

CS 6402 – Advanced Topics in Data Mining

More Project Ideas

Note: This list includes some topics that we will be discussing in lectures very soon, namely social networks, recommendation systems, and anomaly detection in graphs.

Software Projects

- gBoost: fix the Matlab/C version posted at <https://github.com/rkwitt/gboost> or re-implement it in a different language
- Discriminative subgraph mining: implement it using distributed or parallelized computing; note that Dr. Leopold has already implemented it in Python (so doing that wouldn't be novel)
- Recommendation system: develop a recommendation system (possibly in combination with a social network) and analyze it
- Anomaly detection: implement an anomaly detection algorithm and test it on a graph dataset

For any software project that you do, make sure you test it using a **variety of datasets**. Here, variety means size (# vertices and # edges), vertex degree, and # of graphs (in a transaction graph setting).

Research Papers

- Discriminative subgraph mining: Dr. Leopold has a Python implementation she will gladly make available to you; you could find a dataset for which you can classify the graphs as positive vs. negative (for example), and do an analysis to find discriminative patterns
- Cliques, trusses, k-cores, representative frequent maximal subgraphs, etc.: write a program to find these patterns in graphs (shouldn't be too hard given the networkx library in Python); find a dataset to look for these patterns and analyze the results
- Motif finding: use one of the motif-finding programs on a non-bio/chemo dataset (that hasn't been analyzed that way); look for motifs and analyze the results
- Social network analysis: find some network dataset that hasn't been analyzed before (or develop one) and apply various metrics to analyze it; depending on the amount of development you do, this might qualify as more of a software project and less of a research paper
- A survey on a topic in graph data mining that we're not going to cover this semester:
 - Deep Learning for Graphs (e.g., Convolutional Networks, Neural Networks)
 - Knowledge Graphs (e.g., semantic searching facilitated by graphs)
 - Communication/Wireless/Sensor Networks

- Social Networks and Disease Transmission (even though we're going to talk about social networks, this will be allowed since disease transmission is such a relevant topic these days!)

Warning: Before you propose something, do a search for it on the internet to see whether it has already been done. Dr. Leopold will not approve projects that are not novel!