

MHEDAS 2024-2025

---

# A2 - COMMUNITY STRUCTURE OF COMPLEX NETWORKS

## **Authors:**

Ihona Correa de Cabo  
Lloyd Linton Jones

Group A

A decorative graphic consisting of numerous thin, wavy green lines that flow from the bottom left towards the top right, creating a sense of movement and complexity.

# INDEX

1. INTRODUCTION	1
2. NETWORK CHARACTERISTICS	1
3. COMMUNITY DETECTION	5
3.1. Modularity Based Algorithms	5
3.1.1. Greedy Algorithm	5
3.1.2. Louvain Algorithm	5
3.2. Probabilistic Models	6
3.2.1 Stochastic Block Model (SBM)	6
4. ANALYSIS OF DETECTED COMMUNITIES	6
5. CONCLUSION	13

# 1. INTRODUCTION

The aim of this report is to identify and analyse communities in complex networks. In this context, communities are groups of nodes that are more densely connected to each other than to the rest of the network. Detecting them can be useful for understanding the underlying structure and function of complex networks.

For this purpose, a connectome network was chosen from Netzschleuder [1]. The connectome, which can be defined as a comprehensive map of neural connections in the brain, was obtained from MRI datasets of 477 people. Here, nodes are brain regions, and edges are weighted by several factors such as the number of "tracks" that run between two nodes, the fibre length, the fractional anisotropy and the number of occurrences in each of the 477 individuals.

Among the different options from the network catalogue for this project, it has been decided to use the dataset that contained the connectome from all participants (both men and women).

Since the network was a different network from the one in A1, some basic metrics and visualisations are presented in order to gain a deeper understanding of the network before detecting communities. Afterwhich, the different communities of the dataset were detected and analysed.

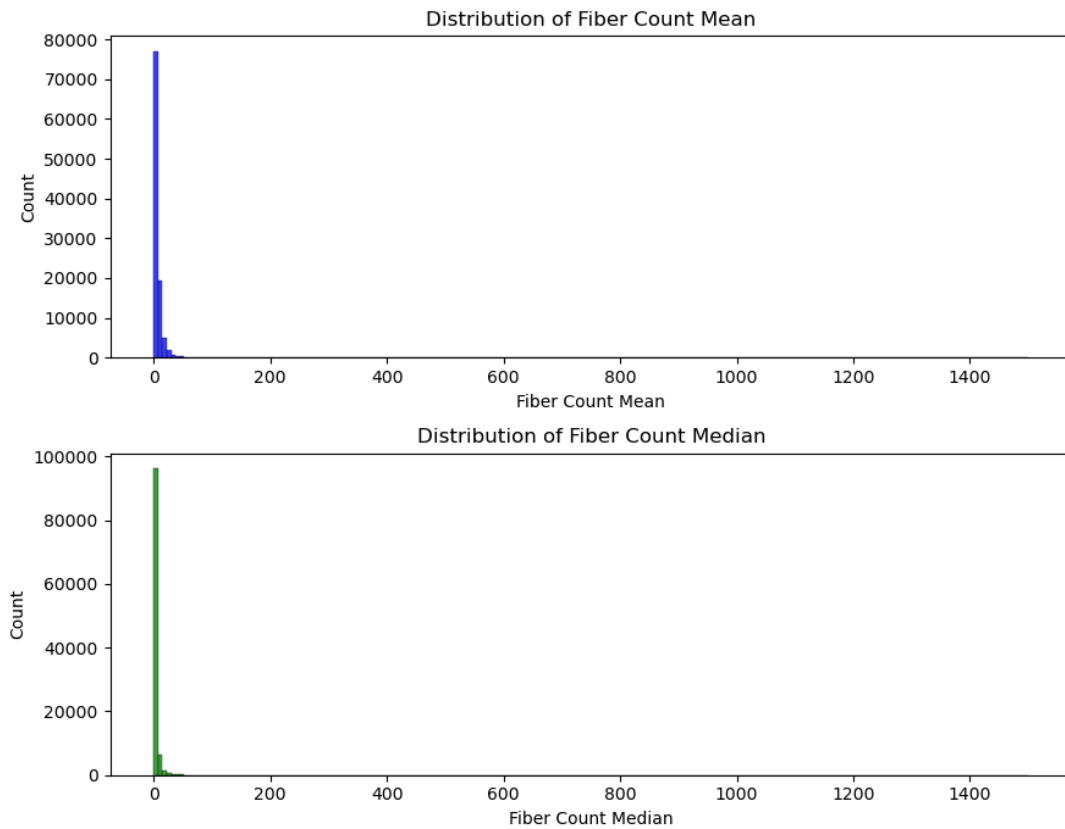
# 2. NETWORK CHARACTERISTICS

The first step was to load the network into the working environment and decide the weights that quantify the interactions between brain regions. Among the possible options, it was decided that the fibre count had more relevance and strength in this context. Fibre count directly quantifies the number of tracts connecting two brain regions, so it can be used to represent how much "flow" can pass between the nodes (brain regions). For instance, a higher number of fibres likely enables more efficient communication between nodes.

