graphs to compute the z-scores, still some interesting information emerged.

Z-scores of the Indices				
Network Name	Average Degree Centrality	Average Closeness Centrality	Average Betweenness Centrality	Global Clustering Coefficient
Cat	-	-78.12233	486893418529103.44	78.82627
Fly	-0.99999	-0.61015	-1.07693	-0.99903
Macaque	-0.33799	-0.07249	-0.04282	-0.46237
Mouse	1.54089	1.32611	-0.79289	1.88028
Human	-0.24987	-100.99570	-377.10841	36079.66666

Looking at the table above the results of the human graph stand out as interesting except for the average degree centrality which was unexpected since looking just at the values it was much higher than the ones computed in the other graphs. All the values for the fly, macaque and mouse graphs seem to be explained by the natural properties of graphs of that size. But it was very unexpected to see the z-scores on the cat graph, just looking at the values it was very much in line with the others but the z-scores point at those values as very unusual.

5.3 Principal Component Analysis

As last step we wanted to display the obtained results in 2D space to be able to visualize them and see if the differences seen in the data were reflected in this way too (Figure 1). Looking at the image it is possible to see how the human graph is separated from the rest, the cat that presented unusual z-score but uniform values has been grouped together with the macaque and the fly graphs. Unexpectedly the mouse graph has been set apart from everyone else even though both the values of the indices and the z-scores were very close to the macaque and fly graphs.

6 Graphical results

In this section we provide a graphical representation of the networks, the distributions of the Degree Centralities, Closeness Centralities and Betweenness Centralities of the 4 smallest networks.

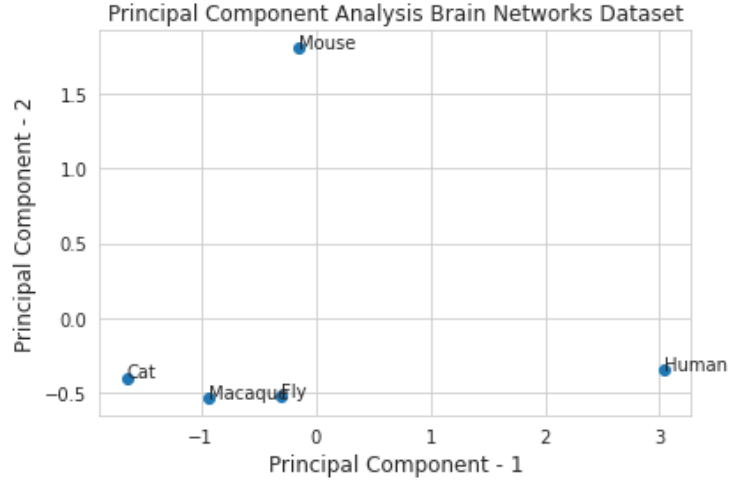


Figure 1: Results of the principal component analysis, this allow us to represent the data on a 2D graph.

Since the human brain network is very large, with the machine available we have not been able to visualize it, so there is no figure for it.

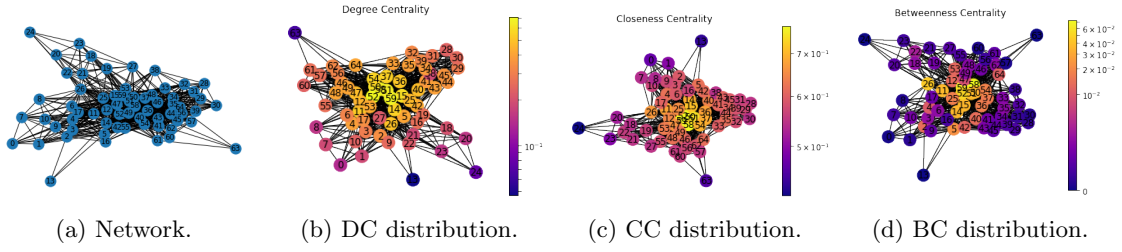


Figure 2: Figures for the cat's brain network.

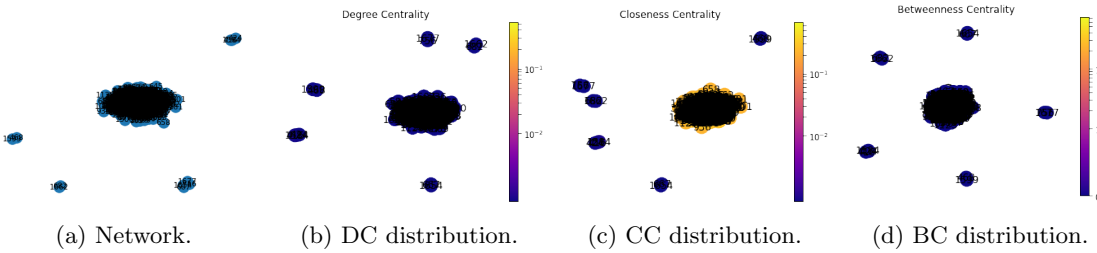


Figure 3: Figures for the fly's brain network.

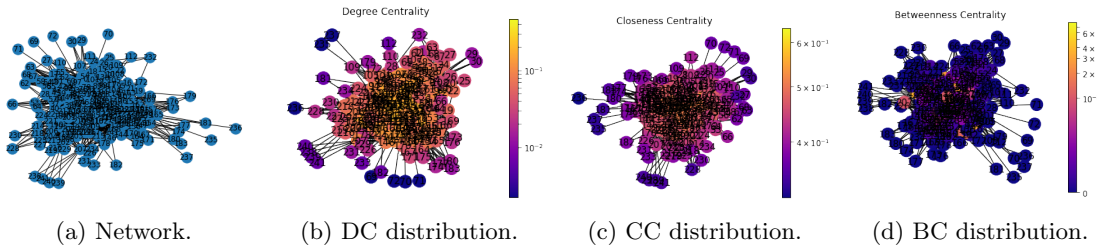


Figure 4: Figures for the macaque's brain network.

