Graph Data Science to Improve Electrophysiology Data Analysis for Connectomics

This masters project aims to leverage data science techniques, including graph data-scientific and machine learning approaches, to improve the existing toolset for identification of micro-connectomic subgraphs. In 1986 John White, Sydney Brenner, and their team in England completed and published a project to document the network of neural connections in a small animal model–the C. Elegans worm [1], [2]. This landmark study can be considered the birth of the field of connectomics–a specialty in bioinformatics and neuroscience that deals with understanding human and other nervous systems by mapping their neurons' connections into a "connectome". The field has progressed with the anticipation that understanding nervous system circuitry will provide insight into high-level behavioral function and nervous system disease. The work done in the 1980's was tedious and constrained to very small organisms. Modern mapping of larger systems is now infeasible without computational approaches.

The original method for collection of a connectome used a microscope, often implementing electron microscopy with computer-aided feature recognition of microscopy images [1], [2], [5], [6]. These assays provide excellent resolution, and can even give insight about extracellular context (e.g., supporting cells) but cannot be collected in-vivo and require significant tissue preparation ex-vivo. More recent (2010-2015) approaches have used fMRI and dMRI to collect human connectomes in-vivo [7]. However, these MRI-based connectomes limit resolution to the voxel ($\approx$ 1mm 3D pixel) rather than to the cell. While not inhibiting association of brain regions to specific behaviors or other phenotypes, constrained resolution fails to measure essential nuance in processes between collections of cells inside each voxel. Additionally, the fMRI and dMRI techniques measure cellular activity only indirectly through mapping variables like blood flow, and cannot capture excitatory, inhibitory, or modulatory activity between cells necessary for precise characterization of inter-cellular relationships.

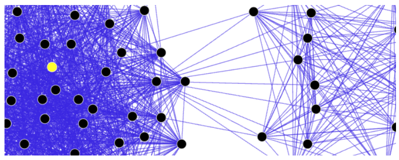

Figure 1: Graph data science is applied to a diverse set of problems in domains including social science, information retrieval, and pharmaceutical development. Image courtesy of: [3], [4]

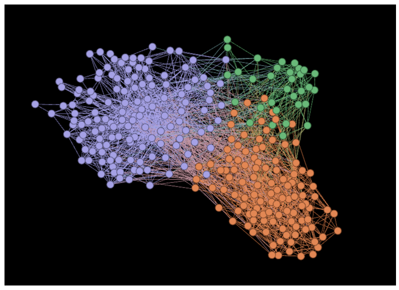

Figure 2: A visualization of the C. Elegans connectome partitioned using a graph autoencoder and k-means clustering, as performed in my CS 575 final course project. Partitions corresponded to sensory, motor, and interneuron divisions of the nervous system

Simultaneous with these advancements in connectomics were developments in in-vivo electrophysiology approaches. Research-grade electrode probes (e.g., Neuropixels) have been developed to record the activity of multiple neurons from a single insertion point in in-vivo animal studies [8], [9] and large-scale datasets have been collected for community use [10], [11]. In addition, medical devices have begun to be developed which will enable similar electrophysiology-based recording and stimulation of nervous tissue [12]–[14]. Such devices have already shown promise in treating nervous system conditions ranging from quadriplegia to epilepsy [14]–[16].

As a final and perhaps more obvious recent development, the last several years have shown explosive growth in machine learning research and commercialization, with the transformer model published in 2017 [17] and diverse applications ranging from chatbots to protein structure prediction [18], [19]. Additionally, complementary data science approaches have continued to develop, such as graph data science (e.g., [20]). However, it seems that there may still be a scarcity of application of these modern data scientific approaches to the other neuroscientific fields. The recent and concurrent progress in all these fields (connectomics, electrophysiology, brain computer interfaces, and data science/machine learning), as well as the apparent/potential lack of computational methodologies is processing this data, serves as inspiration for this master's project, and the beginning of my post-master's research portfolio.

