Cosmic Web and Brain Neural Network Comparison

Roya Joulaei Vijouyeh,[1] Mojtaba Roshana,[2] and Arsalan Namvar[3]

[1]roya.joulaeivijouyeh@studenti.unipd.it
[2]mojtaba.roshana@studenti.unipd.it
[3]arsalan.namvar@studenti.unipd.it

It has been hypothesized that neural systems show topological similarities with the cosmic web, indicating common, or even universal, mechanisms or laws governing the formation of networks across scales, from micrometers to megaparsecs. However, the statistical strength of such similarity is not fully verified. In this project, we first process images from the human cortex, cerebellum, and cosmic web to identify nodes. We sample links according to node distances from an exponential function with exponents r_0 suggested by previous work to obtain ensembles of network realizations for each case. Then we compare them using various metrics, from shortest paths to spectral entropy. Our results indicate that the systems are rather distinguishable.

I. INTRODUCTION

Real-world networks, across a wide range of disciplines, from sociological and biological to man-made, display common topological properties, often referred to as universality in physics— for instance, heterogeneous degree distribution, small-worldness, high modularity, and assortativity. An example of interest is the similarity of the human connectome (brain network) and the cosmic web of galaxies reported by Vazza et al. [1].

Although understanding the commonalities is highly important for identifying general principles that underly the structure and function of complex systems, it might be even more important to characterize, classify and distinguish networks based on their macroscopic features. Therefore, in this report, we take a critical view of the aforementioned study and perform a systematic analysis of the two systems.

Firstly we process images of the galaxy, human cortex, and cerebellum to identify nodes. Secondly, we stochastically sample connections from a probability distribution that exponentially decays with node-node Euclidean distance, $p \propto e^{-r/r_0}$, where the characteristic distance r_0 that is given in ref.[1]. This stochasticity allows for generating ensembles of networks for any set of spatially distributed nodes instead of single shots adjacencies used in ref.[1] that are prone to thresholding error— i.e., $p = 1$ for $r < r_0$ and $p = 0$ otherwise. To compare network ensembles of 30 realizations for each system, we use several classical metrics, including average shortest path, diameter, clustering coefficient, and modularity. Also, we perform spectral entropy analysis across scales.

In contrast with the previous work [1], our results demonstrate that the three systems are clearly distinguishable and even widely different according to certain metrics.
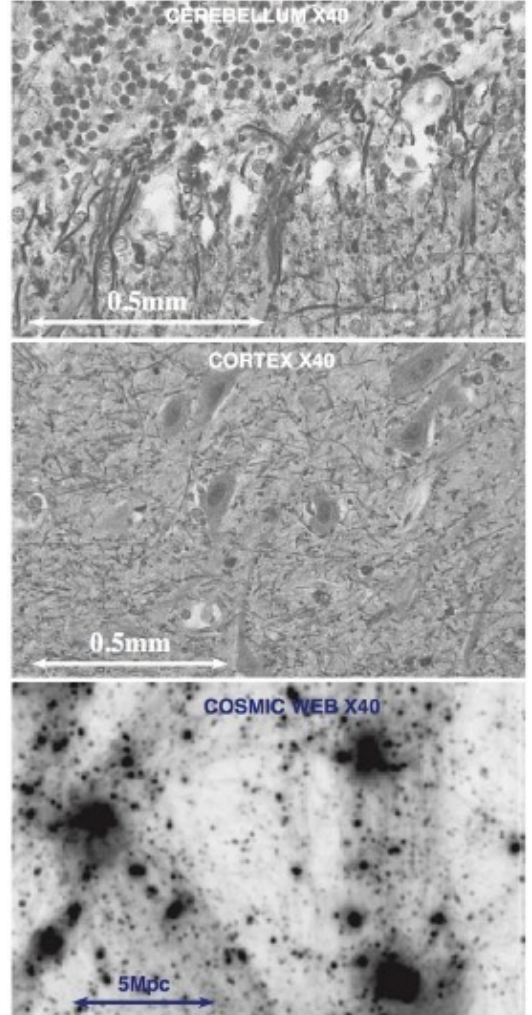

FIG. 1. **The image for each network with their scales.** Visualizing image samples with the scales of the Cortex, Cerebellum, and Cosmic with 40x magnification.[1]

II. DATA DESCRIPTION

In order to obtain these complex networks, we utilize an image representation of each system. Employing image-processing techniques, we identify the nodes within the image and determine the links and adjacency matrix based on the spatial relationships between the

nodes.

A. Microscopy

The study involved examining multiple samples of human cerebral and cerebellar cortex that were preserved using formalin fixation and paraffin embedding. The slices were taken at a depth of 4 μm and were viewed under magnification factors of 4×, 10× and 40×. The researchers used a Neurofilament (2F11) Mouse Monoclonal Antibody (Ventana/CellMarque/Roche) to label the neurofilaments. The samples were processed using Ventana BenchMark Ultra Immunostainers, and a Nikon eclipse 50i microscope was used to visualize them. The researchers avoided using magnifications larger than 40× to achieve better optical depth resolution and to reduce the non-linear response of the optic microscopy. The size of each image in 40× magnification is 1.6 mm.

B. Cosmological Simulations

We used synthetic samples of the cosmic web from a high-resolution $\left(2400^3\right.$ cells and dark matter particles) simulation of a cubic $100^3 \mathrm{Mpc}^3$ cosmic volume $\left(1\mathrm{Mpc} = 3.085 \cdot 10^{24}\right.$ cm), performed with the grid code ENZO [2]. The simulation produces a realistic distribution of dark matter, ordinary matter, and magnetic fields at the present epoch. In order to mimic the "slicing" procedure of brain tissues, we produced 12 different thin slices (with thickness 25Mpc) from the simulated volume, by extracting four slices in perpendicular directions with respect to each of the independent axes of the simulation. We are using the same data of the [1] study at this URL: https://cosmosimfrazza.myfreesites.net/cosmic-web-and-brain-networkdatasets.

III. NODE DETECTION

In each image, the nodes are distinguishable by their high value. In the Cosmic web, the nodes are the dense spots and in the brain samples neurons are characterized nodes. we are using find_peaks algorithm from the Photutils [3] python library which finds local peaks in an image that are above a specified threshold value. our threshold contains the median and standard deviation of the image. The peaks found in the cosmic image are reported in Figure 2.

