NETWORK ANALYSIS

Szymon Zɪɴᴋᴏᴡɪᴄᴢ
Matricola number: 5181814

May 1, 2024

University of Gᴇɴᴏᴠᴀ
Instructor: Professor Mᴀʀɪɴᴀ Rɪʙᴀᴜᴅᴏ

Abstract

This report presents an exhaustive network analysis of the "mouse-kasthuri-graph-v4" dataset, a brain network derived from the somatosensory cortex of a mouse. Utilizing Python libraries, this study explores the structural and functional connectivity through various network metrics to assess the graph's resemblance to theoretical network models, such as random and power-law networks. The dataset comprises 1,029 nodes and 1,559 edges, exhibiting a sparse connectivity with an average degree of approximately 3.03. Initial findings suggest the absence of triangles, indicating a lack of local clustering typical to brain networks. Through detailed computations including degree distribution, betweenness centrality, assortativity, and community detection, this analysis characterizes the graph's topology, identifies crucial nodes, and evaluates network robustness and efficiency. Comparisons with random graph models highlight significant deviations, suggesting a structured yet not densely connected network, typical of many biological systems. The report aims to contribute to the understanding of neurological structures and potentially aid in understanding neurobiological processes and disorders.

Contents

1 Introduction

1.1 Methods

In this study, it was performed an in-depth analysis of the "**mouse-kasthuri-graph-v4**" dataset to uncover the underlying structural and functional properties of the brain network. The following types of network analyses will be conducted:

- **Degree Distribution Analysis:** We calculate the degree of each node to understand the distribution pattern and compare it with random and scale-free networks.

- **Centrality Measures:** This includes:
  - Betweenness centrality to identify nodes that serve as bridges within the network.
  - Closeness centrality to find out how quickly information can spread from a given node to others in the network.
  - Eigenvector centrality to identify influential nodes in the network.

- **Community Detection:** Using algorithms like the Louvain method and Greedy Modularity Optimization to identify clusters or communities within the network.

- **Assortativity Coefficient:** We calculate the assortativity to examine the tendency of nodes to connect with others that are similar in terms of degree.

- **Network Density and Sparsity:** Analyzing the overall density of the network to understand the fraction of potential connections that are actual connections.

- **Path Length Analysis:**
  - Calculation of average shortest path length to understand the efficiency of information transport across the network. However the graph is not fully connected, so average path length for largest connected component was calculated.
  - Network diameter to determine the longest of all the calculated shortest paths in the network.

- **Clustering Coefficient:** To measure the degree to which nodes in the network tend to cluster together.

- **Comparison with Random Network Models:** Comparing observed network metrics with those expected in random networks to contextualize the findings within established network theories.

- **Connectivity Analysis:** Evaluating the robustness of the network by analyzing its connectivity and the size of its largest connected component.

The above analyses will be performed using Python libraries such as NetworkX, supplemented by visualizations created in Gephi to illustrate the network's characteristics visually.

1.2   Dataset Description

The dataset analyzed in this report, referred to as "mouse-kasthuri-graph-v4," originates from the somatosensory cortex of a mouse brain and was made available through the Neurodata's Kasthuri project. This dataset is particularly notable for its application in the study of complex neural networks within the brain.

1.2.1   Source and Collection

The "mouse-kasthuri-graph-v4" is part of a larger repository aimed at providing high-resolution electron microscopy data to the scientific community. The data was collected and processed by the Kasthuri lab and downloaded from networkrepository.com.

1.2.2   Characteristics

The dataset features a network graph with the following properties:
- **Number of Nodes:** 1,029, representing various cellular components within the brain.

- **Number of Edges:** 1,559, indicative of the synaptic connections between these components.

- **Average Degree:** Approximately 3.0301, reflecting the average number of connections per node.

- **Network Type:** Unweighted and undirected, suitable for fundamental network analysis.

The data enables exploration of neural connectivity and serves as a fundamental tool for understanding complex biological network patterns, reflecting both the functional and structural properties of the brain network.

2 Analysis Findings

This section of the report presents the key findings from the analysis of the "mouse-kasthuri-graph-v4" dataset. The data was examined using various network metrics to reveal insights into the topology, efficiency, and characteristics of the brain network. The findings are organized into several subsections based on the type of analysis conducted.

2.1 General Characteristics

The network consists of a total of 1,029 nodes and 1,559 edges. These metrics indicate the scale of the network and the extent of connections among the nodes.

- **Number of Nodes:** 1,029

- **Number of Edges:** 1,559

- **Average Degree:** 3.03

2.2 Degree Distribution

The degree distribution of the network was analyzed to assess the connectivity patterns of the nodes. The distribution shows a skewed pattern, suggesting that while most nodes have a low degree of connectivity, a few nodes act as hubs with significantly higher connections. This pattern is inconsistent with random networks, which typically exhibit a Poisson degree distribution, and is more characteristic of scale-free networks.

