CS 575 Project 2: Complex Contagion

J. Wesley Borden

Mar 11, 2024

Abstract

This project continues network modeling of simulated neurons, with each neuron modeled as a finite state machine, and the eventual goal of modeling realistic mammalian biological neuronal networks using graph data science. Primary contributions of the project include (1) restructuring the neuron FSM to include a separate state for the axon hillock and terminal, with a timer to virtualize the propagation time along the axon and (2) modeling complex contagion with absolute and relative thresholds. The primary outcome was presence of a self-sustaining propagation of signal through the network. Results show that such self-sustaining activation was always possible when the thresholds were low enough, but that many factors may contribute to the self-sustaining activation otherwise, including degree distribution (e.g., the scale free property), density, modularity, and diameter.

Introduction

Between 1950 and 1986, a team of scientists in Europe, led by Dr. Sydney Brenner and John White, initiated a groundbreaking effort to map the neural circuitry of the nematode Caenorhabditis elegans (Original Project, Review). This project, informally known as "The Mind of the Worm" laid the conceptual and technical foundation for the modern field of connectomics. It is now well established that higher-order cognitive and behavioral functions in mammalian brains emerge from the complex interactions within large-scale biological networks of neurons. Recent advances in connectomics have enabled large-scale network modeling of human brain function using non-invasive imaging techniques such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) (e.g., the Human Connectome Project). However, these methods lack the resolution to resolve neuronal connectivity at the single-cell level.

The advent of brain-computer interfaces (BCIs) has introduced new methodologies capable of recording and interfacing with neural circuits at unprecedented resolution (e.g., BrainGate). These technologies hold the potential to not only advance our understanding of the human brain but also to enable transformative therapeutic approaches for neurological and psychiatric disorders. Moreover, by elucidating the underlying principles of biological neural computation, these efforts may inspire new paradigms in machine intelligence, fostering biologically informed advances in artificial intelligence and neural network design (Example).

In this report, I detail my second iteration of a simulation of biological neuronal networks. Neurons are modeled as a combination of two finite state machines (FSMs): one for the axon hillock, where incoming signals are considered to be combined and processed, and one for the axon terminal, where signals can be transmitted to other neurons. A timer between the FSMs substitutes for the axon, where signals are sent in one direction from hillock to terminal. Simulations for several networks were modeled, and characteristics of the simulation were analyzed in context of input network metrics.

Hypotheses

In review of the first iteration of this simulation, Dr. Goodrich hypothesized that:

- "Self sustaining activation is a function of degree distribution and diameter"

I considered this hypothesis and added another of my own:

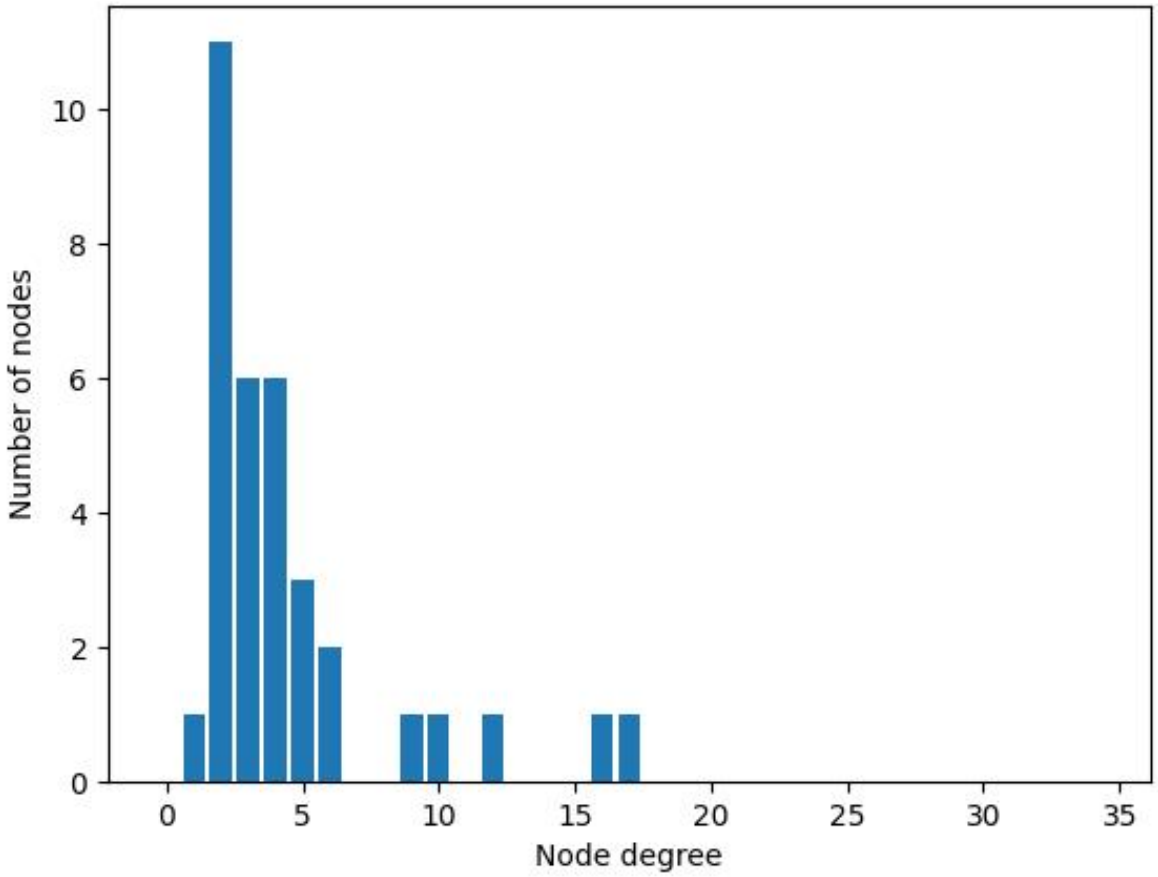
- "Self sustaining activity is promoted by strong modularity in sufficiently large networks, but is inhibited by the small world property"

In retrospect, I now recognize that graphs do not have modularity. Partitions have modularity. In my experiments I defined "Louvain Modularity" as the modularity of highest-modularity partition found with the Louvain algorithm.