Peaks are the maxima above the threshold within a local region. The local regions are defined by either the box size which we set as 10 pixels in this project. Box size defines the local region around each pixel as a square box. If multiple pixels within a local region have identical intensities, then the coordinates of all such pixels are returned. Otherwise, there will be only one peak pixel per local region. Thus, the defined region effectively imposes
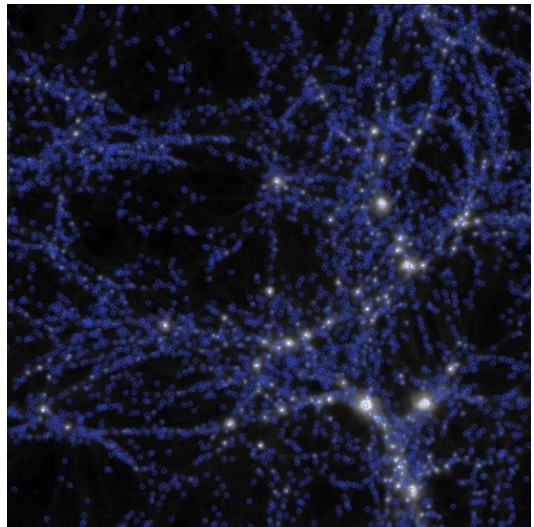

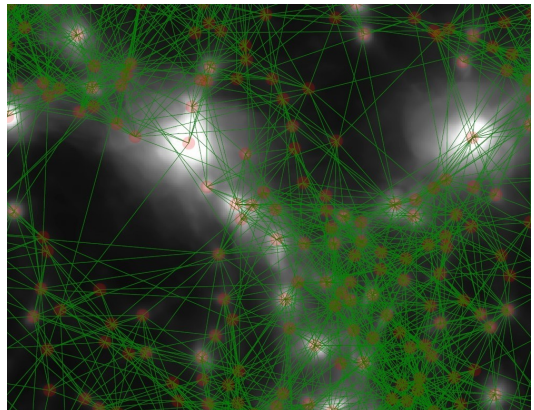FIG. 2. **Nodes in Cosmic web image.** 3992 nodes are found in the cosmic web image using Photutils.

FIG. 3. **Visualization of the generated network.** Plotting the generated network on the cosmic web image. The nodes are shown by red dots and links are shown by green lines.

a minimum separation between peaks unless there are identical peaks within the region. We found 3992 nodes in the cosmic web image, 2492 nodes in the Cerebellum image, and 1560 nodes in the Cortex sample.

IV. NETWORK CONSTRUCTION

Once the nodes have been identified, the network is generated by establishing links based on the distance between nodes within the image. We stochastically sample connections from a probability distribution that exponentially decays with node-node Euclidean distance, $p \propto e^{-r/r_0}$, where distance r_0 obtained by the ref[1]. In case the network is not connected, we take the largest connected component and remove the smaller ones. We chose r_0=28 for the cosmic web equivalent 1.2Mpc in the
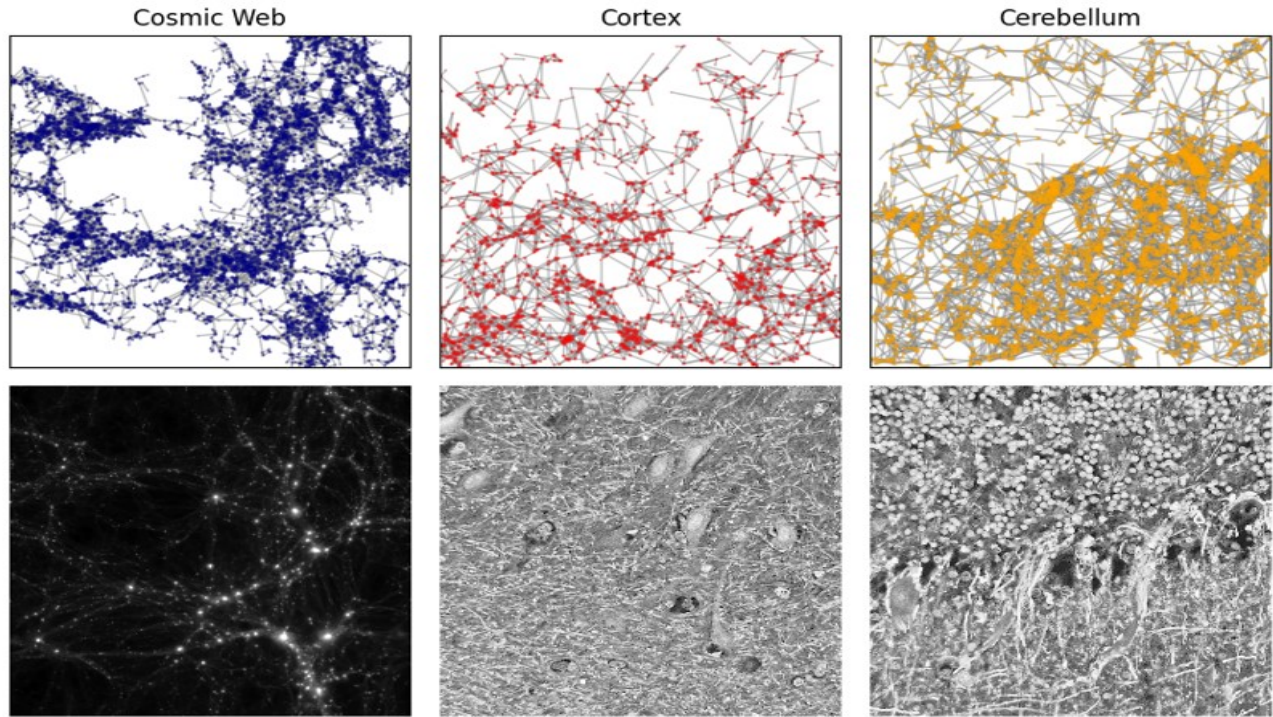
FIG. 4. **Networks constructed from the images.** The top row shows visualizations of one network realization obtained for each system, given by the images represented in the bottom row. The positions of network nodes are aligned with their pixel positions in the image. The node identification and link sampling procedure are described in the Data section.