2.3 Degree Metrics

The degree of a node is a fundamental measure in network analysis, indicative of the node's connectivity. In this network:

- **Maximum Degree:** 123, showing the highest connectivity for a single node.

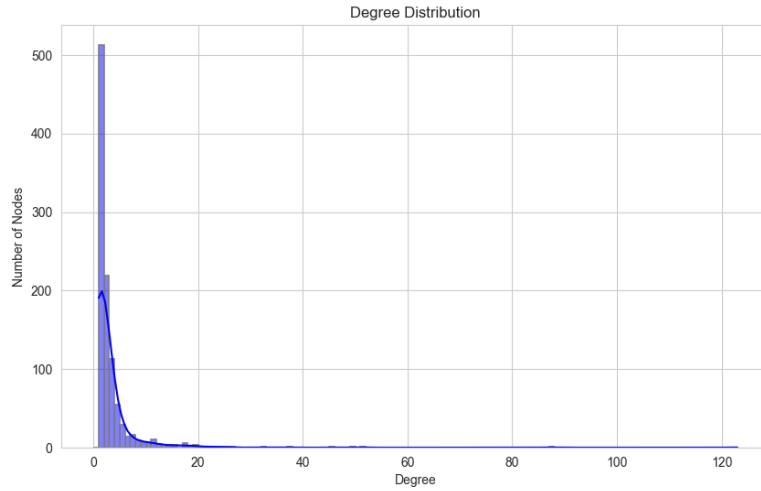- **Minimum Degree:** 1, indicating the lowest connectivity.

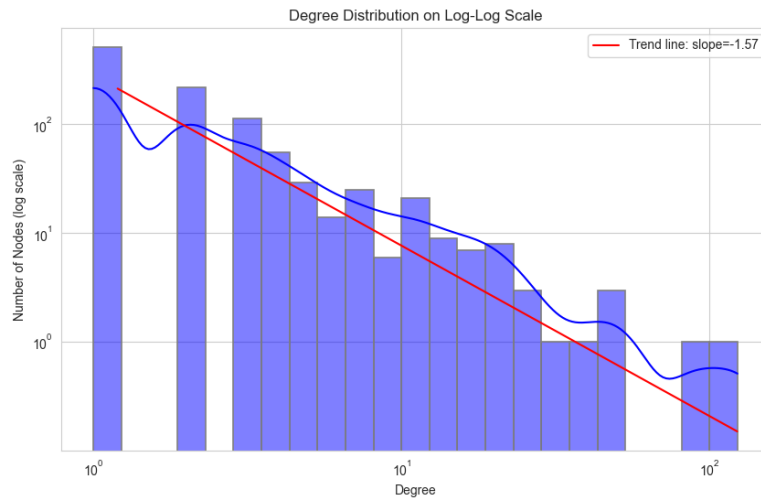Figure 1: Degree distribution of the "mouse-kasthuri-graph-v4" dataset.

Figure 2: Degree distribution of the "mouse-kasthuri-graph-v4" dataset in log scale.

- **Natural Logarithm of Node Count (**$\ln N$**):** Approximately 6.9363, which helps in understanding the logarithmic scale of connectivity.

2.4   Triangle Count

The number of triangles in the network is a measure of local cohesiveness. For this dataset:

- **Number of Triangles:** 0, indicating no closed triplets of connected nodes. This absence of triangles suggests a lack of clustering typically seen in random networks. This is atypical for small-world networks, which generally have high clustering.

2.5   Centrality Measures

Centrality measures were computed to identify critical nodes within the network. Key findings include:

- **Betweenness Centrality:** Nodes with high betweenness centrality were identified, indicating their role as significant points of control and flow within the network. These nodes are critical for the transfer of information across different parts of the brain.

- **Closeness Centrality:** Nodes with high closeness centrality have shorter paths to all other nodes in the network, suggesting their importance in efficient communication.

- **Eigenvector Centrality:** Some nodes showed high eigenvector centrality, which highlights their influence over the network by being connected to other highly connected nodes.

2.6   Centrality Analysis

Centrality metrics are crucial for understanding the influence of individual nodes within a network. This analysis focuses on four key centrality measures: degree, betweenness, closeness, and eigenvector centrality.