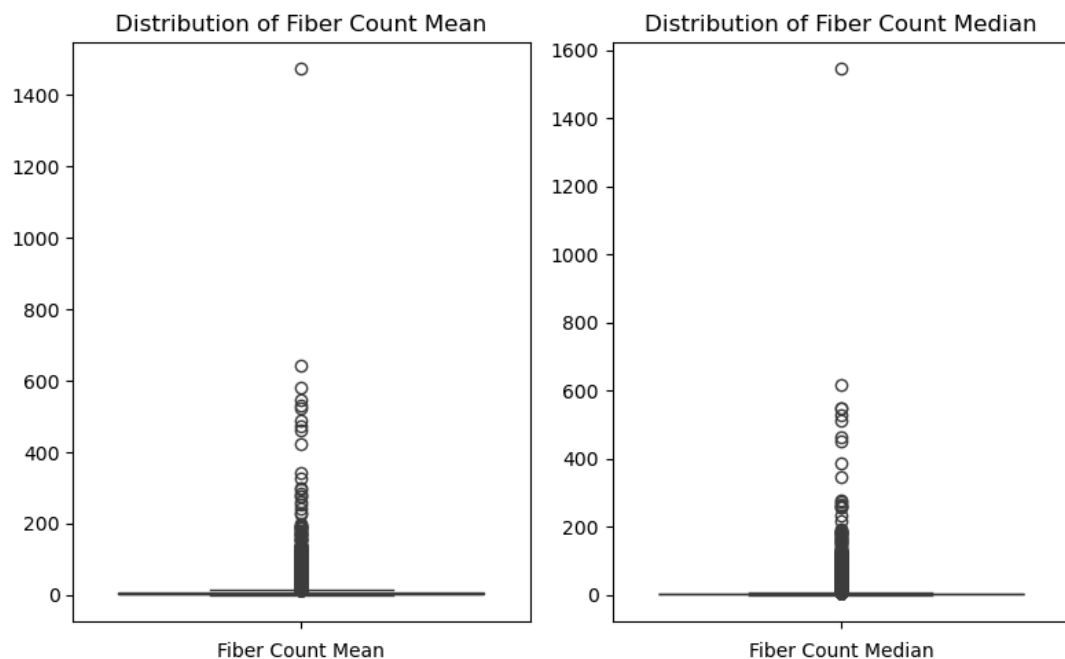
From the network repository, the fibre count mean and median were available. The fibre count mean represents the average number of tracks across the 477 individuals, whereas the other gives the median value of tracks across the participants.

In order to decide between the fibre count mean or median, the distribution of both factors was compared to check for outlier imbalance, ensuring that the metric that best represents that data is used. This comparison is achieved through the histograms in *Figure 1* and boxplots in *Figure 2*.

[1] B. Szalkai et al., *The Budapest Reference Connectome Server v2.0*, *Neuroscience Letters*, 595, 60–62 (2015). DOI: [10.1016/j.neulet.2015.03.071](https://doi.org/10.1016/j.neulet.2015.03.071). Available at: [Budapest Connectome Network - Netzschleuder](https://www.budapest-connectome-network.org/)



**Figure 1. Histogram of fibre count mean and median.**



**Figure 2. Boxplots of fibre count mean and median.**

Both the histograms and boxplots reveal that most of the data is centred at very low values (close to 0). This means that the majority of edges in the brain network have only a small number of connections (or “tracks”) between most brain regions in the dataset. As most

values cluster around the same range, the mean and median will be similar, so either can be used as the weight. Since there are few outliers, fibre counts mean was chosen.

The source, target and weight (fibre count mean) is stored in a data frame for further analysis.

This network is a mono-layered undirected network. Therefore, the connection between nodes (source and target) is symmetric. The source and target columns define the relationship between brain regions, whereas the weight quantifies the strength of those relationships. Moreover, it is a unipartite network, so there is only one single type of node.

Next, an undirected graph was created from the edge list using the networkx library. Then, some basic parameters and metrics of the network were obtained. *Table 1 and 2* contain a summary of these metrics.

