Course Title: Graph Theory for Biological Network Analysis

Note: The first offering of this course will be Winter 2022, not Fall 2021.

Eventually this should be a “real” course, but since Fall 2021 is the first offering, the below is *approximate*. Evaluation (ie., how grading will occur) is also up in the air, but is likely to involve about 3-5 assignments, with a combination of graph theory, programming, and data analysis of real-world data, and a larger final project with some flexibility of what that involves.

Proposed Syllabus:

- Bio 101 for Computer Scientists
    - genes and proteins; protein function + interactions; relation to disease
    - Cellular process: genes -> transcription -> mRNA ->translation -> folded protein
    - effect of disease on protein interactions; virus proteins; how drugs help at the molecular level; the drug-disease-protein-gene network.
    - Types of biological networks
        - gene-gene interactions
        - gene regulation
        - DNA-DNA interactions
        - protein-protein interactions
        - drug-disease interactions
        - Genome Wide Association Studies (GWAS)
        - any combination of the above
    - The missing data problem
        - Lab methods for direct + indirect detection of protein-protein interactions
        - noise in biological networks
        - false positives + false negative edges
        - detection of missing edges (“edge prediction”)
    - computation is cheaper than wet lab work
    - sometimes near-trivial applications of graph theory can uncover relationships that are elusive to detect in the lab (the L3 story)

- Graph theory review
    - types of graphs: directed, undirected, weighted
    - using node + edge colors to depict multiple data types simultaneously
    - review of basic algorithms: BFS, DFS, Dijkstra

- Recent advances that have aided analysis of biological networks:
    - graphlets; network motifs; local + global network alignment
    - statistical modeling of biological networks:
        - Erdos-Renyi; Small-world; Geometric; Scale-free

- Detecting missing data (edge prediction) in biological networks
    - relation to graphlets + network motifs

    - recent results

- Cross-species comparison of biological networks
    - “model” organisms: fruit flies, mice, rats
    - evolutionary relationships + divergence
    - network alignment as “subgraph isomorphism with noise”
    - local network alignment
        - detecting + comparing protein complexes across species
        - detecting similar function of proteins across species
        - protein function, sequence similarity, network similarity
    - global network alignment
        - separating similar protein complexes across the network by global placement
    - partial review of existing methods
    - separating search algorithms + objective functions in optimization
    - importance of network topology + choosing the right measure of topological similarity
    - comparison of topological measures of similarity (both theoretical + empirical)
    - Prediction of protein function via network analysis + alignment