2.6.1   Degree Centrality

Degree centrality measures the number of direct connections a node has. The node with the highest degree centrality in our network is node 6, which indicates that it has the most direct connections to other nodes, making it a critical point of connectivity. The top 10 nodes by degree centrality are:

Node	Centrality Score
6	0.1196
83	0.0846
92	0.0496
0	0.0477
35	0.0438
218	0.0360
14	0.0311
66	0.0253
37	0.0243
718	0.0233

Table 1: Top nodes by degree centrality

2.6.2   Betweenness Centrality

Betweenness centrality identifies nodes that serve as bridges along the shortest paths between other nodes. Node 6 also exhibits the highest betweenness centrality, underscoring its role in facilitating communication across the network. The top nodes by betweenness centrality are:

Node	Betweenness Score
6	0.3047
83	0.1943
92	0.1078
0	0.0931
35	0.0738
218	0.0625
14	0.0491
37	0.0382
402	0.0377
111	0.0363

Table 2: Top nodes by betweenness centrality

2.6.3   Closeness Centrality

Closeness centrality measures how quickly a node can access all other nodes in the network. Again, node 6 ranks highest, indicating it has the shortest paths to all other nodes, which enhances its communication speed. Top nodes by closeness centrality include:

Node	Closeness Score
6	0.3126
83	0.2964
92	0.2774
0	0.2766
147	0.2736
561	0.2713
210	0.2697
111	0.2676
496	0.2651
218	0.2641

Table 3: Top nodes by closeness centrality

2.6.4   Eigenvector Centrality

Eigenvector centrality considers not only the quantity of direct connections a node has but also the quality, as determined by the centrality of its neighbors. Node 6 dominates this measure as well, suggesting it is connected to many highly connected neighbors. The top performers are:

Node	Eigenvector Score
6	0.6212
83	0.2051
92	0.1472
0	0.1124
210	0.0906
147	0.0884
561	0.0866
496	0.0821
112	0.0788
378	0.0777

Table 4: Top nodes by eigenvector centrality

This analysis highlights the significant role of node 6 in this network, with its exceptional centrality across all measures pointing to its central role in network dynamics.

2.7   Community Detection

Community detection algorithms were applied to identify modular structures within the network. The Louvain method revealed several densely interconnected communities, suggesting functional compartmentalization within the brain. These communities could correspond to groups of neurons that co-activate during specific types of neural processing.
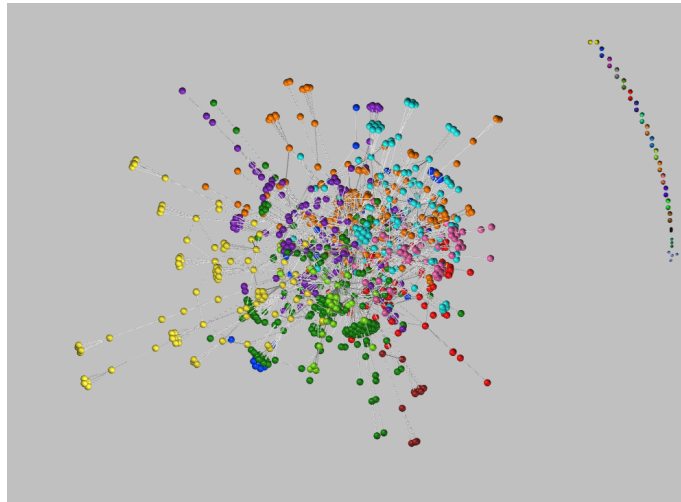

Figure 3: Community structure of the "mouse-kasthuri-graph-v4" dataset identified using the Louvain method.

- **Number of communities: 40-41**, suggests that the network is composed of several distinct modules or clusters. This modular structure is typical in brain networks, where different modules often correspond to different functional areas or tasks.

- **Modularity: 0.645**, suggests a strong division of the network into communities, with dense connections within communities and sparse connections between them. In practical terms, it suggests that the rat brain network has significant clustering of nodes (such as neuronal clusters or brain regions) where there are more interactions within the same cluster and relatively fewer interactions between different clusters.

- **Coverage : 0.71**, high coverage suggests that the community divisions in the network effectively capture a large amount of the interaction or relationships that occur across the network.

- **Performance : 0.95**, high performance score indicates that there are many links within communities and few between them, suggesting that the communities are well-separated and densely connected internally.

2.8 Network Density and Assortativity

The analysis also covered network density and assortativity:

- **Network Density:** 0.0029 The network exhibits a low density, which is typical for large biological networks where not all nodes are directly connected meaning that the network is **sparse**.

- **Assortativity:** -0.2363, The network showed a negative assortativity coefficient, disassortative mixing. This type of mixing happens when nodes in a network tend to connect with other nodes that are unlike them.

2.9 Path Analysis

The network's path analysis indicated:

- **Average Shortest Path Length:** 4.9136 , The average path length across the network is relatively short, facilitating quick transmission of signals across the network. This is relatively short given the total number of nodes **1029** and diameter of the network **12**. This suggests a small-world characteristic since the path length is relatively low compared to the network size.

