Effects of False Positive and False Negative Connections on Network-Level Metrics

Mariia Dolmatova

ABSTRACT

Accurate construction of brain connectomes critically depends on appropriate thresholding of connectivity matrices and choice between specificity and sensitivity in the research. However, no consensus exists on optimal strategies for finding a suitable balance. In particular, the trade-off between false positive (FP) and false negative (FN) connections can substantially bias network topology and graph-theoretical metrics. While previous work has proposed an approximate 1:2 FP-to-FN ratio as a suitable compromise for binary undirected networks, the generalizability of this recommendation to other network types remains unclear.

This study systematically investigates the effects of FP and FN perturbations on real-world connectome data from *Caenorhabditis elegans*. Starting from a directed, weighted anatomical connectivity matrix derived from electron microscopy, FP and FN connections were introduced incrementally to simulate progressive network perturbation. Network-level metrics—including clustering coefficient, global efficiency, modularity, and number of modules—were tracked dynamically across weighted and binary as well as directed and undirected network representations.

Across most network types, clustering coefficient decreased with increasing FP and FN perturbations, while global efficiency exhibited opposing responses to FP and FN, consistent with prior findings. However, weighted undirected networks showed attenuated FP-induced increases in efficiency and distinct clustering behavior. Modularity and module count displayed high variability and lacked consistent trends, particularly in directed networks, suggesting fundamentally different modular organization compared to undirected cases. Notably, an increase in the number of modules under FN perturbations was observed in weighted undirected networks, potentially reflecting topological fragmentation following the removal of critical connections. Together, these findings indicate that thresholding strategies optimized for binary undirected networks may not be directly transferable to weighted or directed connectomes, underscoring the need for network-type–specific or alternative approaches.

1. Introduction

The correct identification and thresholding of connectivity matrices is a crucial step in connectome analysis. Despite its importance, there is currently no consensus on an optimal approach. One of the biggest issues is finding an optimal balance between specificity and sensitivity: deciding to allow either more false positive (FP) connections by increasing sensitivity, or false negative (FN) connection by choosing specificity in the research. This question is crucial since it is now evident that inappropriate inclusion or exclusion of connections can substantially alter network topology and bias graph-theoretical metrics, potentially resulting in misleading conclusions (Rubinov & Sporns, 2010; Zalesky et al., 2016). Understanding how FP and FN errors affect network measures is therefore essential for the reliable interpretation of connectome analyses.

This challenge is further complicated by the variety of network types encountered in neuroscience. Many commonly used imaging techniques, such as fMRI or EEG are used to measure functional connectivity and mostly produce undirected connectivity matrices since they are based on correlational data, whereas techniques such as electron microscopy allow for the reconstruction of directed networks. Yet, the distinction is not strict and some functional data might be represented with directed connectivity (such as cross-correlation or Granger casualty) and vice versa, such as commonly used DTI and diffusion-MRI outcomes (Bastos & Schoffelen, 2016). MRI and other correlational methods matrices often yield fully dense matrices and require thresholding, frequently resulting in binary networks (Bullmore & Sporns, 2009; Rubinov & Sporns, 2010), whereas structural connectivity techniques such as DTI generate weighted matrices reflecting properties like synapse counts or streamline numbers, making weight preservation important for subsequent analyses. A lot of research was done on different binary networks and an approximate 1:2 FP-to-FN ratio was proposed (Zalesky et al., 2016) to balance the

influence on the network, however, the behavior of weighted networks under similar perturbations remains less well characterized.

This project focuses on studying the effects of FP and FN connections on real-world connectome data from *Caenorhabditis elegans*. Specifically, we examine how progressive introduction of FP and FN edges influences multiple network-level metrics across weighted and binary, as well as directed and undirected, network representations. By systematically comparing these network types, this study aims to clarify how different connectome constructions respond to connectivity errors and to inform methodological choices in connectome analysis.

2. Materials and methods

2.1 The data

The data was taken from the updated C. elegans weighted connectome (Vashney et al, 2011) that was an extension of the existing full-scale nervous system of C. elegans obtained through serial partitioning and electron microscopy (White et al, 1986). The network consists of 279 nodes, each corresponding to a neuron in the nervous system and each connection denotes the presence of a synaptic contact, the weight represents the amount of such contacts. The initial matrix is directed and has a density of 2.83%.