cosmic image and r_0=24 for the brain samples which is 32µm in real scale. We then built the adjacency matrix of nodes, A_{ij}, i.e., a matrix with rows/columns equal to the number of detected nodes, with value $A_{ij} = 1$ the nodes will be connected with probability $p \propto e^{-r/r_0}$ or $A_{ij} = 0$ otherwise. This stochasticity allows for generating ensembles of networks for any set of spatially distributed. This method selects from $N \sim 3992$ nodes in our cosmic web slices, with an average number of $\langle k \rangle \sim 4.68$ connections per node. For the cerebellum slices we measured $\langle k \rangle \sim 7.14$ with $N \sim 2492$, while for the cortex we measured $\langle k \rangle \sim 11.41$ for the $N \sim 1560$ identified nodes. We generate 30 realizations for each image with this method which will be different because of the stochasticity of the linking algorithm. Then apply all the network measurements for all of these generated networks to have more accurate results and avoid thresholding error. The links and nodes can be seen in Figure3 for the Cosmic web, and also all the networks in Figure4.

V. METRICS

In order to compare our networks we use both classical and modern metrics which pave the way for better analysis. we compute the average shortest path, diameter, clustering coefficient, modularity, and the spectral entropy.

A. Classical metrics

Average shortest path: A path is a sequence of links connecting two nodes. Among every pair of nodes, it is possible to find the path that has the minimum number of links. This path is called the shortest path. To calculate the average shortest path, we first find the shortest path between each pair of nodes in the network, and then take the average of all these shortest paths. Networks with shorter average shortest paths generally allow for faster and more efficient communication between nodes. In contrast, networks with longer average shortest paths can be less efficient and may have more bottlenecks or barriers to information flow. The formula for calculating the average shortest path in a network with N nodes is:

$$\langle L \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} d(i,j)$$

where $\langle L \rangle$ is the average shortest path, d(i,j) is the shortest path length between node i and node j, and the summation is taken over all pairs of distinct nodes i and j in the network.

Diameter : In network science, the diameter of a network is a measure of its size, and it is defined as the
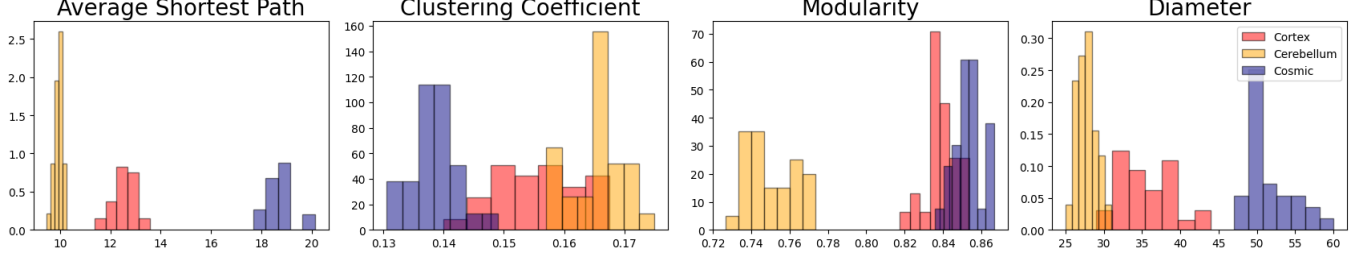
Statistical Comparison of Ensembles

FIG. 5. **Statistical comparison of ensembles.** The network ensembles are compared in terms of the average shortest path, clustering coefficient, modularity, and diameter, and their probability distribution is reported here. While some similarities between the cortex and galaxy are shown in terms of clustering coefficient and modularity, according to the average shortest path and diameter, the three systems are plainly different.

longest shortest path between any two nodes in the network.Formally, the diameter of a network $G = (V, E)$ is defined as:

$$\text{diameter}(G) = \max_{u,v \in V} d(u, v)$$

where $d(u, v)$ denotes the shortest path between nodes u and v in the network. Intuitively, the diameter of a network represents the maximum distance between any two nodes in the network, and it provides a measure of how connected or spread out the nodes are. Networks with a small diameter are often considered more efficient, as they enable faster communication and information exchange between nodes.Finding the diameter of a network can be computationally expensive, especially for large and complex networks.

Clustering coefficient : The clustering coefficient is a measure of the extent to which nodes in a network tend to form clusters or tightly-knit groups. Specifically, the clustering coefficient of a node measures the proportion of its neighbors that are connected to each other, out of all possible connections between them.There are several ways to define the clustering coefficient, but one common definition is as follows: The clustering coefficient C_i of a node i in a network is given by:

$$C_i = \frac{2t_i}{k_i(k_i - 1)}$$

where t_i is the number of triangles (i.e., three nodes connected to each other) that include node i, and k_i is the degree of node i, which is the number of edges incident to it. The clustering coefficient of a network is then defined as the average of the clustering coefficients of all its nodes:

$$C = \frac{1}{N} \sum_{i=1}^{N} C_i$$

where N is the number of nodes in the network.The clustering coefficient is a useful measure of the local

structure of a network, as it can provide insights into the formation of cliques or tightly-knit groups of nodes. In social networks, for example, high clustering coefficients are often observed in communities of friends or colleagues who tend to interact more frequently with each other than with outsiders.

Modularity : Modularity is a measure of the extent to which a network can be partitioned into communities or modules. In other words, it quantifies how well the nodes in a network are organized into groups that are densely connected within themselves and sparsely connected with the rest of the network. The modularity of a network is defined as:

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

where A_{ij} is the weight of the edge between nodes i and j, k_i and k_j are the degrees of nodes i and j, respectively, m is the total weight of the edges in the network, c_i and c_j are the communities to which nodes i and j belong, respectively, and $\delta(c_i, c_j)$ is the Kronecker delta, which equals 1 if nodes i and j belong to the same community and 0 otherwise.

