



---

# Graph Theory and Null Models Assignment

---

## AUTHORS

ChangHui Simone Lin - [changhuisimone.lin@unicam.it](mailto:changhuisimone.lin@unicam.it)

February 18, 2026

# Assignment

## Point 1

**Task:** You can build the brain network of *C. elegans* from the files `NeuronConnect.xls` (`NeuronTypes` give you additional information) or using directly the `brainLightedAdj` file.

**Solution:** The brain network of *C. elegans* was constructed by parsing the connectivity data from `NeuronConnect.csv`. To focus strictly on the neuron-to-neuron brain network, all Neuromuscular Junction (NMJ) connections were filtered out.

The network was built as a directed graph with the following logic:

- **Gap Junctions (EJ):** Treated as bidirectional electrical synapses; edges were added in both directions ( $i \rightarrow j$  and  $j \rightarrow i$ ).
- **Chemical Synapses (Sp, S, R, Rp):** Treated as directed edges from the pre-synaptic to the post-synaptic neuron.
- **Weights:** Multiple connections between the same pair of neurons were aggregated by summing their weights.

The resulting network topology consists of 279 Nodes ( $N$ ) and 4577 Edges ( $E$ ). This node count ( $N = 279$ ) aligns with the standard set of somatic neurons typically analyzed in the *C. elegans* connectome (Varshney et al., 2011), confirming that non-somatic or pharyngeal neurons/muscles were successfully excluded.

## Point 2

**Task:** Considering the overall nematode brain network (gap junction + synapse), calculate the degree distribution, the distance distribution, the node degree correlation (you can plot the degree vs. knn as shown in class), the clustering, the diameter.

**Solution:** The fundamental topological metrics of the constructed network (gap junctions + synapses) were computed. The results are summarized below:

- **Average Degree ( $\langle k \rangle$ ):** The network is relatively dense for a biological graph, with an average degree of  $\langle k \rangle = 32.81$ .
- **Diameter ( $D$ ):** The diameter is 5, indicating a globally compact structure where the farthest two nodes are only 5 hops apart.
- **Average Path Length ( $L$ ):** The characteristic path length is short,  $L = 2.44$ . This low value facilitates rapid signal integration and communication across the organism.
- **Clustering Coefficient ( $C$ ):** The average clustering coefficient is  $C = 0.34$ . This value indicates a strong tendency for triangular motifs and local processing units, significantly higher than what would be expected in a random graph of similar density.

- **Node Degree Correlation:** The correlation between a node's degree ( $k$ ) and the average degree of its neighbors ( $k_{nn}$ ) is visualized in Figure 1b. The plot exhibits a negative trend (disassortativity), where high-degree hubs tend to connect with low-degree nodes. This structure typically offers robustness against random failures but fragility to targeted attacks on hubs.

The distance distribution and degree correlation profile are visualized in Figure 1.

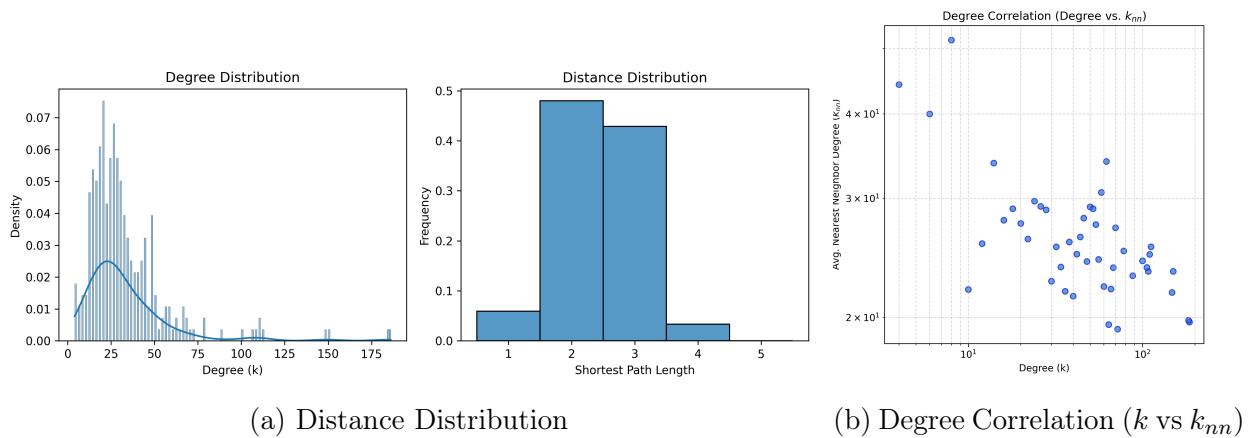


Figure 1: Topological properties of the *C. elegans* brain network. (a) The distance distribution (shortest path lengths) is unimodal and centered around path lengths of 2-3. (b) The negative slope in the  $k_{nn}$  plot confirms disassortativity, characteristic of biological neural networks.