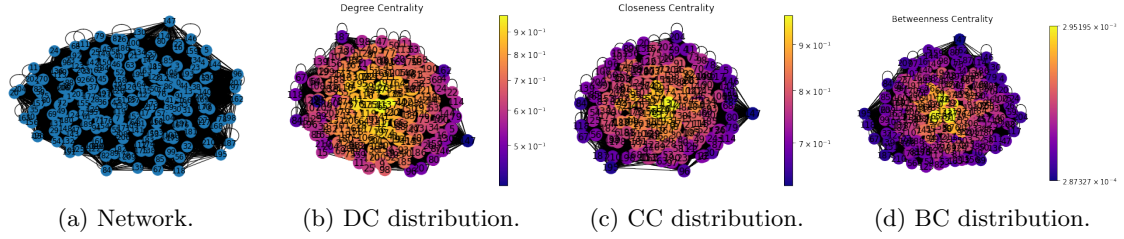


Figure 5: Figures for the mouse's brain network.

7 Conclusion

To conclude our analysis, the goal of this project was to compare various brains and ultimately assess if the human intelligence can be explained at least in part by the human brain superiority. The brain is surely the most complex between the species we analyzed, as expected. In fact, the size alone is orders of magnitude bigger than the others, but on top of that almost all indices computed point to more interesting features belonging to the human brain with respect to the others. It is also quite interesting to denote that the macaque, a species that should be a bit similar to humans (as we share a common ancestor with chimpanzees) has a less complex brain, which is closer to the fly, according to the PCA analysis. Finally, of course our analysis is not enough to justify the human intelligence supremacy over every other species of the animal kingdom, but we believe that it is at least interesting to denote that graph theory concepts and algorithms results agree with biological sciences.

8 Code

We provide our code for this project in the following GitHub repository:

https://github.com/NiccoloGuglielmin/Learning_from_Networks_Project_2223.git .

The code available in this repository is the one we runned exploiting Google Colaboratory machines which are based on a Linux system, platform *Linux* – 5.10.133 + *-x86_64* – *with* – *Ubuntu* – 18.04 – *bionic* with already installed *Python3.7.15* .

In this repository you can find also the link to download the data folder with the brain networks and the instructions to run the code in a local machine.

References

- [1] Katrin Amunts et al. “BigBrain: An Ultrahigh-Resolution 3D Human Brain Model”. In: *Science* 340.6139 (2013), pp. 1472–1475.
- [2] Katrin Amunts et al. “BigBrain: An Ultrahigh-Resolution 3D Human Brain Model”. In: *Science* 340.6139 (2013), pp. 1472–1475. URL: <https://networkrepository.com/bn-cat-mixed-species-brain-1.php>.
- [3] Katrin Amunts et al. “BigBrain: An Ultrahigh-Resolution 3D Human Brain Model”. In: *Science* 340.6139 (2013), pp. 1472–1475. URL: <https://networkrepository.com/bn-fly-drosophila-medulla-1.php>.
- [4] Katrin Amunts et al. “BigBrain: An Ultrahigh-Resolution 3D Human Brain Model”. In: *Science* 340.6139 (2013), pp. 1472–1475. URL: <https://networkrepository.com/bn-macaque-rhesus-brain-1.php>.
- [5] Katrin Amunts et al. “BigBrain: An Ultrahigh-Resolution 3D Human Brain Model”. In: *Science* 340.6139 (2013), pp. 1472–1475. URL: <https://networkrepository.com/bn-mouse-brain-1.php>.
- [6] Ryan A. Rossi and Nesreen K. Ahmed. “The Network Data Repository with Interactive Graph Analytics and Visualization”. In: *AAAI*. 2015. URL: <https://networkrepository.com>.

9 Detailed contribution of each member

We organized our work in the following way:

- we met together to find links to datasets, repositories, papers and examples of already done analysis about general topics. We also decided the motivation for our project and we wrote the project proposal document we submitted;
- we started individually to write some code about simple analysis of networks in Python to understand what was simple and what needed more work to do, in order to build experience and to have some code to start the project, using Python packages NetworkX and NetworKit;
- we met again together to develop a more deep analysis of what has been decided to be proposed for the project in order to adjust main points; then we wrote the project midterm document we submitted;
- we met another time to define the structure of the code in order to assign to each member a precise set of activities to accomplish. The division of the activities and, accordingly, of the code has been done assigning to Andrea the parts for Code Setup, computations of Cliques, Graphlets, Z-scores and the conclusions for the analysis of the smallest 4 networks; Niccolò developed the functions for the drawings of any network, the part for the computation of all centralities for the Cat, Fly, Macaque and Mouse brain networks (Degree, Closeness, Betweenness centralities), the computation of the Clustering Coefficients for these networks; Angelo developed the analysis of the human network and concluded the code implementing the PCA analysis to represent our extracted feature data in a 2D plane. During the development, sometimes we met and some other times we worked alone;
- we met to decide the structure of the final report document, so that each member has been able to write the section about the code he developed. Since we used a shared document to write the report, we all had access to the document in order to help and give hints to other members, if needed. At the end, we met again to revise the report and we submitted this final version.

Since when needed each component of the group also helped other members in both the theoretical part and the coding part and we worked most of the time in presence together, it's very complicated to assess the precise fraction of work done by each member; it is roughly split evenly between all 3 members.