Feature	Result
Number of nodes	1015
Number of edges	105293
Average degree	207.48
Radius	2
Diameter	4
Average clustering coefficient	0.7114
Assortativity	0.01351

**Table 1. Basic features of the network.**

Centrality Metric	Result	
	Mean	Median
Betweenness	0.00094	0.00023
Closeness	0.51707	0.51238
Eigenvector	0.02822	0.02717
PageRank	0.000985	0.000942

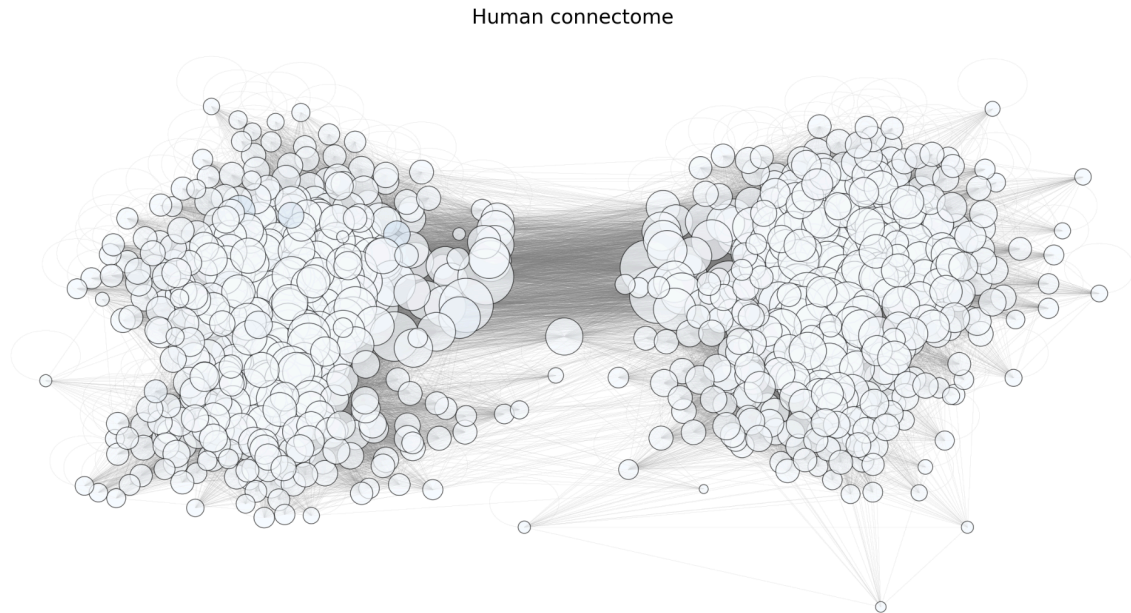
**Table 2. Centrality measures of the network.**

The network consists of 1,015 nodes and 105,293 edges, forming a dense and connected structure. The average degree of the nodes is approximately 207.48, indicating a high level of connectivity between brain regions. The diameter of the network, which represents the longest shortest path, is 4, and the radius, representing the shortest maximum distance from any node, is 2, suggesting a compact network. The assortativity coefficient of 0.0135 suggests a weak positive correlation between the degrees of connected nodes.

Centrality measures provide insights into the roles of individual nodes. The mean and median betweenness centrality, at 0.00094 and 0.00023 respectively, indicate that most nodes lie on relatively few shortest paths. Closeness centrality averages 0.5171, showing

that nodes are, on average, relatively central within the network. Eigenvector centrality (mean: 0.0282, median: 0.0272) and PageRank (mean: 0.000985, median: 0.000942) further support this compactness, emphasising that influence and connectivity are distributed evenly across most nodes.

Moreover, the network was visualised to see any potential groups or clusters before the community detection (*Figure 3*):



**Figure 3. Network visualisation.**

From the above visualisation, it is evident that there are already two distinguishable groups within the given network. These groups might correspond to the right and left hemisphere of the connectome.

### 3. COMMUNITY DETECTION

The aim of this section is to explore and apply several community detection algorithms to identify the communities of this network.

#### 3.1. Modularity Based Algorithms

Modularity is a metric that measures the quality of a division of a network into communities. It evaluates how well the network is partitioned compared to a random network that has the same degree distribution. Modularity ranges between -1 and 1, where highest values indicate stronger community structure and a value of 0 suggests that there is no community structure beyond random chance.

##### 3.1.1. Greedy Algorithm

The greedy algorithm aims to maximise modularity. It starts with each node in its own community and iteratively merges communities to increase modularity until no further improvement is possible. It has been chosen due to its simplicity and the quick detection it provides for relatively small networks like this one. For this network, where there is already an intuition of two larger communities, this easy algorithm gives an approximate solution for modularity optimization.

After applying the greedy algorithm, the modularity score was 0.4585, which means that there is a community structure.