B. Spectral entropy

Spectral entropy is a network complexity measure obtained from the eigenvalues of the graph Laplacian matrix [4]. Let the adjacency matrix describing the connections be A, where each element A_{ij} of the matrix is 1 if an edge exists between nodes i and j and is 0 otherwise. Let D be the degree diagonal with off-diagonal elements of 0 and the diagonal element D_{ii} encoding the degree of node i. The graph Laplacian follows $\mathcal{L} = A - D$ with eigenvalues $\lambda_i, i = 1, 2, ...N$, where N is the number of nodes. The spectral entropy follows:

$$S_\tau = -\sum_{i=1}^{N} \frac{e^{-\tau\lambda_i}}{Z} \ln \frac{e^{-\tau\lambda_i}}{Z},$$

where τ is a resolution parameter enabling a multiscale analysis, and $Z = \sum_{i=1}^{N} e^{-\tau\lambda_i}$.
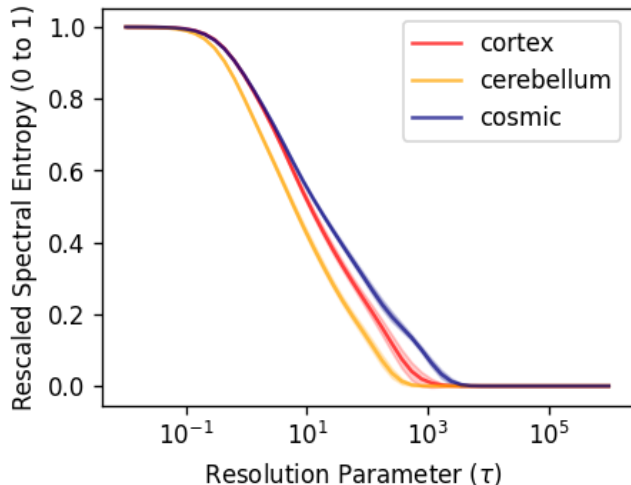

FIG. 6. **Spectral entropy analysis.** Lines indicate the mean spectral entropy of the cortex, cerebellum, and galaxy, and the shadows indicate the standard deviations, where τ is a resolution parameter that indicates the scale of analysis. At small scales, the three systems are indistinguishable. At large scales, they are distinct. However, at middle scales, the cortex shows high similarity with the galaxy.

The spectral entropy is bounded between $\ln N$ and 0. Therefore, we rescale it as $S_\tau \to \frac{S_\tau}{\ln N}$ to allow for comparisons across systems of different sizes [5].

VI. RESULTS

We use the classical network metrics introduced in the Methods to compare network ensembles and report their probability distribution in Fig. 13. Some overlap is observed between the cortex and the galaxy. Yet, the average shortest path and diameter analysis reveals their significant difference.

Interestingly, the cerebellum shows the characteristics of small-world networks compared to the other two cases by having the lowest average shortest path and diameter while also exhibiting a high clustering coefficient. According to all four metrics, the cortex stands in the middle of the two other systems, slightly more similar to the cerebellum.

We also perform a multiscale spectral entropy analysis (See Fig. 6, where τ indicates the scale of analysis. For instance, small τ means that the complexity analysis is focused on nodes and the first neighbors, while large τ considers longer paths. As expected, all systems are similar at extremely small scales, confirming previous studies of spectral entropy [5]. However, their differences are pronounced at middle to large scales.

Interestingly, the cerebellum's spectral entropy decays more rapidly, with respect to the resolution parameter. Furthermore, it is worth noting that the error bars of different systems don't touch each other after a certain value of τ.

The codes which we did for this study are available on GitHub at this URL:
https://github.com/mojee13/Comparison-Between-the-Neuronal-Network-and-the-Cosmic-Web

VII. CONCLUSION

In this project, we processed images from the human cortex, cerebellum, and galaxy to locate nodes and build ensembles of networks representing each case, by sampling links from an exponential probability distribution $p \propto e^{-r/r_0}$, where r_0 is a characteristic length given by the previous sudies [1].

We compare the ensembles using different metrics such as average shortest path, diameter, clustering coefficient, modularity, and spectral entropy.

Overall, we confirm some similarities between the cortex and the galaxy. However, our analysis of the average shortest path, diameter, and spectral entropy indicates significant differences between the human brain and the galaxy, suggesting that these systems are quite distinct.

VIII. EXTENDED ANALYSES

We added other network analyses on our networks, including:

1. **Homophily (See Fig. 7)**: We take the intensity of node pixels in the images to be an attribute, being Low, Mid and High intensities, and plot the topological and spatial representations of each network. Then, we calculate the homophily and, also, assortativity. The positive assortativity values we obtained indicate that the networks are to some extent assortative. In an assortative network, high-degree nodes connect with each other avoiding low-degree nodes.