Objectives and Deliverables

While making a significant research grade contribution in this space will be beyond the scope of this masters project, I aim to replicate existing scientific methods and refine my understanding of these field. I will do this by pursuing the following objectives and creating associated deliverables. I will work on the most basic objectives first, and complete as many of the objectives as my time constraints for the project allow.

1. (Objective) Understand basic neural signal processing as well as current approaches for electrophysiology data processing and general principles for existing related work in micro-connectomics.
   - (Deliverable) A literature review of works related to the topics introduced in this proposal (2-3 pages).
   - (Deliverable) A revised/forked python package based on the International Brain Lab electrophysiology data processing python package(s), with a jupyter notebook tutorial demonstrating the neural signal processing tools.
2. (Objective) Design, implement, and test 3-5 algorithms for electrophysiology-based identification of connectome subgraphs with International Brain Lab small animal data. Algorithm design may consider strategies such as pairwise statistics or graph convolutional networks.
   - (Deliverable) A set of python classes and/or modules, and a jupyter notebook demonstrating these algorithms.
   - (Deliverable) A report of test results with metrics for time-efficiency and reliability for each algorithm*.
3. (Objective): Develop an agent-based simulation and/or other visualization of the identified connectome subgraphs, potentially including dynamic firing patterns*.
   - (Deliverable) A simulation or visualization (such as with Matplotlib or PyQT) of the identified subgraphs and any associated dynamics*.

*These objectives/deliverables are most ambitious, and may be removed from scope (or iterated on) depending on time constraints of the project.

Plan for Weekly Meetings

Dr. Goodrich and I will plan to meet at his office weekly on Tuesdays at 1:00 PM to review the status of this project, troubleshoot roadblocks, and plan iterative progress related to the project's questions and deliverables during Spring and Summer (2025) terms.

Three-Sentence Summary

The last 20 years have shown significant developments in the fields of connectomics, electrophysiology, brain computer interfaces, and data science/machine learning. It seems that there is tremendous opportunity to use these tools together synergistically and for the mutual benefit of all of the fields. My aim in this project is to better understand the landscape of these fields and to build algorithms to identify connectome subgraphs based on electrophysiology data.

One-Sentence Summary

I am working to understand electrode-based recording of the brain and use it to build a network of neurons and computationally analyze the brain's function.

References

[1] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, "The structure of the nervous system of the nematode caenorhabditis elegans," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 314, no. 1165, pp. 1–340, Nov. 1986, ISSN: 0962-8436. DOI: 10.1098/rstb.1986.0056.

[2] S. W. Emmons, "The beginning of connectomics: A commentary on white et al. (1986) 'the structure of the nervous system of the nematode caenorhabditis elegans'," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, no. 1666, p. 20 140 309, Apr. 2015, PMCID: PMC4360118, ISSN: 1471-2970. DOI: 10.1098/rstb.2014.0309. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4360118/.

[3] DarwinPeacock and Maklaan, *Social network diagram (segment)*, Licensed under CC BY 3.0. A fictional social network diagram consisting of 165 nodes and 1851 edges, auto-generated by script., Dec. 2014. [Online]. Available: https://en.wikipedia.org/wiki/File:Social_Network_Diagram_(segment).svg.

[4] Wikipedia contributors, *Social network analysis*, https://en.wikipedia.org/wiki/Social_network_analysis, Accessed: 2025-05-01, May 2025.

[5] K. Amunts, C. Lepage, L. Borgeat, *et al.*, "Bigbrain: An ultrahigh-resolution 3d human brain model," *Science*, vol. 340, no. 6139, pp. 1472–1475, 2013.

[6] T. J. Sejnowski, "Nanoconnectomics," in *Micro-, Meso- and Macro-Connectomics of the Brain*, H. Kennedy, D. C. V. Essen, and Y. Christen, Eds., Accessed via NCBI Bookshelf, Cham (CH): Springer, Mar. 2016, ch. 1, pp. 1–10. DOI: 10.1007/978-3-319-27777-6_1. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK435767/.

[7] J. S. Elam, M. F. Glasser, M. P. Harms, *et al.*, "The human connectome project: A retrospective," *NeuroImage*, vol. 244, p. 118 543, Dec. 2021, Epub 2021 Sep 8, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2021.118543.