##### 3.1.2. Louvain Algorithm

The Louvain algorithm is a hierarchical clustering algorithm that maximises modularity through the following steps:

1. Initial Assignment: Each node in the network is assigned to its own community.
2. Iterative Improvement: The algorithm iterates through the nodes and evaluates the modularity gain by moving each node to its neighbouring communities.
3. Community Aggregation: After evaluating all nodes, the algorithm aggregates the nodes into their new determined communities.
4. Steps 2 and 3 are repeated until no further improvement in modularity can be achieved.

In the final community structure, nodes in the same community are more densely connected to each other than to nodes in other communities.

Whereas the greedy algorithm focuses on a global approach, considering modularity changes across the whole network at each step, the Louvain uses local modularity optimization.

The Louvain algorithm has an advantage of being fast and scalable and can output hierarchical communities (for example hemispheres and subregions in those hemispheres, if that was the case). This algorithm is considered more robust and often achieves higher modularity scores, like in this case.

The modularity of the inferred partition with the Louvain algorithm was 0.4948, slightly higher than with the greedy algorithm, suggesting a better community structure.

## 3.2. Probabilistic Models

### 3.2.1 Stochastic Block Model (SBM)

The Stochastic Block Model (SBM) is a generative probabilistic model used to represent networks with a predefined community or block structure. It assumes that nodes can be grouped into distinct blocks, and the probability of an edge between two nodes depends only on the blocks to which they belong. SBM is particularly useful for studying networks with underlying group-level connectivity patterns, like in the case of the network analysed since it is known there are two hemispheres which have been clearly represented by the prior visualisations and analysis.

To generate a network using the SBM, nodes are first assigned to blocks according to some predefined rule or probability distribution. A connectivity matrix is then defined, where each entry specifies the probability of an edge between nodes in two given blocks. Edges are formed independently according to these probabilities, resulting in a network with specified interblock and intrablock connectivity. By varying the block sizes and connection probabilities, networks can be created with differing structural properties resembling the known real-world community structures.

The relationship between SBM and modularity lies in their complementary roles in community detection and evaluation. While SBM focuses on generating networks with predefined community structures, modularity is a metric used to evaluate the quality of community partitions by comparing the density of internal connections within communities to what would be expected in a randomised network. Networks generated using SBM can be analysed using modularity-based approaches to assess how well a given partition reflects the true community structure. Conversely, modularity optimization algorithms, such as Louvain or Greedy, can help uncover communities in empirical networks, which can then be compared to the ground truth provided by SBM-generated data. In this way, SBM provides a controlled framework for testing and validating community detection algorithms, while modularity offers a means to evaluate their performance or assess network structures post generation.

## 4. ANALYSIS OF DETECTED COMMUNITIES

This section shows visualisations of the community structures that were detected by the aforementioned algorithms as well as the analysis of these communities. These detected communities are then compared to the reference labelling that is provided in the network repository.

The metadata for the nodes contains several columns that could potentially be used for reference labelling, but the hemisphere column was chosen because it is more evenly distributed and it aligns better with the network structure during the initial visualisation.



dn_hemisphere		dn_region	
label	count	label	count
right	508	cortical	1000
left	507	subcortical	15

**Table 3. Label counts based on reference factors**

A new data frame is formed using this metadata with index being the node and the hemisphere each node is assigned to which can be either right or left, as it is seen in *Table 3*).

Then, for each of the three algorithms described above, a data frame is created that contains the nodes and the detected communities assigned to each. These data frames reveal that the greedy algorithm detected only two communities or classes (*Table 4*, left), whereas the Louvain detected four (*Table 4*, right), and finally it is known that the SBM had two communities as analysis was conducted assuming there were two blocks.

Greedy Algorithm Communities		Louvian Algorithm Communities	
Community	Count	Community	Count
0	513	0	309
1	502	1	199
		2	312
		3	195

**Table 4. Algorithm community classification node counts**

Merging each of the algorithm data frames with the metadata by using the common index (node), obtains a merged dataset that contains the index (node), the community assigned to that node by the specific algorithm and the label from the metadata. Finally each merged data frame is grouped by community then reference label to count how many nodes are assigned to each category grouping. This allows misclassified nodes to be identified.