2.2 Processing and software

This data was used as a reference graph for subsequently introducing FP and FN connections one by one in order to generate perturbed networks. While FN simulation simply chose one of the existing edges with any weight and removed it, FP simulation drew the weight of each new edge from the empirical distribution of existing edge weights. Once a connection was established or removed, the new edited matrix was becoming the reference for a future analogical modification

with constant reassigning. This was repeated 300 times to simulate gradual increase in the network perturbation and study the effects of FN and FP connections in dynamics. On each step, the clustering coefficient (C), global efficiency (GE), modularity and the amount of modules were documented to measure the network behavior. C, GE, and modularity are standard network-level metrics that are well suited for dynamic analysis; the number of modules was added in order to get more insight on the change in modular topology of the network. The processing was performed in Matlab (2025b) with the usage of BCT toolbox.

2.3 Extrapolation of the method

In order to account for different network types, the initial matrix was then binarized to simulate binary undirected connectome, and the whole process was repeated. In the binary case the weights for FP were equal 1. Later the initial weighted matrix was symmetrified and binarized again to also simulate symmetrical weighted and unweighted data and was fed into the same pipeline.

3. Results

Each of the measurements changes for all 4 networks were presented in graphs that capture the dynamical change as the perturbation grows.

3.1 Directed matrices analysis

Both weighted and binary cases clustering coefficients have shown to follow a similar descending trend with the increasing amount of both FP and FN connections. The global efficiency has shown the opposite changes for FP and FN and the FP induced curve was much steeper. The number of modules and modularity measurements had significant variations on each step and did not show any distinguishable trend (Figure 1).