2. **Comunity detection and the number of communities (See Fig. 8 and Fig.14)** Firstly, we perform **Louvain** method, which is a hierarchical agglomerative clustering algorithm used to optimize modularity. It starts by assigning each node to its own community and then iteratively merges communities in a way that maximizes the increase in modularity. At each iteration, the algorithm considers each node and evaluates the increase in modularity resulting from moving the node to each
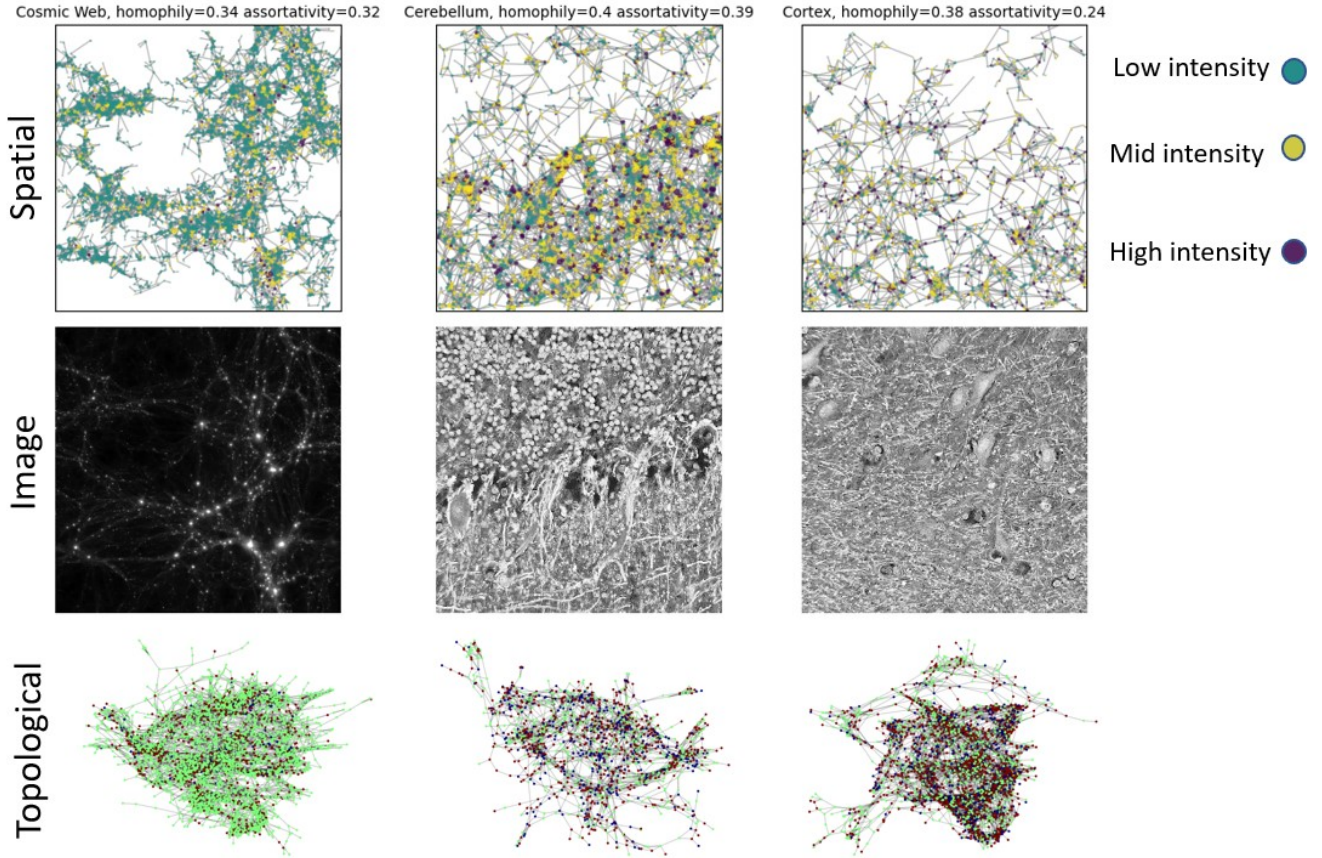
FIG. 7. **Homophily.**we illustrate Spatial, Image, and Topological representation for each network. For the cosmic web homophily is 0.34 and assortativity is 0.32. In terms of the Cerebellum, we can see that it has the highest homophily (0.4) and assortativity of which is 0.39. The homophily of the Cortex network stands at 0.38 and the assortativity of which is 0.24.

neighboring community. This process is repeated until no further improvement in modularity can be achieved by merging communities. Secondly, we use **Girvan-Newman** algorithm. The Girvan-Newman algorithm is a hierarchical clustering algorithm based on the concept of edge betweenness, measuring the number of shortest paths that pass through a given edge in a network. The algorithm starts by calculating the edge betweenness for all edges in the network and removing the edge with the highest betweenness. This process is repeated iteratively, with the network being split into smaller and smaller components as edges are removed. The resulting dendrogram can be cut at different levels to obtain different network partitions into communities. The Girvan-Newman algorithm is computationally expensive but effectively detects communities in various networks, including social networks, biological networks, and transportation networks. Note that due to the large size and connectivity of our networks, our systems could not run the full Girvan-Newman for two of our networks. However,

in the cortex, the algorithm converged, and we reported the results. Besides, we use the **Greedy modularity optimization** algorithm that starts with an initial network partition and then iteratively optimizes modularity by moving nodes from one community to another. At each iteration, the algorithm considers each node and evaluates the increase in modularity resulting from moving the node to each neighboring community. The node is then moved to the community, which results in the largest increase in modularity. This process is repeated until no further improvement in modularity can be achieved by moving nodes between communities.

3. **Robustness (See Fig. 9)**: We analyze the robustness of each network against random failures and targeted attacks led by degree, betweenness, eigenvector, closeness, and PageRank centrality.