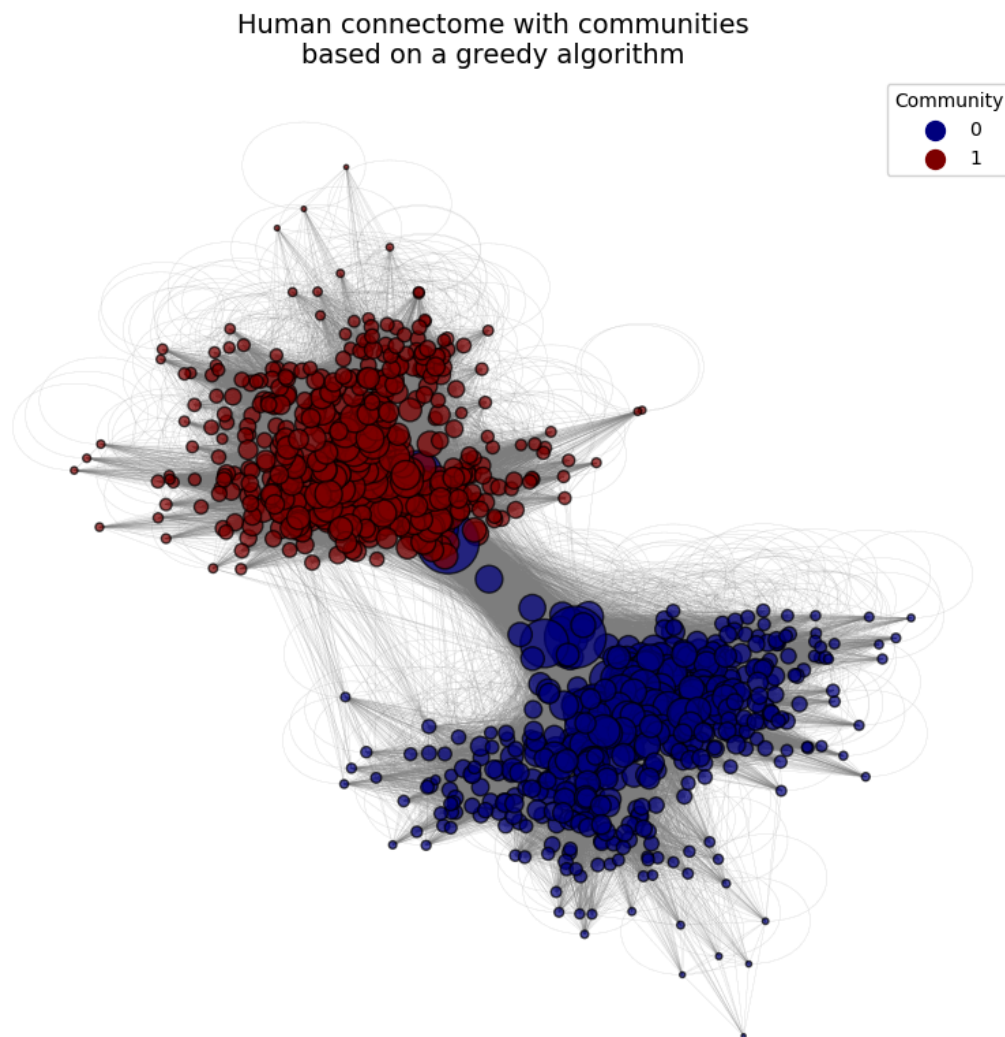
For the **greedy algorithm**, it is possible to see that there are 5 nodes that were misclassified (they were assigned to the right class “0” but actually belong to the left “1”, shown in *Table 5*.)

Community	Hemisphere	Count
0	left	5
0	right	508
1	left	502

**Table 5. Greedy algorithm community classifications by reference label**

After visualising the communities detected by the algorithm (*Figure 4*), it is clear that there are some misclassified nodes (blue) that should be assigned to the left hemisphere (red).

Nevertheless, the greedy algorithm correctly detects the two reference classes that correspond to the left and right brain hemisphere.



