Graph Data Science to Improve Electrophysiology Data Analysis for Connectomics

In this project I aimed to gain an introductory understanding of subfields at the intersection of data science and neuroscience, including neural signal processing, electrophysiology data processing, and general principles in micro-connectomics. I also aimed to understand and replicate existing algorithms to build a cellular-resolution network of a biological neural network. Deliverables are included in the project repository, and here I briefly reflect on highlights and key takeaways from the project.

As I prepared for this project I anticipated the significance of this interdisciplinary space. However, I don't know that I anticipated how significant recent advances have been. Specifically, I found that multiple advanced methods in statistics and data science have been applied to these problems, including approaches from information theory, systems theory, graph theory, and deep learning. I found that there was much more content in the math applied through these methods than I could learn or replicate in this project, and I saw directions that I could work in to build on this project in the future. At one point in the project Dr. Goodrich and I had a conversation about how the more we learn the more we realize we don't know–so we should seek to become more humble as we grow. I found that as a meaningful soft skill I plan to carry through my career.

In another conversation with Dr. Goodrich, he recommended that I emphasize depth-first over breadth-first research, meaning that as I see sizeable scope for potential research projects, I should focus a niche and lean on others' work. As one who is naturally inclined to understand details, this is an important learning for me. I began to develop discretion for details that were important for me to understand and details where I could rely on others' specialties. This was especially relevant for the more recent signal-processing and spike analysis methods I studies, where I could see that the methods were built on many previously characterized concepts.

In preparation for this project I worked with Dr. Goodrich in CS 575 to build a relevant foundation of network science. As we did, I began to find that the biological network in neural tissue was organized in a pattern that could be described with data science metrics. The network produced in this project validated that prior learning. Specifically, I found that biological neuronal networks have strong core-periphery structure, and possibly satisfy the small-world property (more samples are needed to validate), but that within a given anatomical structure there is little organization by centrality or community.

This project was an excellent opportunity for me to build my background in a topic I am passionate about. I look forward to building on these technical, practical, and soft skill learnings with other opportunities in the future.

Hours

Category	Time (HH:MM)
Study projects, methods, and code; adapt code for my use	47:30
Study publications and write literature review	38:00
Build Demo of IBL Data	12:45
Study and replicate spike train analysis algorithms	18:45
Maintain repository and environment	9:15
Review status periodically	4:15
Produce report and presentation	3:30
Total	**134:00**