3.1 Symmetrified (undirected) matrices analysis

Unlike all the other network types, the weighted undirected network did not show FP and FN shared descending trend, with FN varying around the same values. General efficiency has shown similar results with the directed networks, and for nodularity the clear descending slope is seen for binary matrices with FP connections, a slightly less clear but still noticeable slope is also seen for the weighted matrix FPs. The FN modularity measurements have a high variance around similar values and do not show any distinct trend. The number of modules appears to increase gradually in the weighted matrices with the increase of FN, but the high variability suggests that this trend may not be statistically meaningful. The number of modules of the binary matrix oscillated between 4 and 5 (Figure 2).
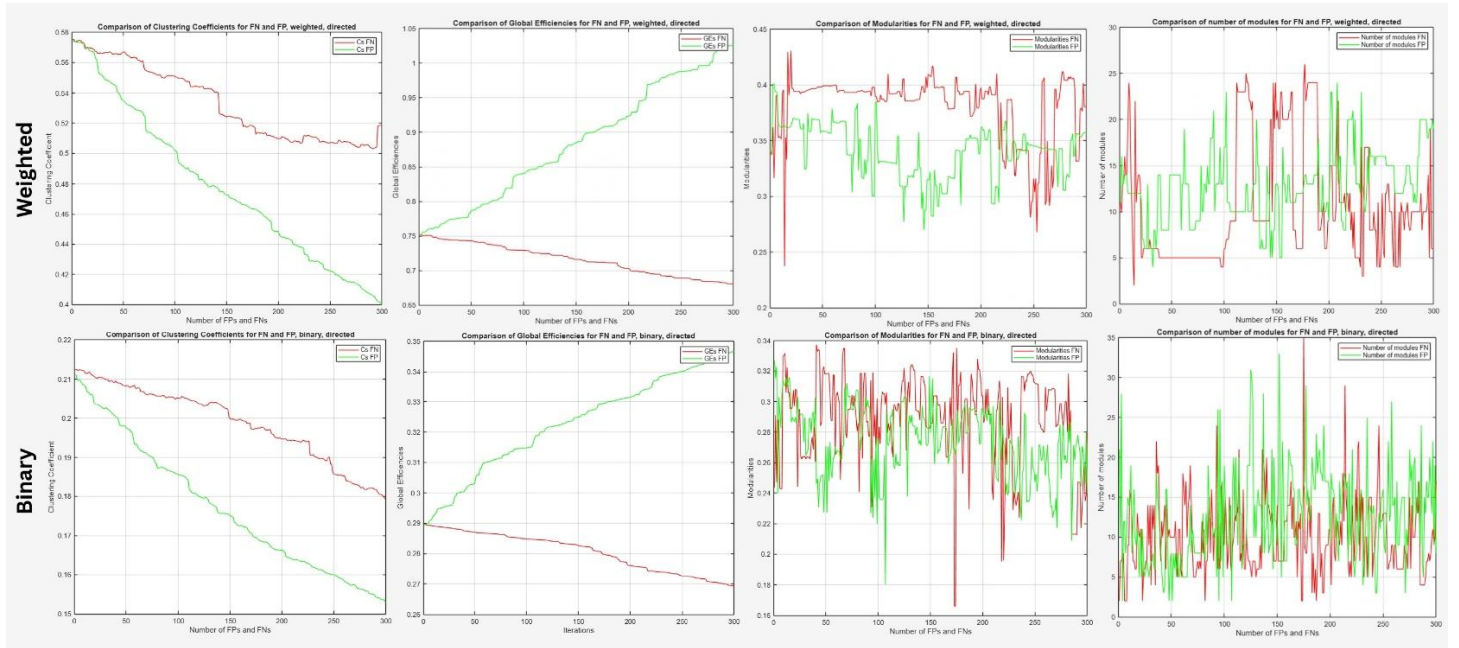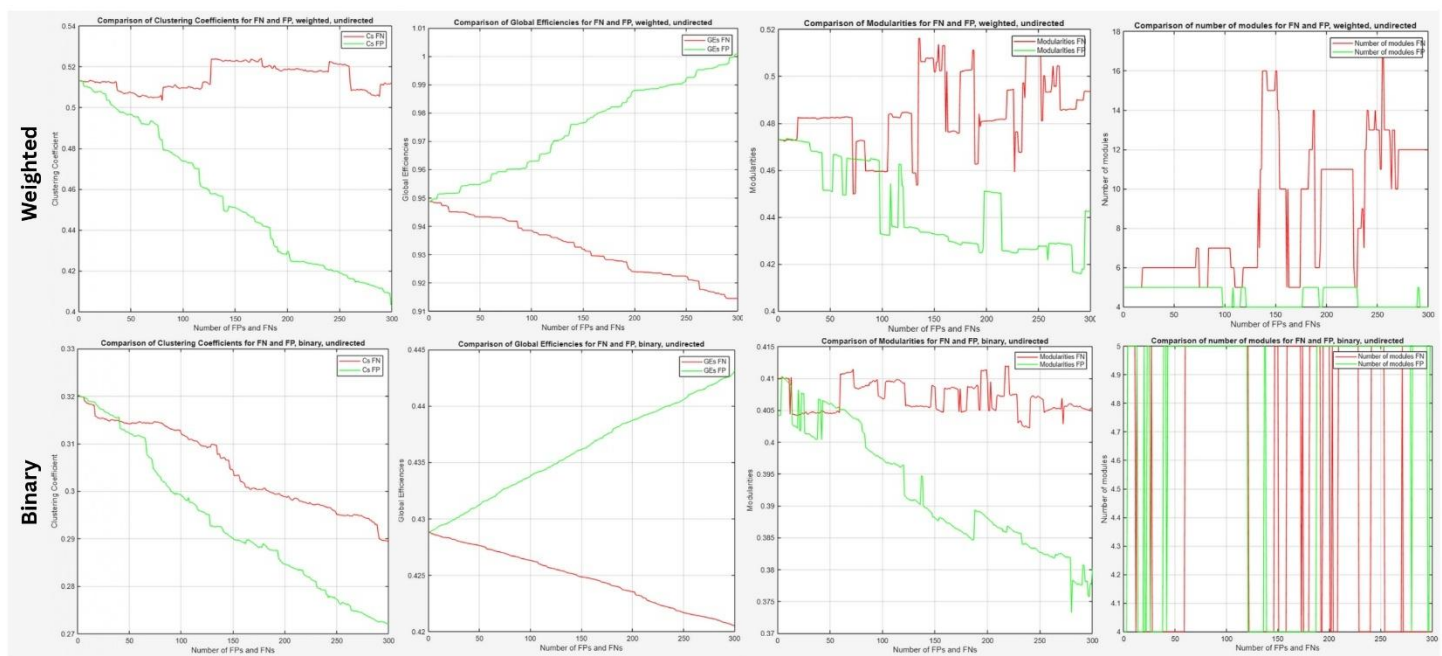