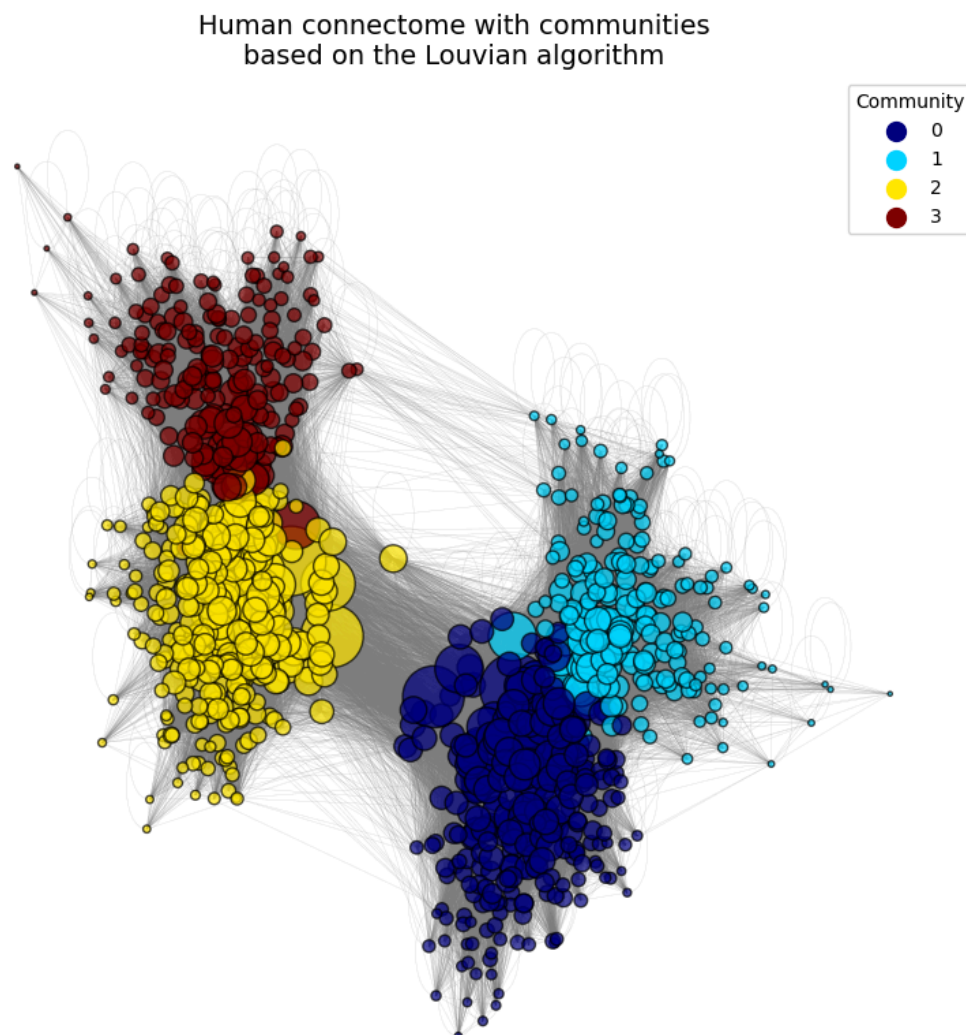
**Figure 4.** Visualisation of the communities detected by the greedy algorithm.

On the other hand, the **Louvain algorithm**, assigned nodes according to the following community structure, shown in *Table 6*.

Community	Hemisphere	Count
0	right	309
1	right	199
2	left	312
3	left	195

**Table 6.** Louvian algorithm community classifications by reference label

It is evident that the Louvian algorithm detected four communities despite there being only two reference labels. The following image, *Figure 5*, shows a visualisation of these detected communities.



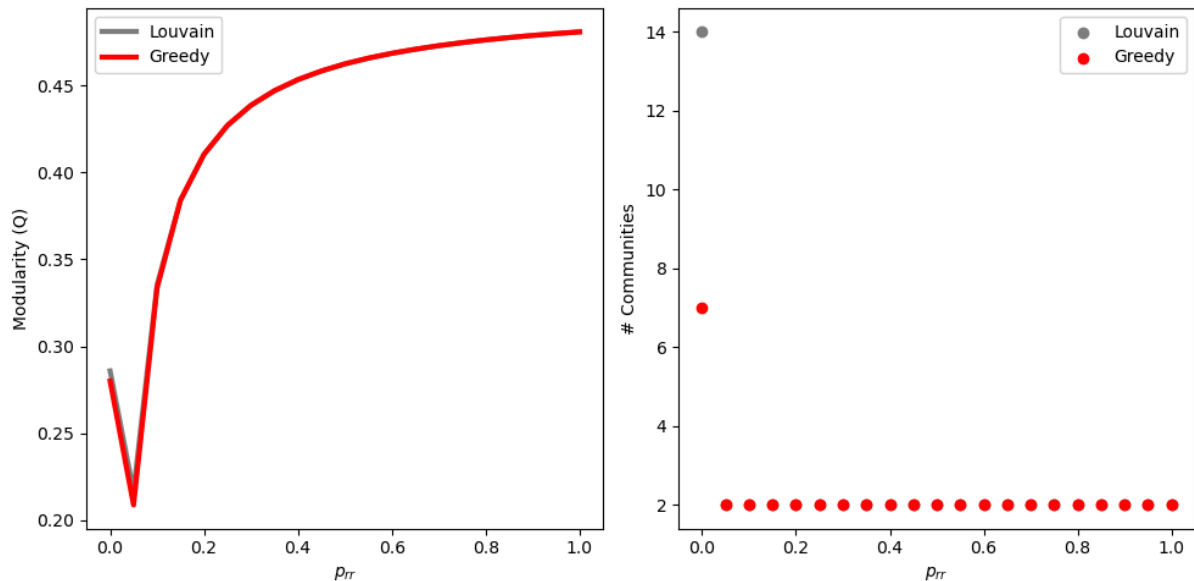
**Figure 5. Visualisation of the communities detected by the Louvain algorithm.**