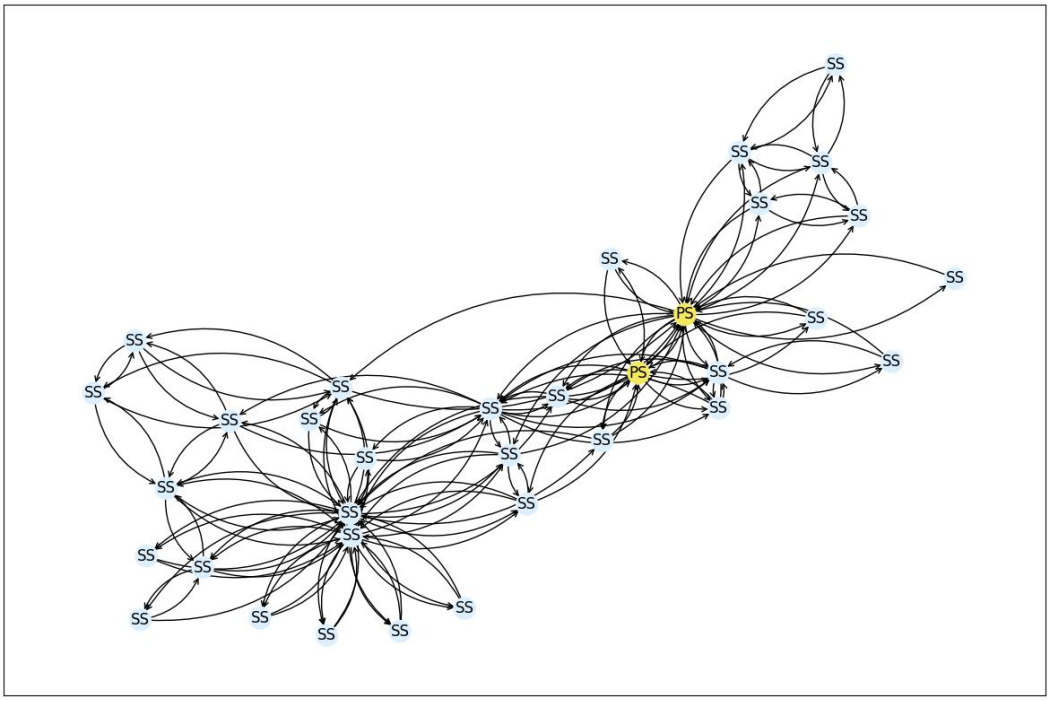
Discussion

I performed pairwise tests of networks described in the project specifications, each varied by which agents were initially activated. Each network simulation was tested with absolute or relative thresholds of 2, 3, 4, 10%, 20%, and 30% of neighboring agents. Initial and final images of each simulation were saved, and agents in each state over time were plotted. Degree distribution was also plotted for each input graph. For example, for the karate graph, with agents 1 and 2 initially activated, and a relative threshold of 10%, the following images were generated:

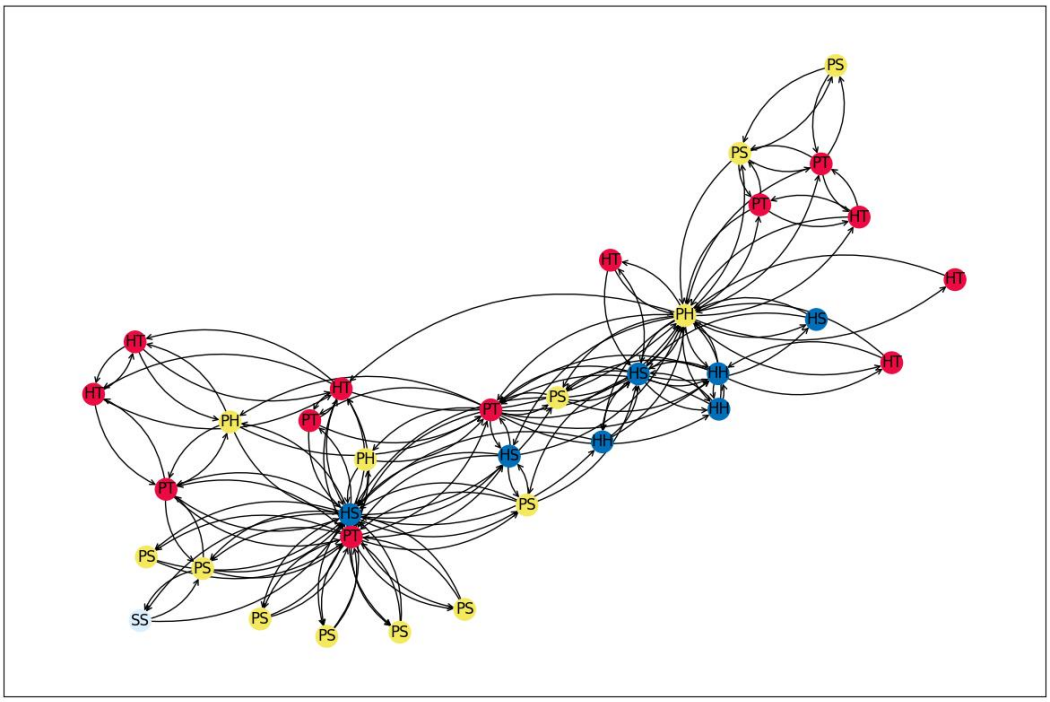
Degree distribution:



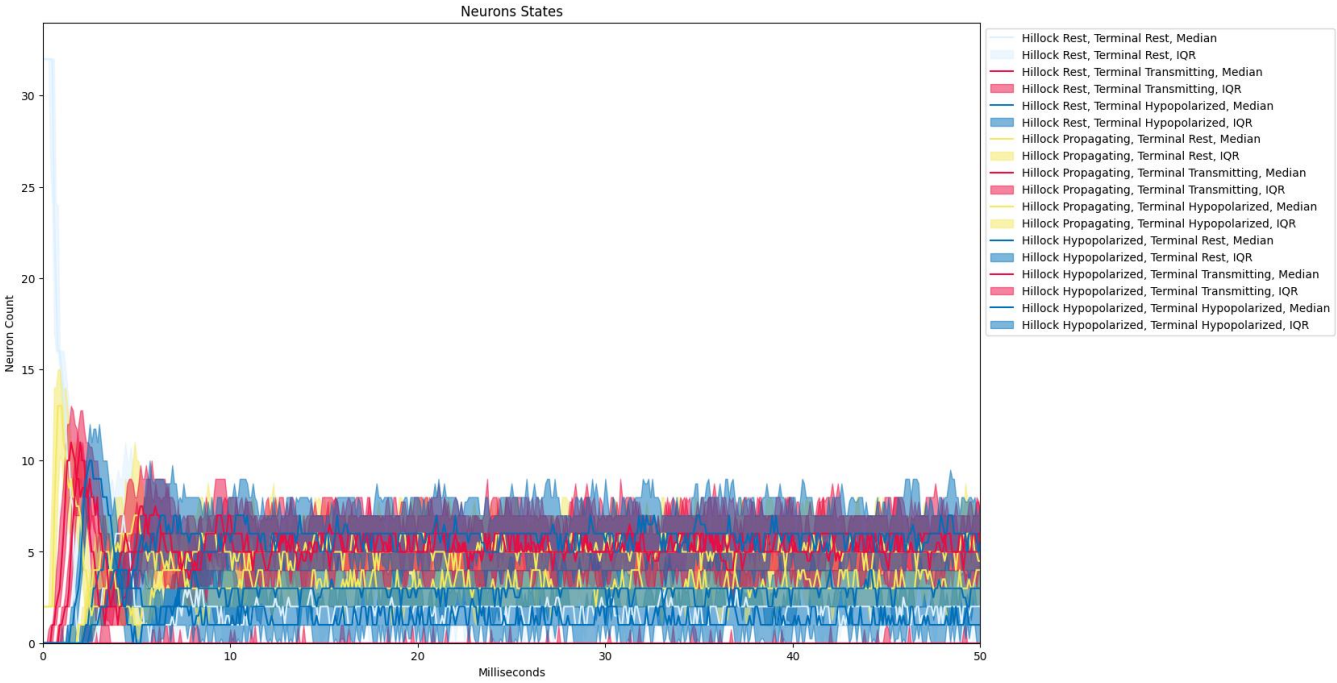
Initial State:



Final State:

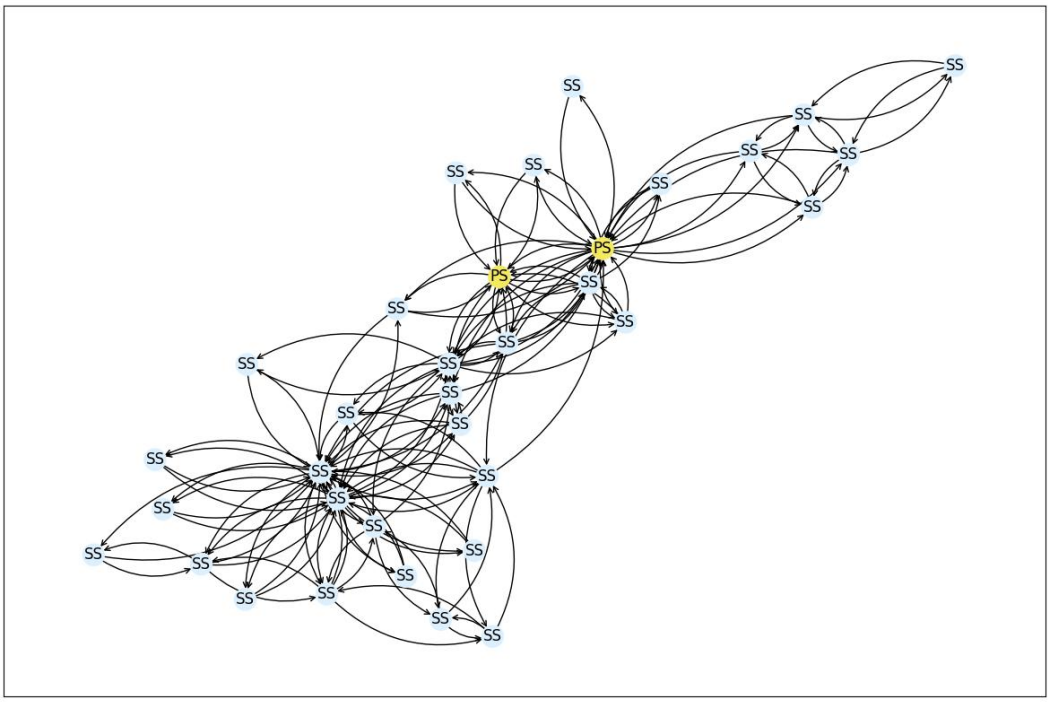


States over time:

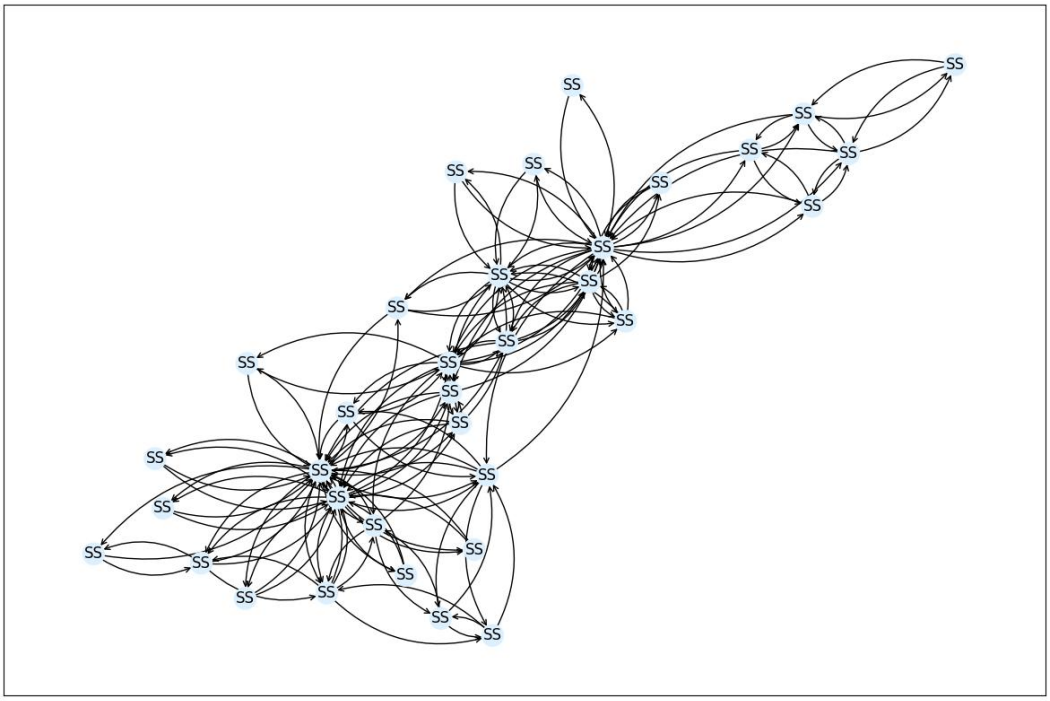


The above is an example of self-sustaining activation, because the signal continues to propagate through the network over time. In my prior iteration, I found that some networks were not able to maintain self-sustaining activation. For example, for the same karate graph with agents 1 and 2 initially activated, but an absolute threshold of 4 agents, the signal was not self sustaining:

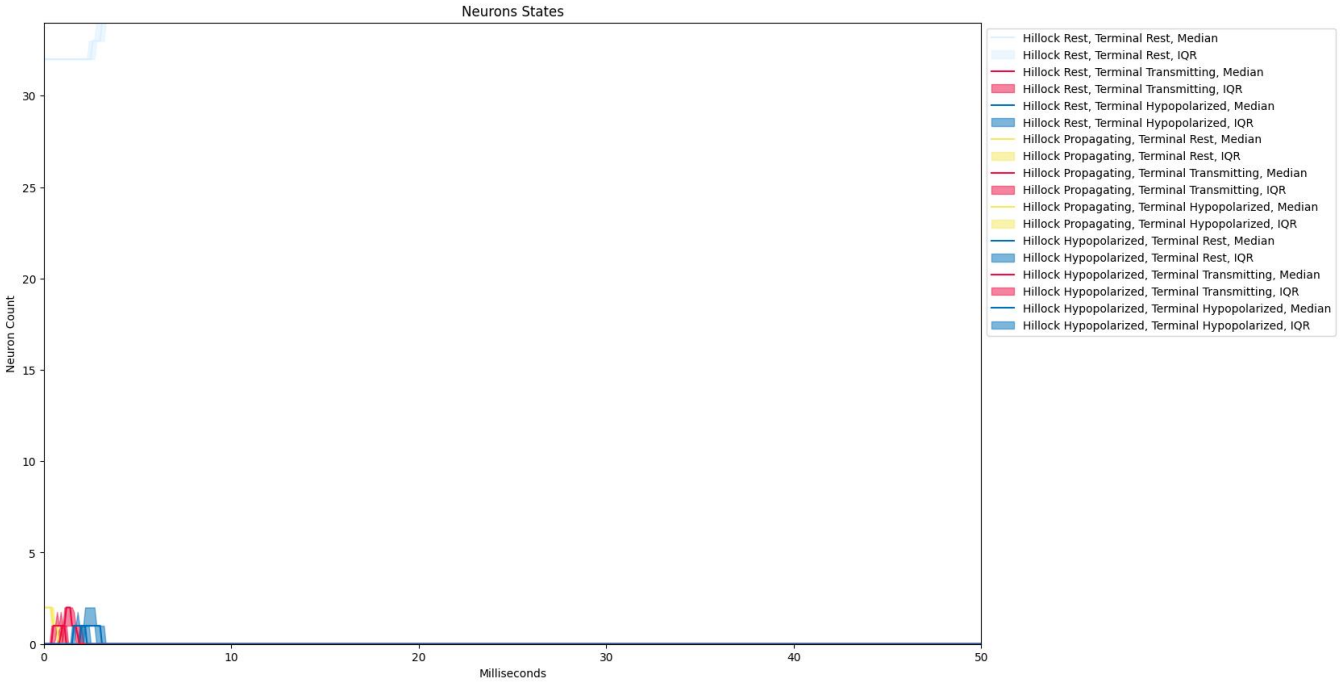
Initial State:



Final State:



States over time:



This totaled to 138 experiments. I organized the results on Microsoft Excel as follows:

run_name	graph	initial_nodes	maximum_degree	average_degree	diameter	radius	density	louvain_modularity	clustering_coeff	average_degree*dia	Deg. Dist	Sust_Abs_2	Sust_Abs_3	Sust_Abs_4	Sust_Rel_10	Sust_Rel_20	Sust_Rel_30
small_world_410_20-random-nodes	small_world_410	random 20	5.00	2.00	61.00	31.00	0.00	0.90	0.000	122.00	small world	fail	fail	fail	sustained	sustained	sustained
small_world_410_20-highest-degree	small_world_410	highest 20	5.00	2.00	61.00	31.00	0.00	0.90	0.000	122.00	small world	fail	fail	fail	sustained	sustained	sustained
dublin_50-random-nodes	dublin	random 50	50.00	13.49	9.00	5.00	0.03	0.71	0.456	121.39	right skewed bell curve	sustained	sustained	sustained	sustained	sustained	fail
dublin_50-highest-degree	dublin	highest 50	50.00	13.49	9.00	5.00	0.03	0.71	0.456	121.39	right skewed bell curve	sustained	sustained	sustained	sustained	sustained	fail
small_world_100_4-random-nodes	small_world_100	random 4	8.00	4.00	7.00	5.00	0.04	0.59	0.195	28.00	small world	fail	fail	fail	sustained	sustained	fail
small_world_100_4-highest-degree	small_world_100	highest 4	8.00	4.00	7.00	5.00	0.04	0.59	0.195	28.00	small world	fail	fail	fail	sustained	sustained	fail
small_world_100_4-lowest-degree	small_world_100	lowest 4	8.00	4.00	7.00	5.00	0.04	0.58	0.195	28.00	small world	fail	fail	fail	sustained	sustained	fail
scale_free_410_20-highest-degree	scale_free_410	highest 20	60.00	3.98	6.00	4.00	0.01	0.50	0.038	23.88	scale free	sustained	fail	fail	sustained	sustained	sustained
scale_free_410_50-random-nodes	scale_free_410	random 50	60.00	3.98	6.00	4.00	0.01	0.50	0.038	23.88	scale free	sustained	fail	fail	sustained	sustained	sustained
scale_free_410_50-highest-degree	scale_free_410	highest 50	60.00	3.98	6.00	4.00	0.01	0.50	0.038	23.88	scale free	sustained	fail	fail	sustained	sustained	sustained
figure_19-4_1-2	figure_19-4	1, 2	5.00	3.41	7.00	4.00	0.21	0.46	0.463	23.88	bell curve	fail	fail	fail	sustained	sustained	sustained
figure_19-4_6-7-12	figure_19-4	6, 7, 12	5.00	3.41	7.00	4.00	0.21	0.46	0.463	23.88	bell curve	fail	fail	fail	sustained	sustained	sustained
figure_19-4_7-8	figure_19-4	7, 8	5.00	3.41	7.00	4.00	0.21	0.46	0.463	23.88	bell curve	fail	fail	fail	sustained	sustained	sustained
scale_free_100_4-lowest-degree	scale_free_100	lowest 4	25.00	3.92	6.00	4.00	0.04	0.46	0.121	23.52	scale free	sustained	fail	fail	sustained	sustained	fail
scale_free_100_4-highest-degree	scale_free_100	highest 4	25.00	3.92	6.00	4.00	0.04	0.46	0.121	23.52	scale free	sustained	fail	fail	sustained	sustained	fail
scale_free_100_4-random-nodes	scale_free_100	random 4	25.00	3.92	6.00	4.00	0.04	0.45	0.121	23.52	scale free	fail	fail	fail	sustained	sustained	fail
circulant_20-12_1-2	circulant_20-12	1, 2	4.00	4.00	5.00	5.00	0.21	0.44	0.500	20.00	all 4	fail	fail	fail	sustained	sustained	fail
circulant_20-12_1-10	circulant_20-12	1, 10	4.00	4.00	5.00	5.00	0.21	0.44	0.500	20.00	all 4	fail	fail	fail	sustained	sustained	fail
karate_1-2	karate	1, 2	17.00	4.59	5.00	3.00	0.14	0.42	0.571	22.94	right skewed bell curve	sustained	fail	fail	sustained	sustained	fail
karate_33-34	karate	33, 34	17.00	4.59	5.00	3.00	0.14	0.42	0.571	22.94	right skewed bell curve	sustained	fail	fail	sustained	sustained	fail
karate_2-random-nodes	karate	random 2	17.00	4.59	5.00	3.00	0.14	0.42	0.571	22.94	right skewed bell curve	sustained	fail	fail	sustained	sustained	fail
circulant_20-1234_1-2	circulant_20-1234	1, 2	8.00	8.00	3.00	3.00	0.42	0.29	0.643	24.00	all 8	sustained	fail	fail	sustained	sustained	fail
circulant_20-1234_1-2-3-4	circulant_20-1234	1, 2, 3, 4	8.00	8.00	3.00	3.00	0.42	0.29	0.643	24.00	all 8	sustained	fail	fail	sustained	sustained	fail