4. **Degree distributions (See Fig. 10 and Fig. 11)**: we plot the logarithmic degree distribution of each network and fit the exponents. Also,
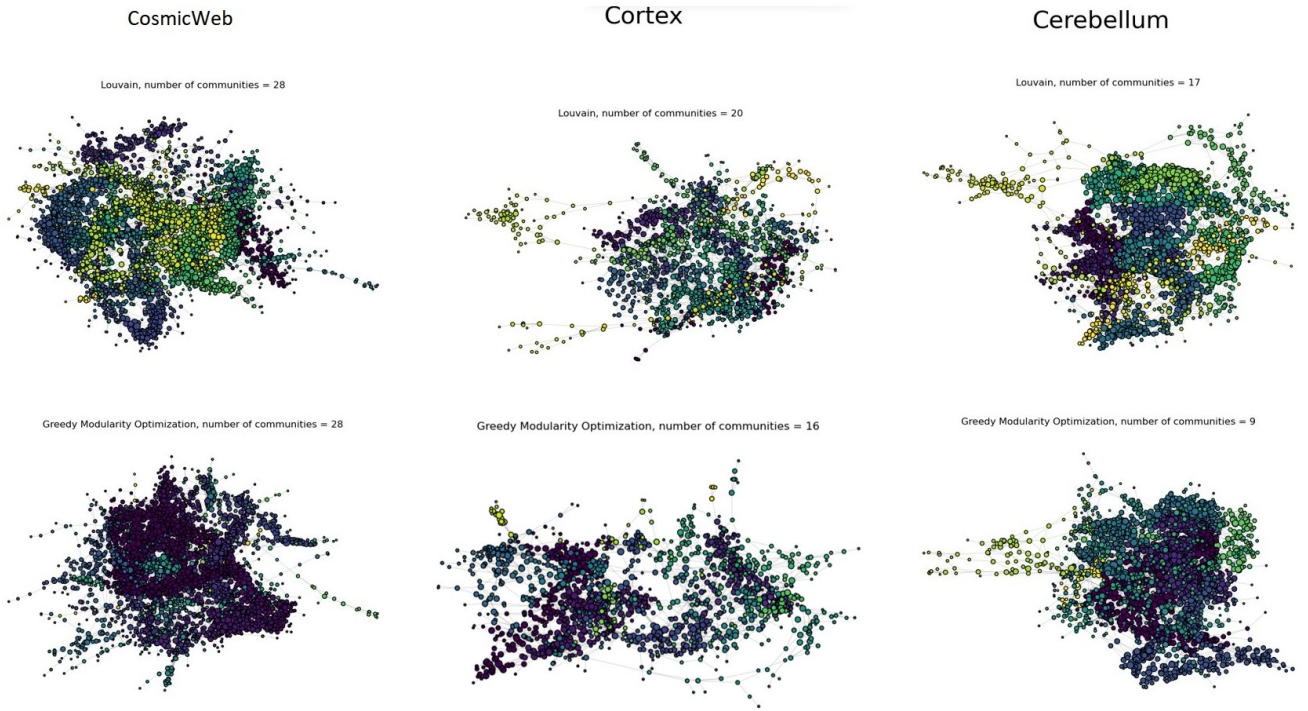
FIG. 8. **Community detection with Louvain and Greedy Modularity algorithms.** We find 28 communities for the cosmic web, 20 communities for the cortex, and 17 for Cerebellum with the Louvain algorithm which is shown in the first row. However, the Greedy Modularity finds 28,2,9 for Cosmicweb, cortex, and cerebellum, respectively, which can be seen in the second row.
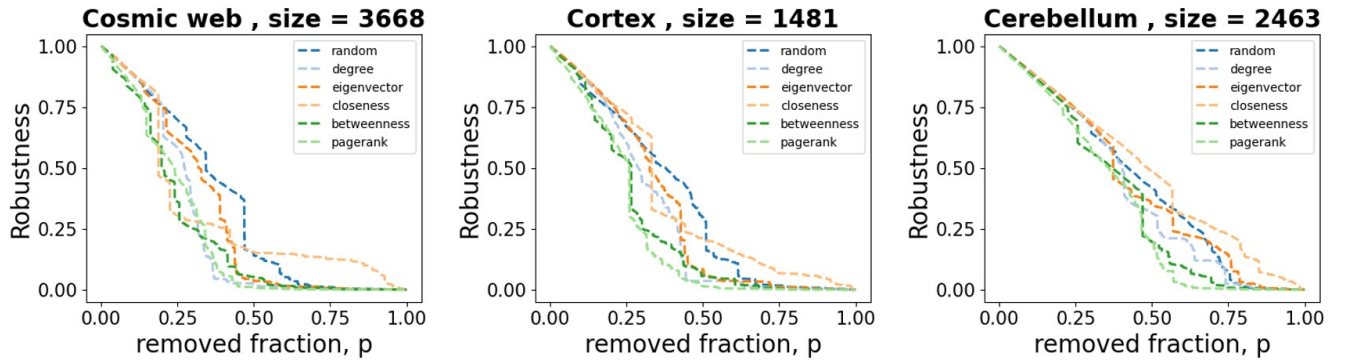

FIG. 9. **Robustness analysis.** Analyzing the robustness of each network with random failure and targeted attacks(random, degree, eigenvector, closeness, betweenness, PageRank). PageRank and betweenness targeted attacks have better performance. Cerebellum seems to have the highest robustness to random failures and targeted attacks.

we plot the degree distribution histogram of each network. The networks are not heterogeneous.

5. **PageRank versus Degree Centrality (See Fig. 12)**: We plot the degree centrality versus page rank for the three networks and perform a linear regression.

6. **Degree Correlation (See Fig. 15)**: we calculated the degree correlation matrix is the product of the probability of each degree and showed that all three networks are assortative which indicates that nodes of comparable degree tend to link to each other.
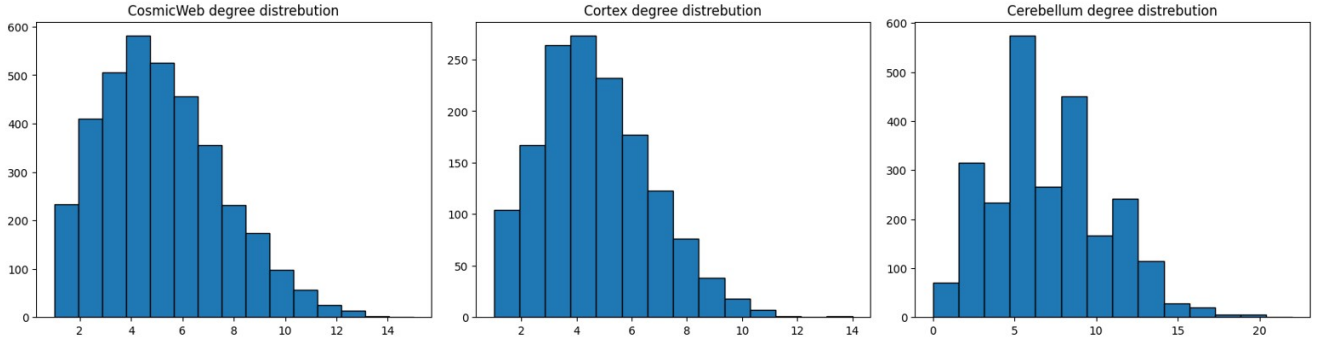
FIG. 10. **Degree distributions for three networks.** the degree distribution is shown in the histogram, which represents the homogeneity and heterogeneity of each network. The histograms resemble poison distribution.
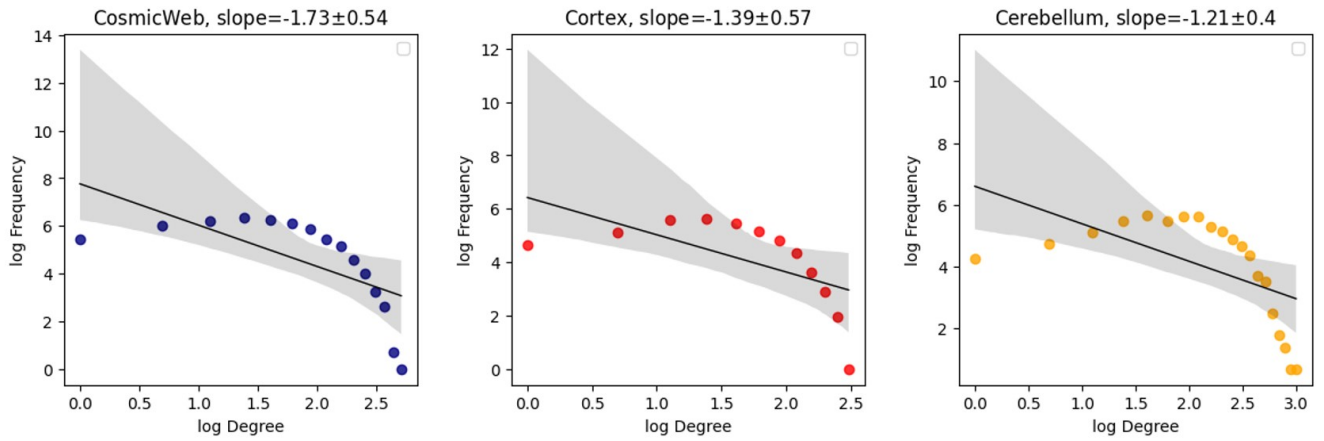