[8] J. J. Jun, N. A. Steinmetz, J. H. Siegle, *et al.*, "Fully integrated silicon probes for high-density recording of neural activity," *Nature*, vol. 551, no. 7679, pp. 232–236, 2017, ISSN: 1476-4687. DOI: 10.1038/nature24636. [Online]. Available: https://doi.org/10.1038/nature24636.

[9] A. C. Paulk, Y. Kfir, A. R. Khanna, *et al.*, "Large-scale neural recordings with single neuron resolution using neuropixels probes in human cortex," *Nature Neuroscience*, vol. 25, pp. 252–263, 2022, ISSN: 1546-1726. DOI: 10.1038/s41593-021-00997-0. [Online]. Available: https://doi.org/10.1038/s41593-021-00997-0.

[10] I. B. Laboratory, "Data release - Brainwide map - Q4 2022," Nov. 2022. DOI: 10.6084/m9.figshare.21400815.v7. [Online]. Available: https://figshare.com/articles/preprint/Data_release_-_Brainwide_map_-_Q4_2022/21400815.

[11] Y. Zhang, Y. Wang, M. Azabou, *et al.*, *Neural encoding and decoding at scale*, 2025. arXiv: 2504.08201 [q-bio.NC]. [Online]. Available: https://arxiv.org/abs/2504.08201.

[12] E. Musk and Neuralink, "An integrated brain-machine interface platform with thousands of channels," *Journal of Medical Internet Research*, vol. 21, no. 10, e16194, Oct. 2019, ISSN: 1438-8871. DOI: 10.2196/16194. [Online]. Available: https://www.jmir.org/2019/10/e16194/.

[13] N. S. Card, M. Wairagkar, C. Iacobacci, *et al.*, "An accurate and rapidly calibrating speech neuroprosthesis," *New England Journal of Medicine*, vol. 391, no. 7, pp. 609–618, 2024. DOI: 10.1056/NEJMoa2314132. eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMoa2314132. [Online]. Available: https://www.nejm.org/doi/full/10.1056/NEJMoa2314132.

[14] M. Vilela and L. R. Hochberg, "Applications of brain-computer interfaces to the control of robotic and prosthetic arms," in *Handbook of Clinical Neurology*, vol. 168, Elsevier, 2020, pp. 87–99. DOI: 10.1016/B978-0-444-63934-9.00008-1.

[15] E. B. Geller, "Responsive neurostimulation: Review of clinical trials and insights into focal epilepsy," *Epilepsy & Behavior*, vol. 88, pp. 11–20, Nov. 2018, ISSN: 1525-5050. DOI: 10.1016/j.yebeh.2018.06.042. [Online]. Available: https://doi.org/10.1016/j.yebeh.2018.06.042.

[16] C. N. Heck, D. King-Stephens, A. D. Massey, *et al.*, "Two-year seizure reduction in adults with medically intractable partial onset epilepsy treated with responsive neurostimulation: Final results of the rns system pivotal trial," *Epilepsia*, vol. 55, no. 3, pp. 432–441, Mar. 2014, Epub 2014 Feb 22. DOI: 10.1111/epi.12534. [Online]. Available: https://doi.org/10.1111/epi.12534.

[17] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, *Attention is all you need*, 2023. arXiv: 1706.03762 [cs.CL]. [Online]. Available: https://arxiv.org/abs/1706.03762.

[18] J. Jumper, R. Evans, A. Pritzel, *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, pp. 583–589, 2021, ISSN: 1476-4687. DOI: 10.1038/s41586-021-03819-2. [Online]. Available: https://doi.org/10.1038/s41586-021-03819-2.

[19] Wikipedia contributors, *Alphafold — alphafold 2 (2020)*, https://en.wikipedia.org/wiki/AlphaFold, Accessed: 2025-05-02, May 2025.

[20] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, *Graph attention networks*, 2018. arXiv: 1710.10903 [stat.ML]. [Online]. Available: https://arxiv.org/abs/1710.10903.