### Point 3

**Task:** Fit the distributions (degree and distance) focusing on the tail. Comment the results.

**Solution:** To determine if the *C. elegans* brain network exhibits scale-free or small-world characteristics, I analyzed the tails of the degree and distance distributions using statistical fitting techniques.

- **Degree Distribution (Scale-Free Nature):** I fitted the tail of the degree distribution ( $k > 10$ ) to a power-law model  $P(k) \sim k^{-\gamma}$ . As shown in Figure 2a, the data follows a linear trend on a log-log scale, which is the hallmark of a scale-free network. The fitted exponent is  $\gamma \approx 1.44$ . This relatively low exponent (typically  $2 < \gamma < 3$  for random scale-free models) indicates a "heavy-tailed" distribution with the presence of significant **hubs**: highly connected neurons that serve as central communication bridges.
- **Distance Distribution (Small-World Nature):** The distance distribution (Figure 2b) shows a rapid decay in the tail, consistent with an exponential fall-off. The

maximum shortest path length (Diameter) is 5. This short diameter and rapid decay confirm the **small-world effect**, where the average path length  $L$  is significantly smaller than the network size ( $L \ll N$ ). This implies the network is globally compact, allowing for efficient signal propagation despite the sparse connections.

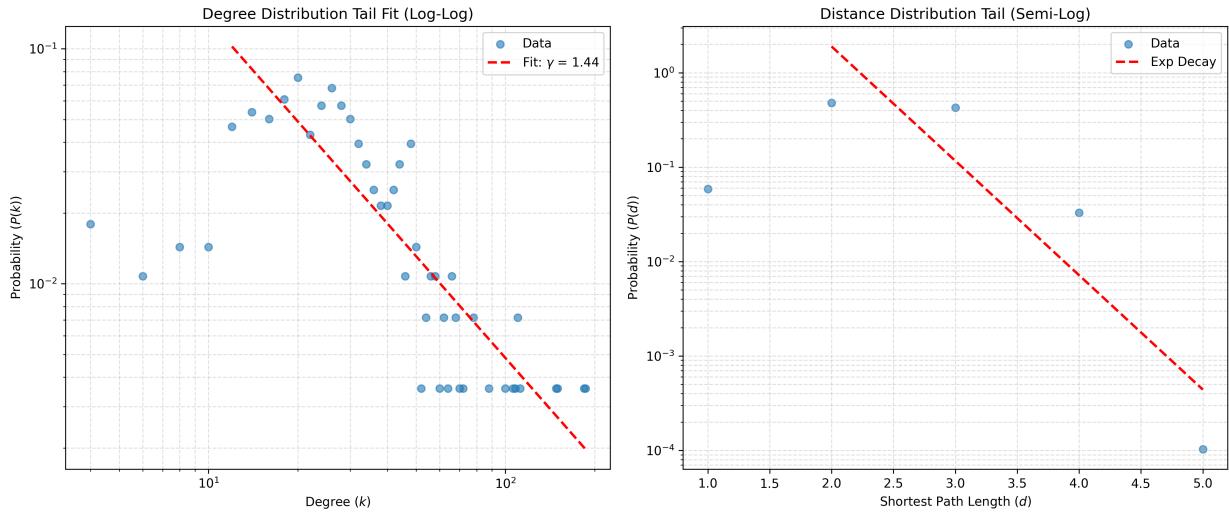


Figure 2: Tail analysis of topological distributions. (a) The degree distribution (log-log scale) fits a power law with  $\gamma \approx 1.44$  for  $k \geq 11$ , indicating a scale-free topology with rich-club hubs. (b) The distance distribution (semi-log scale) decays rapidly, with a maximum path length (diameter) of 5, confirming the compact small-world nature of the network.

## Point 4

**Task:** Randomize the network, i.e. compare the results with the corresponding ER null model. What properties are like random and what properties are different? Remember that you need to have a statistics of the results of the random null model (e.g. do 100 realization and calculate the mean + std or use analytical results if you want).

**Solution:** To assess the significance of the network's topological features, I compared the empirical C. elegans brain network with an Erdos-Renyi (ER) null model. I generated 100 realizations of random graphs preserving the same number of nodes ( $N = 279$ ) and edges ( $E = 4577$ ) as the empirical network. The results are summarized in Table 1.