This may suggest that the algorithms are capturing structures at a different granularity. The result from the greedy algorithm detects the largest-scale structure in the network, which in this case corresponds to the division of the brain into hemispheres. The Louvain algorithm, on the other hand, detected four communities, splitting hemispheres into sub-smaller communities. Although we do not have that level of detail in the metadata, the two hemispheres may have been subdivided into structural subregions (for example frontal and occipital lobes within each hemisphere). Therefore, it can be seen how Louvain's hierarchical nature allows us to detect finer structures within the larger communities.

As described before, the stochastic block model (SBM) provides a probabilistic approach to detect and understand community structures within networks. The previous analyses using the greedy and Louvain algorithms identified two communities broadly corresponding to

brain hemispheres but with some variations in the detected structures. While the greedy algorithm grouped nodes into two major communities matching hemisphere divisions from the metadata, the Louvain algorithm identified four communities, potentially reflecting subregions within hemispheres. The SBM complements these methods by introducing predefined blocks that can represent these communities probabilistically. This allows the validation of prior findings and the exploration of connectivity patterns.

The following graphs, shown in *Figure 6*, represent the changes in modularity and number of communities as the probability of finding connections inside a given block is varied.



**Figure 6. Changes in modularity and number of communities with  $P_{rr}$**

*Figure 6* illustrates the relationship between intra-block connection probability ( $P_{rr}$ ) and two key metrics: modularity and the number of detected communities. As  $P_{rr}$  increases, modularity initially rises, indicating stronger and more cohesive community structures. Beyond a threshold, modularity plateaus, reflecting a stable community division. Similarly, the number of communities decreases with increasing  $P_{rr}$ , as the network transitions from fragmented blocks to more consolidated structures. This behaviour demonstrates that higher intra-block probabilities lead to fewer, well-defined communities.

*Table 7* summarises the SBM community classifications, revealing a near-perfect alignment between detected communities and the reference hemispheric labels. Most nodes are correctly classified into their respective communities, with a single misclassified node. This accuracy highlights the robustness of SBM in reflecting known structural properties of the connectome.

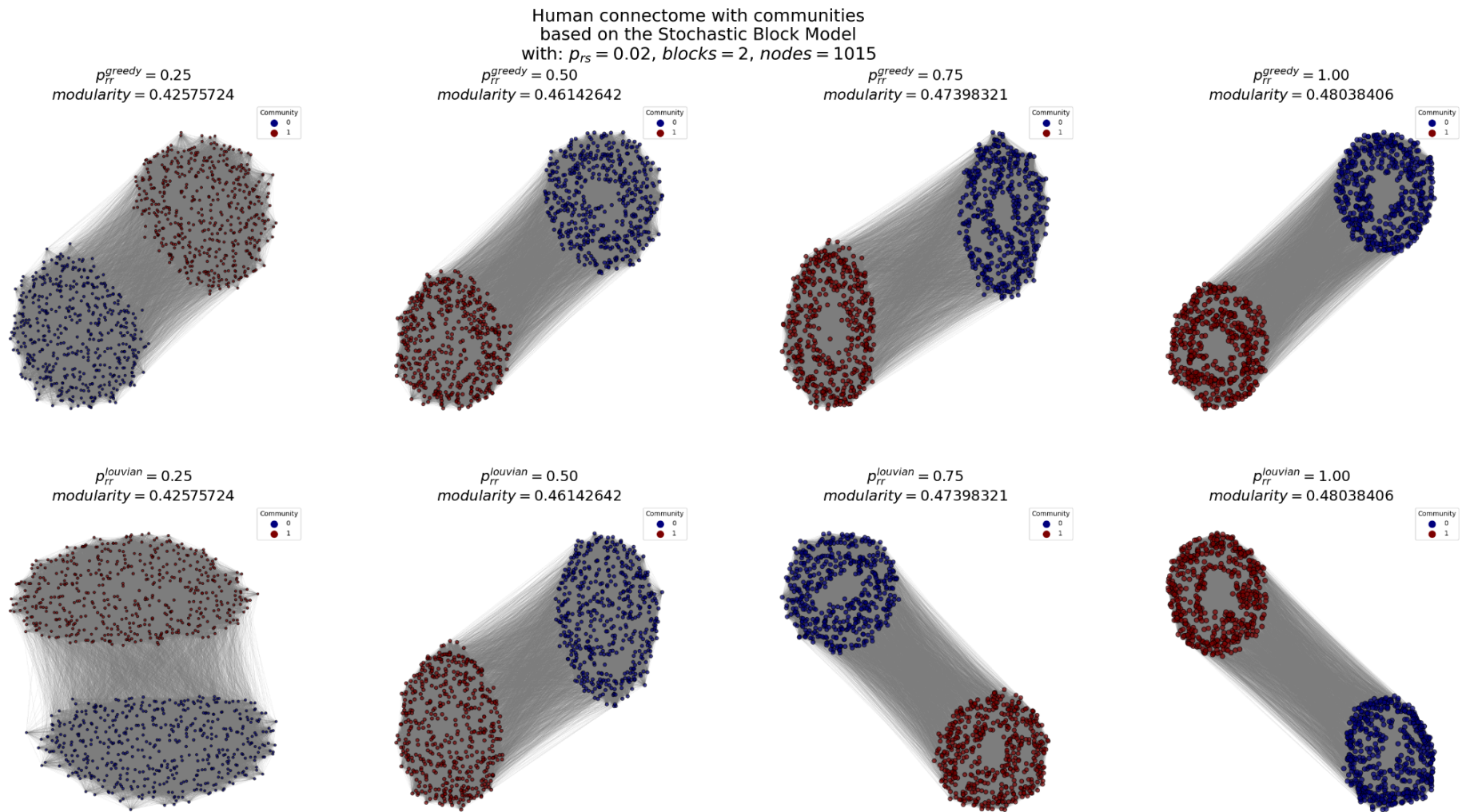
Community	Hemisphere	Count
0	right	507
1	left	506
1	right	1