FIG. 11. **logarithmic Degree distributions for three networks.** the logarithmic degree distribution of three networks is reported, a curve is fitted and the shadow describes the std error. The degree distribution is not a power law.

[1] F. Vazza and A. Feletti, The quantitative comparison between the neuronal network and the cosmic web, Frontiers in Physics **8**, 491 (2020).

[2] G. L. Bryan, M. L. Norman, B. W. O'Shea, T. Abel, J. H. Wise, M. J. Turk, D. R. Reynolds, D. C. Collins, P. Wang, S. W. Skillman, *et al.*, Enzo: An adaptive mesh refinement code for astrophysics, The Astrophysical Journal Supplement Series **211**, 19 (2014).

[3] L. Bradley, B. Sipocz, T. Robitaille, E. Tollerud, C. Deil, Z. Vinícius, K. Barbary, H. M. Günther, A. Bostroem,

M. Droettboom, *et al.*, Photutils: Photometry tools, Astrophysics Source Code Library , ascl (2016).

[4] M. De Domenico and J. Biamonte, Spectral entropies as information-theoretic tools for complex network comparison, Physical Review X **6**, 041062 (2016).

[5] A. Ghavasieh and M. De Domenico, Multiscale information propagation in emergent functional networks, Entropy **23**, 1369 (2021).
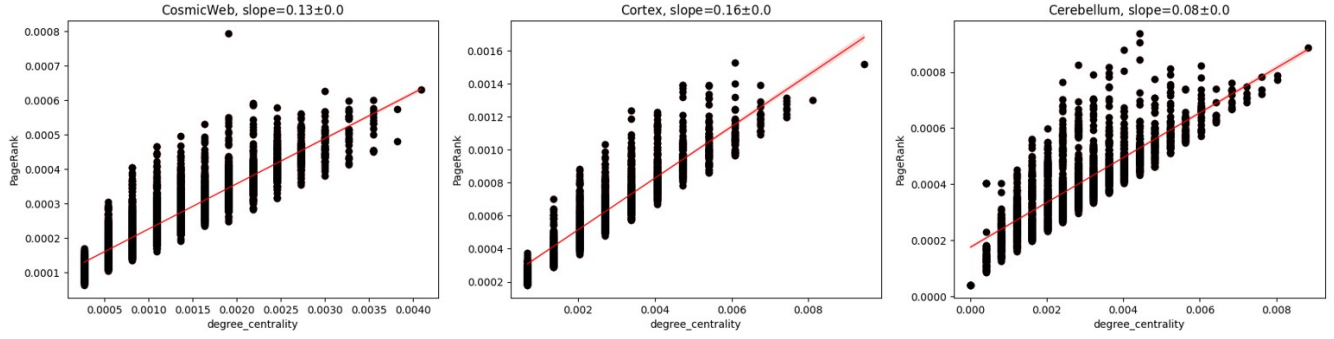
FIG. 12. **Degree centrality versus PageRank.** we plot the degree centrality versus page rank for our three networks and linear fit on it. Also mentioned is the slope of the fitting. we do not have any Hubs in these networks as we can see.
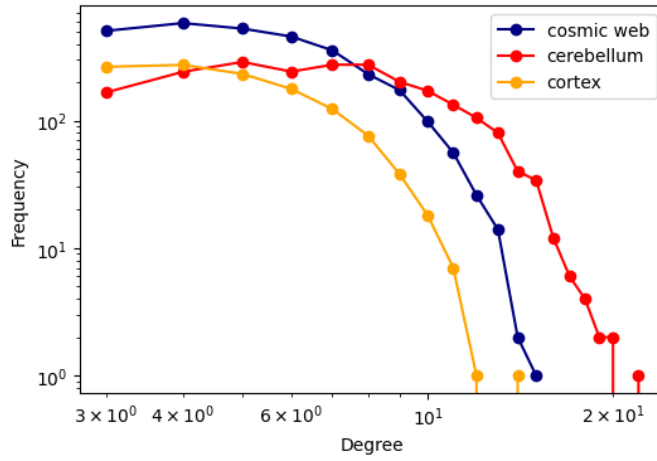

FIG. 13. **comparison of logarithmic degree distribution.** this figure shows the degree distribution of each network. the red dots which belong to the cerebellum has some nodes with higher degrees to compare with the cortex and cosmic web.
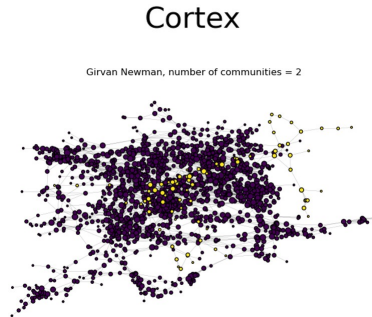