- **Network Diameter(maximum separation between any two nodes):** **12** , relatively larger diameter suggests that the network exhibits small-world properties.
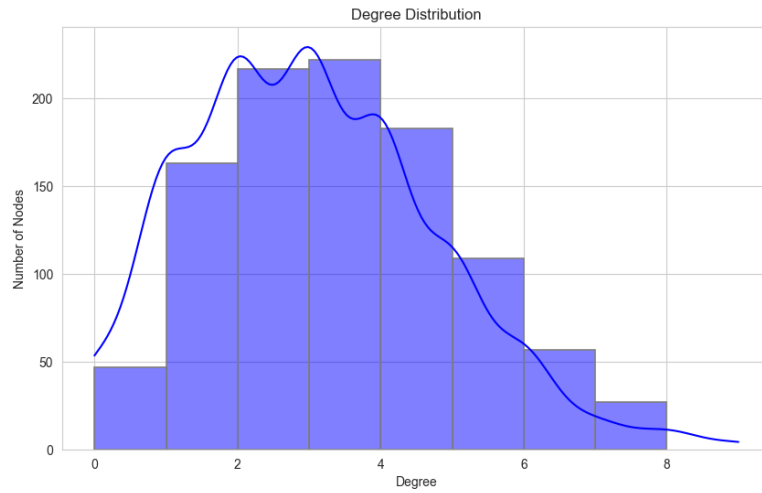
3 Comparison with Theoretical Models

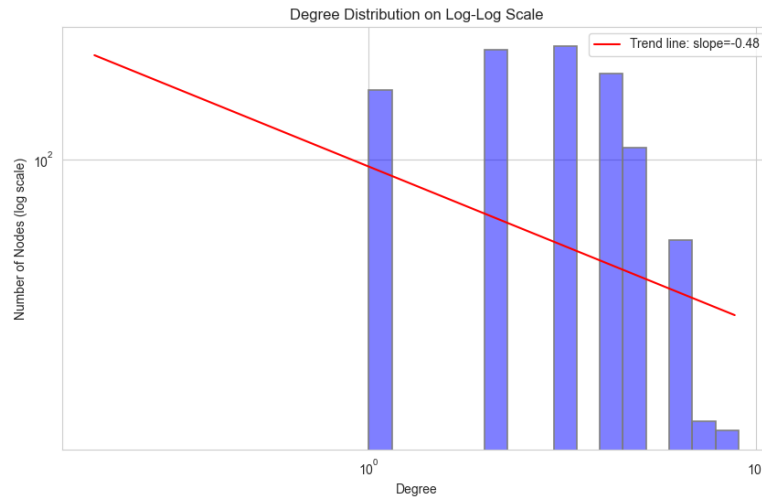Figure 4: Degree distribution of a random network.

Figure 5: Degree distribution of a random network in log scale.

In comparing the structure of the network to theoretical models, we can employ following metrics:

- **Clustering Coefficient:** The average clustering coefficient of our network is **0.0**, substantially lower than that of a comparable random network, which is 0.0015. This significant disparity suggests that the network's structure is decidedly non-random and follows a different organizational pattern. The absence of triangles in our network further substantiates this observation, as random networks typically exhibit a higher clustering coefficient due to the probabilistic formation of triangles. which could be attributed to specific biological functions or developmental patterns in the somatosensory cortex.

- **Degree Distribution:** The random network follows a poisson distribution, while the mouse brain network exhibits a scale-free distribution. This difference in degree distribution indicates that the network is not random but rather follows a power-law distribution, which is typical of many biological networks.

- **Path Length and Connectivity:** The average clustering coefficient of the largest connected component (LCC) in the random network is 0.0049, and the average clustering coefficient of the entire random network is 0.0046. These values are typically low in random networks, as nodes are less likely to form tightly-knit groups or clusters.

  The number of connected components in the random network is 53 while the mouse brain network has 20.

The markedly lower clustering in our network indicates a deliberate avoidance of closed triplets, which could reflect specific functional or evolutionary adaptations in the structural organization of the mouse brain.

4 Conclusion

The structural analysis of the mouse brain network reveals distinctly non-random characteristics when compared to theoretical random network models. The average clustering coefficient of the mouse brain network is **0.0**, significantly lower than that observed in comparable random networks, which have an average clustering coefficient of approximately 0.0016. Additionally, the original network structure deviates from randomness in terms of connectivity, as evidenced by the number of connected components— **20** in the mouse brain network compared to **53** in a random network of similar size. These findings suggest a tailored structural configuration in the mouse brain network that likely supports specific neurological functions or efficiency. The absence of clustering and reduced number of components may reflect evolutionary adaptations for optimized signal processing and energy conservation. Such unique structural features highlight the specialized nature of biological networks, underscoring their deviation from simple random connectivity and pointing towards a more complex, functionally driven architecture.