Project Title: The Impact of Quality Metrics on Communities Detected in Complex Networks
Student Names: Jennifer Nguyen (Pomona College '17), Christina Tong (Pomona College '17), Anastasia Voloshinov (Pomona College '17)
Project Mentor: Professor Tzu-Yi Chen (Pomona College)

Goals / Purpose of the Project

Networks (also known as graphs) can be used to model real-world systems, where nodes represent entities of the system and edges show the interactions between entities. A group of vertices that are more connected to each other than to others are called communities, or clusters. In many applications, we are interested in identifying communities within these networks based solely on the interactions observed. Quality metrics are used to determine how well-formed these communities are, with good metric scores generally reflecting networks where connections are more dense within communities than between communities.

The majority of commonly-used algorithms for detecting communities optimize a metric known as modularity, but other metrics such as coverage, performance, and silhouette index could also be used within those same algorithms. While existing literature focuses mainly on comparing the quality metrics' evaluations of already computed communities, our research implements these metrics within community-detection algorithms, thus affecting the actual process in which communities are created. The goal of this research is to examine the impact of replacing modularity with other quality metrics in two commonly-used community detection algorithms: the Louvain method and the Clauset-Newman-Moore (CNM) algorithm. Do these algorithms output consistently different clusterings when using the other quality metrics and, if so, in what ways are the clusterings different?

Process Used in Completing the Research

We began the year by getting familiar with existing research in complex networks. With the help of Professor Chen, we identified algorithms and metrics to explore. We did a lot of background reading and research on complex networks, community detection, evaluation metrics, and the strengths and weaknesses of the various metrics. Additionally, we collected the graphs for our test suite from places like the Stanford Network Analysis Project (SNAP). After several months, we settled on a broad research question: Do the Louvain and CNM algorithms output consistently different clusterings when using the other quality metrics, and in what way(s) are the clusterings different?

We operated on a "domain expert" model, assigning each metric to a team member for deeper investigation and implementation into the Louvain algorithm. We also assigned each of the algorithms to a team member for deeper investigation. We began by using the C++ implementation of the Louvain algorithm. We designed high-level implementations of each our metrics, followed by pseudo-code. However, after further consideration, we switched to the Java implementation of the Louvain method, which had a better abstraction of serialization than the C++ implementation. The clusterings obtained from the C++ implementation could still be compared to the output clusterings from the Java implementation because we only changed the implementation and not the method itself. By the end of winter break, we completed implementations of Louvain with coverage, Louvain with performance, and Louvain with silhouette, overcoming various difficulties including peculiarities specific to each metric and reading and writing the input and output.

During the second semester, we ran our modified algorithms (Louvain-performance, Louvain-silhouette, Louvain-coverage, CNM-coverage) and the original algorithms (Louvain-modularity, CNM-modularity) over our suite of graphs. The graphs without ground-truths, or graphs without already known clusterings, were generally much larger, ranging from 1,589 to 40,421 vertices, and 5,484 to 351,382 edges. As a result, we grouped them into families in order to better detect patterns in the output clusterings.

We selected the statistical "correctness" metrics that we would use to compare the output clusterings against each other; these metrics essentially compare the difference(s) between two clusterings. The metrics are: normalized mutual information, variation of information index, split-join distance, Rand index, and adjusted Rand index. Using these correctness metrics, we evaluated the output clusterings from graphs with ground-truths against the ground-truths themselves.