Figure 1. Clustering coefficients (Cs), Global efficiencies (GEs), modularities and numbers of modules changes in response to the amount of false positive (FP) and false negative (FN) connections for directed C. elegans graph. Upper row displays the measurements for weighted matrix, lower – binary, green line shows FP connections, and red line – FN.

4. Discussion

The global efficiency dynamics in response to both FN and FP perturbations across all examined networks showed the expected behavior, consistent with previous studies (Zalesky et al., 2016). However, global efficiency in the weighted undirected network exhibited a weaker FP-induced increase relative to the corresponding FN-related decrease. Similarly, the dynamics of the clustering coefficient followed the expected decreasing trend with increasing FP and FN perturbations in all cases except for the weighted undirected network.

Figure 2. Clustering coefficient, global efficiency, modularity and number of modules changes in response to the amount of FP or FN connections for C. elegans graph after symmetrization (undirected). Upper row displays the measurements for weighted matrix, lower – binary, green line shows FP connections, and red line – FN.

In contrast to previous findings, modularity in the directed networks did not display a descending trend with increasing FP connections, yet interestingly it was still distinguishable in the undirected matrices which was previously observed for the human connectome (Zalesky et at., 2016). The number of modules dynamic was not showing any trend, potentially apart from the weighted undirected network. The unexpected rise in the number of modules in response to increasing FN connections might be explained by significant topological changes brought by the loss of crucial connections (Bordier et al., 2017).

These results suggest that while previously proposed 1:2 FP-to-FN ratio recommendation for data cleaning might work well for binary undirected networks, it might not be the best suit some other network types, especially weighted cases. Alternative approaches proposed in the literature suggest include determining connectivity thresholds based on the fragmentation point of the largest connected component (Bordier et al., 2017), which was shown to work well for the real-life applications (Bordier et al., 2018).

The consistent discrepancies observed in the modularity coefficient across network types suggest fundamentally different modular organization in directed networks. These differences might indicate that thresholding strategies optimized for undirected networks may not be directly transferable, motivating further investigation into the selection of optimal thresholds for directed connectivity.

5. References

Bastos, A.M., Schoffelen, J.-M., 2016. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.* 9, 175. https://doi.org/10.3389/fnsys.2015.00175

Bordier, C., Nicolini, C., Bifone, A., 2017. Thresholding functional connectivity matrices using percolation theory. *NeuroImage* 163, 247–258. https://doi.org/10.1016/j.neuroimage.2017.09.060

Bordier, C., Nicolini, C., Bifone, A., 2018. Graph analysis and modularity of brain functional connectivity networks: Searching for the optimal threshold. *NeuroImage* 181, 687–702. https://doi.org/10.1016/j.neuroimage.2018.07.026

Bullmore, E., Sporns, O., 2009. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. https://doi.org/10.1038/nrn2575

Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* 52, 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Varshney, L.R., Chen, B.L., Paniagua, E., Hall, D.H., Chklovskii, D.B., 2011. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* 7, e1001066. https://doi.org/10.1371/journal.pcbi.1001066

White, J.G., Southgate, E., Thomson, J.N., Brenner, S., 1986. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B* 314, 1–340. https://doi.org/10.1098/rstb.1986.0056

Zalesky, A., Fornito, A., Bullmore, E.T., 2016. Connectome sensitivity or specificity: Which is more important? *NeuroImage* 142, 407–420. https://doi.org/10.1016/j.neuroimage.2016.06.035

Statement of the usage of AI

ChatGPT and Gemini were used for style-consistent rephrasing, grammar and spelling check, synonym search, in some cases it was used as a parser for literature finding and rarely as a text summarizing tool for preliminary article screening. These AI tools were also used several times for code error correction or function suggestions, however it was mostly inefficient or incorrect and were not implemented.