**Table 7. SBM community classifications by reference label**

*Figure 7* visualises the communities detected by the SBM algorithm, illustrating the clear separation of nodes into two dominant communities corresponding to the left and right brain hemispheres. The uniformity in node classification, where almost all nodes are accurately assigned to their respective hemispheres, reflects the algorithm's ability to capture the inherent modular structure of the network. The minimal overlap or misclassification observed, with only one node placed incorrectly, further highlights the cohesiveness of the detected communities.

This visualisation aligns with the trends observed in *Figure 6*, where modularity ( $Q$ ) increases with higher intra-block connection probabilities ( $P_{rr}$ ) and stabilises as  $P_{rr}$  approaches 1. For example, at  $P_{rr} = 0.25$ , the modularity for the SBM communities is 0.4258. As  $P_{rr}$  increases to 0.75, modularity improves to 0.4740, and finally reaches 0.4803 at  $P_{rr} = 1.0$ . These values indicate that the SBM algorithm is better able to detect cohesive communities as intra-block connectivity strengthens.

In *Figure 7*, the visualisation of SBM communities at  $P_{rr} = 1.0$  demonstrates this alignment. The two detected communities map cleanly onto the hemispheric divisions with minimal misclassification, validating the inherent modularity of the connectome. These results highlight the SBM model's strength in accurately reflecting predefined block structures and provide a useful benchmark for evaluating community detection algorithms. Louvain's ability to detect hierarchical subregions and the greedy algorithm's focus on larger-scale modularity further complement this probabilistic framework. Together, these analyses confirm the modular organisation of the human connectome, with SBM providing clarity on how varying  $P_{rr}$  influences community structure detection.



**Figure 7.** Visualization of the communities detected by the SBM algorithm.

## 5. CONCLUSION

The analysis of the human connectome network successfully demonstrated the modular organisation of the brain into distinct communities, corresponding to the left and right hemispheres. Using multiple algorithms — greedy, Louvain, and the stochastic block model (SBM) — provided complementary insights into community structure detection. The greedy algorithm effectively captured the larger-scale modularity of the network, identifying two primary communities. In contrast, the Louvain algorithm revealed finer-grained substructures, likely corresponding to subregions within hemispheres, achieving consistently higher modularity values.

The SBM provided a probabilistic approach to validate these findings, with results closely aligned to the expected hemisphere divisions. As intra-block connection probability ( $P_{rr}$ ) increased, the modularity also rose, stabilising at a high value (0.4803 at  $P_{rr} = 1.0$ ), indicating well-defined communities. Visualisations of the SBM at different  $P_{rr}$  values further illustrated how connectivity strength affects community detection, complementing the results of modularity-based algorithms.

Overall, this analysis shows the importance of using diverse methods for community detection in complex networks. Each algorithm contributed unique perspectives, from modularity to hierarchical structures, providing a holistic understanding of the connectome's modular nature.

These findings can be expanded on by exploring the interplay between structural and functional connectivity in the brain by leveraging these and similar community detection methods using the electrical connectivity factor in the metadata to uncover further insights into neural organisation.