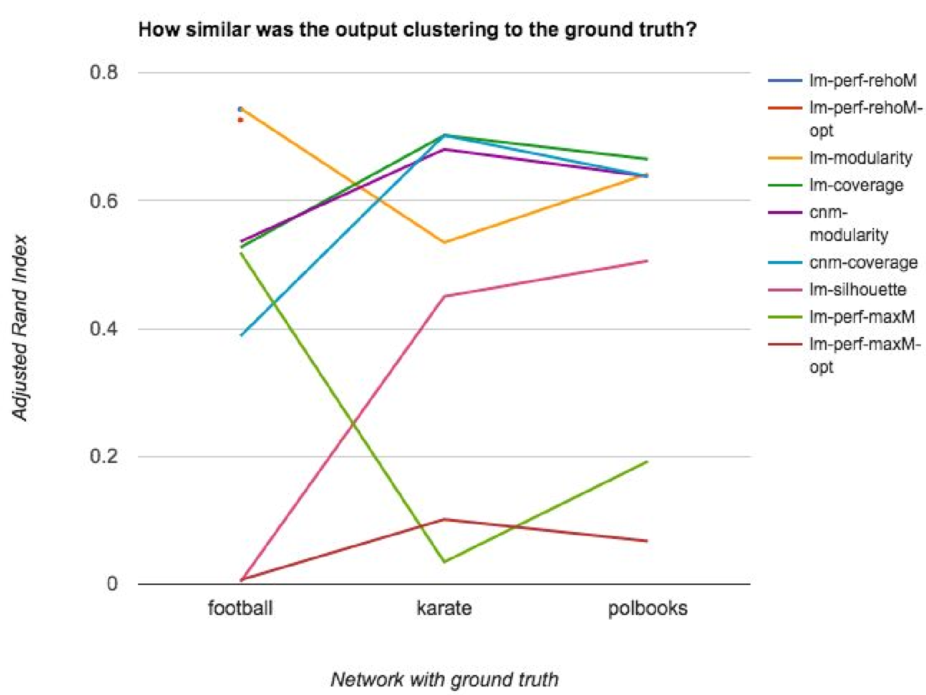
In March, our poster was accepted to the 2016 Tapia conference, and we integrated feedback from reviewers into our ongoing work, and in April, we presented at the Southern California Celebration of Women in Computing and were again able to integrate feedback into our future poster proposals and research.

Towards the end of the semester we finished up data collection. Using the correctness metrics, we evaluated the output clusterings from graphs without ground-truths. Additionally, we ran more validation testing on the output graphs, to ensure that each algorithm variation actually outputs the clustering that maximized the respective metric.

During the summer, we began consolidating our results into a comprehensive paper, which we intend to submit to a conference.

Conclusions and Results Achieved

The graph below visualizes the accuracy of the algorithm variations' output clusterings, as measured by the adjusted Rand index, for all the networks with ground-truths. Scores range from 0 to 1, where scores closer to 1 indicate output clusterings that are more similar to the ground truths.



The lines demonstrate the success (or lack thereof) of the various algorithms in reproducing the ground-truth communities, as well as their performance relative to each other. After calculating the adjusted Rand index from these three ground-truth graphs, we can see the relative performance of the algorithm variations. There are cases where the new algorithm variations produced clusterings closer to the ground truth than the commonly-used Louvain-modularity and CNM-modularity algorithms. For example, the Louvain-coverage variation outperformed all other algorithms on the karate and polbooks networks. However, some of the new algorithm variations, such as Louvain-silhouette, produced clusterings that were substantially less similar to the ground truths than Louvain-modularity and CNM-modularity. These conclusions can be generalized to the other correctness metrics.

We are currently working on the analysis of networks that do not have ground truths. We hope to identify trends within and between the network families.

Web Pages Developed

We used a blog to publish updates on our research.

Publications and Presentations

- *The Impact of Quality Metrics on Communities Detected in Complex Networks,* Jennifer Nguyen, Christina Tong, and Anastasia Voloshinov. Presented in student talk session at the Southern California Celebration of Women in Computing, an ACM-W event. Ventura, CA. April 2016
- *The Impact of Quality Metrics on Communities Detected in Complex Networks,* Jennifer Nguyen, Christina Tong, and Anastasia Voloshinov. Poster to be presented at the ACM Richard Tapia Celebration of Diversity in Computing. Austin, TX. September 2016. (http://tapiaconference.org/schedule/thursday-september-15-2016/630pm-800pm/poster-session-reception-2/2016-student-posters/complex-networks)
- *The Impact of Quality Metrics on Communities Detected in Complex Networks,* Jennifer Nguyen, Christina Tong, and Anastasia Voloshinov. Poster to be presented at the Grace Hopper Celebration of Women in Computing. Houston, TX. October 2016.