Based on these results, and in consideration of our initial hypotheses, I have the following conclusions:

- Most importantly, if the activation threshold is sufficiently low, the network will always be self-sustaining. This is demonstrated where a relative threshold of 10-20% always resulted in self-sustaining activation, regardless of other conditions.
- In testing varied initially-active agents, I found that which agents were initially active had no effect on whether the signal would propagate to equilibrium. The only exception to this was the case where initially activating four random nodes in a scale-free network, with an activation threshold of two neighbors, was insufficient to reach a steady state. In that case, initially activating the four highest or lowest degree neighbors was sufficient to reach equilibrium. I anticipate this was due to the highest and lowest degree nodes being located in ideal locations to affect the rest of the network. This also demonstrates that the problem of maintaining a steady state is often a problem of reaching sufficient activation across the entire network.
- I cannot find an association between graph diameter, density, clustering coefficient, or modularity and self-sustaining activation. I anticipate that this is not because the metrics are not related, but rather that the underlying relationship is more nuanced. Future work may attempt to elucidate these complex relationships.
- The small-world property was found to inhibit sustained activation in most cases, while right-skewed degree distributions (scale free or not; e.g., karate, dublin) were associated with sustained activation. This is consistent with our class's consideration of these properties and their impact on complex contagion.

Future Work

Reflecting on this phase of model development, it is evident that self-sustaining activation arises as a complex emergent property influenced by several graph-theoretic attributes. Preliminary observations suggest a relationship with metrics such as clustering coefficient, graph diameter, density, modularity, and degree distribution, though the precise interplay among these factors remains nuanced. Future work could leverage machine learning approaches, such as graph neural networks (GNNs), to more effectively classify or quantify a graph's propensity toward self-sustaining activation. Additionally, insights derived from machine learning models could provide a more mechanistic understanding of how high-level network properties, such as clustering and density, correlate with activation dynamics.

Another important avenue for improvement is the calibration of hyperparameters governing state transition dynamics within the finite state machines (FSMs). At present, transition durations are sampled from Gaussian distributions with rough estimates for means and standard deviations. However, incorporating real-world electrophysiological data to inform these distributions would substantially enhance the biological plausibility of the model. This could reduce the gap between simulated and observed neural dynamics.

A further refinement involves adopting more biologically realistic network topologies. The current model employs undirected graphs, which assume bidirectional communication between nodes. Extending the model to directed graphs would better capture the asymmetric nature of synaptic transmission. Additionally, exploring multigraph structures, where edges can represent distinct classes of neurons or synapse types, could yield richer and more realistic activation dynamics. Incorporating these structural complexities would also facilitate the modeling of heterogeneous cell populations and diverse synaptic interactions, leading to additional insights.

LLM Acknowledgement

This report was generated using LaTeX, adapted from prior LaTeXfiles which had been written with the help of LLM's such as ChatGPT for formatting, etc. ChatGPT provided LaTeXformatting advice to questions such as "could you help me fix the format of this snippet? ...[LaTeXsnippet] ... This csv has this content: ...[csv file contents] ... And I'm getting the error ... [TeXcompiler error]". LLM's (ChatGPT, Google Search AI Overview) may have provided information for the development of code during the project (e.g., to answer questions about function call syntax), but did not directly write the code. Finally, ChatGPT was used to refine the report's tone, directly updating the text of the report as specified in the project description. My work reflects my best efforts at learning the content and building a promising simulation while leveraging the LLM's for things like productivity, speed, and technical writing tone.