FIG. 14. **Community detection with Girvan-Newman algorithm.** We find 2 community for Cortex with Girvan-Newman method.
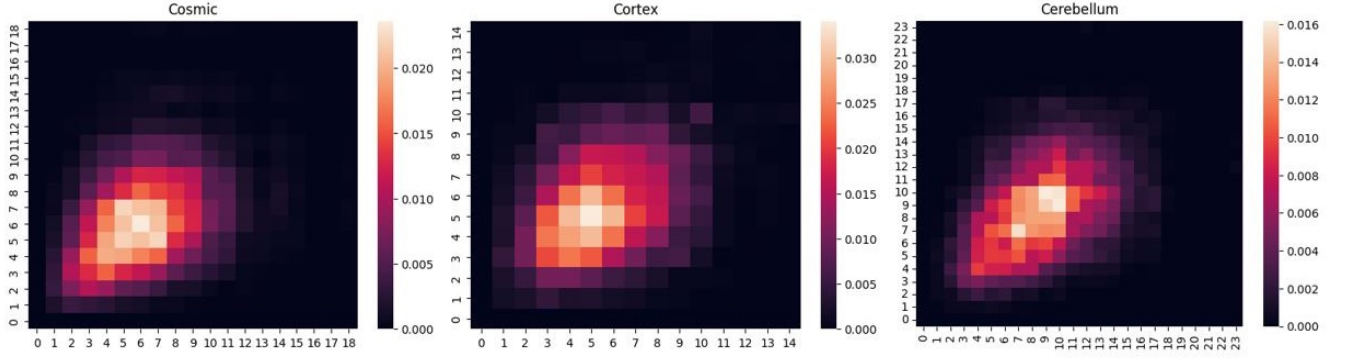
FIG. 15. **Degree correlation matrix.** it shows us that all three networks are assortative which indicates that nodes of comparable degree tend to link to each other. the cerebellum shows that it is the most assortative with the highest average degree of 10 and the cortex shows the least (almost neutral) with the lowest average degree of 5 whereas the cosmic web is somewhere in between with an average degree of 6.
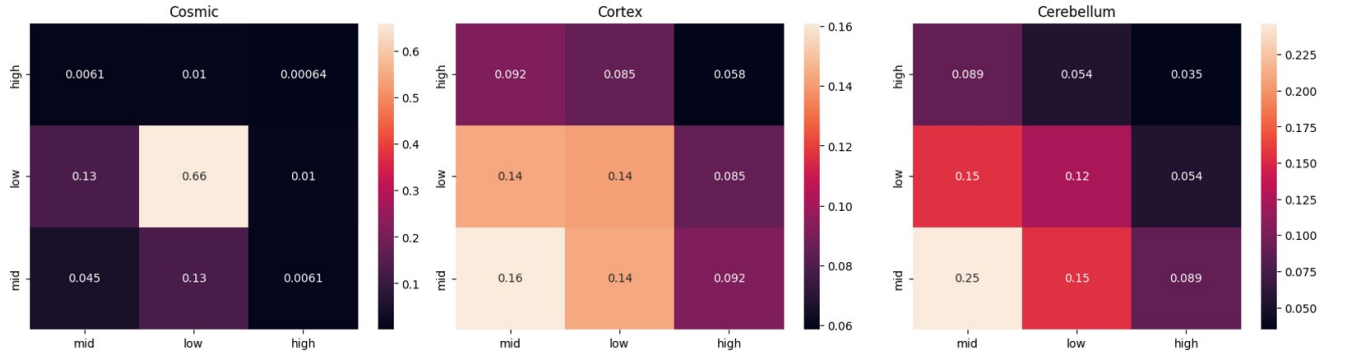

FIG. 16. **Correlation maps with neighbors for homophily** These plots show the correlation for each attribute in each network. Nodes with 'low' attributes have the most correlation with themselves in Cosmic Web and in the Cortex as the same Cerebellum, the 'mid' has the most correlation with themselves.
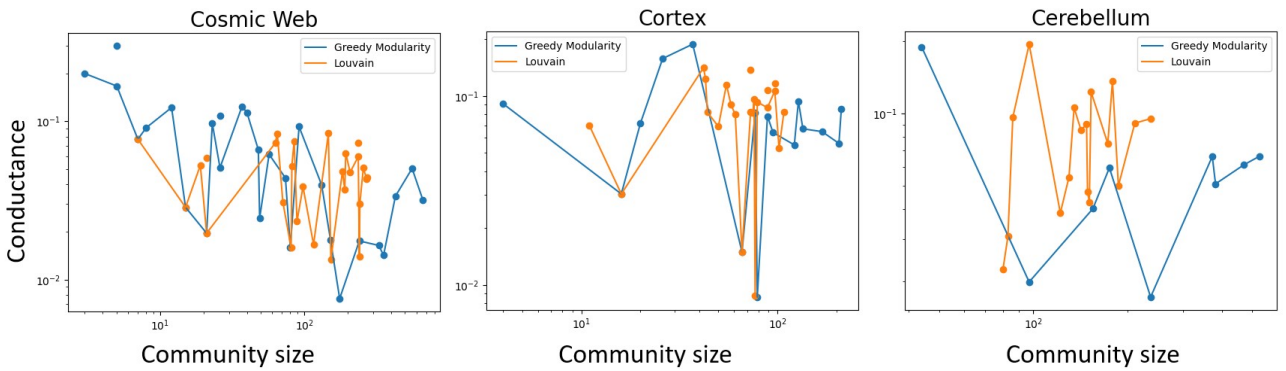

FIG. 17. **Comunity Profile.** Each plot shows the community profile for each network. We detect communities with Louvain and Greedy Modularity algorithms and find the minimum conductance for each community size.
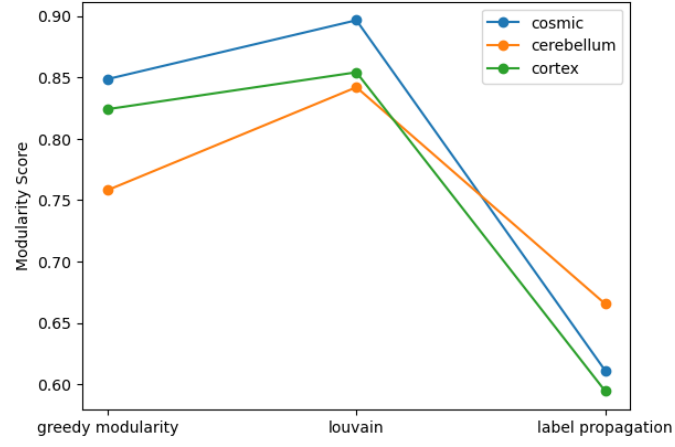
FIG. 18. **Modularity Score.** We find modularity scores with Louvain and Greedy Modularity and Label Propagation community detecting methods for each network. The Louvain method got the most modularity and also greedy modularity has a high score but the label propagation has less score than others. Also, the cosmic web has the most modularity, and the cerebellum has the less score except in the label propagation method.