Metric	Empirical	Random (Mean $\pm$ Std)	Analytical	Z-Score
Clustering Coeff. ( $C$ )	0.3371	$0.0590 \pm 0.0008$	0.0590	364.96
Avg. Path Length ( $L$ )	2.4356	$2.2988 \pm 0.0023$	2.0129	60.34

Table 1: Comparison of topological metrics between the empirical network and the ER null model (100 realizations). Analytical predictions are calculated using  $C \approx \langle k \rangle / N$  and  $L \approx \ln N / \ln \langle k \rangle$ .

**Interpretation:**

- **Clustering Coefficient:** The empirical clustering coefficient ( $C \approx 0.34$ ) is significantly higher than that of the random null model ( $C_{rand} \approx 0.06$ ), with a massive Z-score of  $\approx 365$ . This deviation confirms that the brain network is not random but highly structured, characterized by local cliques and functional modules.
- **Average Path Length:** The empirical path length ( $L \approx 2.44$ ) is slightly larger than the random equivalent ( $L_{rand} \approx 2.30$ ) but remains on the same order of magnitude. Although the Z-score (60.34) indicates statistical significance due to the low variance of the random model, the value satisfies the small-world condition  $L \ll N$  (where  $N = 279$ ).
- **Small-World Property:** The combination of high clustering ( $C \gg C_{rand}$ ) and short path length ( $L \approx L_{rand}$ ) classifies the *C. elegans* brain as a Small-World Network consistent with the Watts-Strogatz model. This topology supports efficient global information transfer (low  $L$ ) while retaining specialized local processing (high  $C$ ).

**Point 5**

**Task:** learn and apply a community detection algorithm to the networks, and comment the results.

**Method:** To identify the mesoscopic structure of the *C. elegans* brain network, I employed the **Louvain Method**, a heuristic algorithm designed to maximize the network's Modularity ( $Q$ ). As defined in the lecture notes, Modularity measures the density of links inside communities compared to links between communities. It is calculated as:

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

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  $j$ .
- $m$  is the total weight of all edges (or number of edges in unweighted graphs).
- $\delta(c_i, c_j)$  is the Kronecker delta, equal to 1 if nodes  $i, j$  belong to the same community, and 0 otherwise.

The algorithm iteratively optimizes this value by first moving nodes locally to increase  $Q$  and then aggregating communities into meta-nodes, repeating until convergence.

**Results and Interpretation:** The community structure detected in the undirected brain network is visualized in Figure 3.

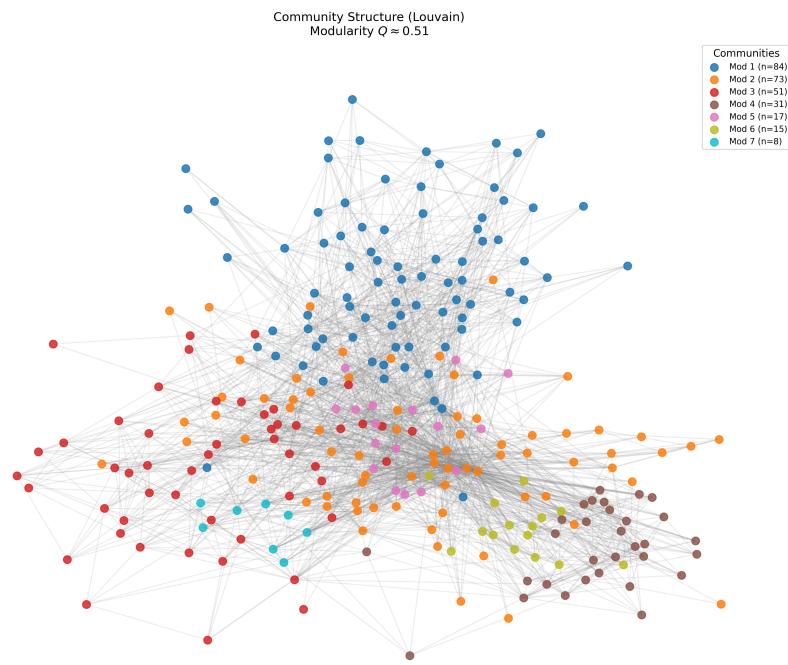


Figure 3: Community structure of the *C. elegans* brain network detected via the Louvain algorithm. Nodes are colored according to their modular assignment.

- **Modularity Score ( $Q$ ):** The analysis yielded a modularity of  $Q \approx 0.50$ . In the context of biological networks, values in the range  $[0.3, 0.7]$  indicate a strong non-random community structure. This suggests the brain is not a homogeneous web but is organized into distinct functional modules.