

CSoI Annual Progress Report - The Polarization of Information

Charles Dickens, Pamela Bilo Thomas, Ram Hari Dahal, Prajjwal Dangal

August 2019

General Team Updates

During the fall semester some of the team members needing to pull back their efforts in order to focus on other matters. The team did gain one student who showed interest in the project and made excellent contributions. As a result of the changes, meetings were less frequent and our mid-year goal of presenting at ASONAM 2019 was not met.

Team Communications

For daily communications, the team continued to utilize the collaboration tool, *Slack*. As for slightly more formal updates *Zoom*, a video communications platform, was used. During these meetings we addressed the progress that was made and planned the next steps.

For project management and source control, resources (including posters, papers, and presentation materials) and code for the project are hosted on GitHub repositories under the Polarization of Information organization. GitHub supports project boards, which our teams used to plan and assign tasks.

Progress Towards Results

The team has implemented a semi-supervised approach to modeling a network of information sources. Twitter granted the team access to the premium search API sandbox for this project, which was a great test-bed for our algorithm.

In Spring 2019, the team focused on scaling the analysis to include information outside of Twitter. To do this, a workflow for constructing a network model for the news articles was needed.

Charles worked with Andrew Millard, an undergraduate electrical engineering student at UH Manoa, to develop a workflow for collecting and grouping news articles related to current topics. The two constructed a pipeline for periodically populating a MySQL database with normalized tables designed to scale for functionality on the teams road-map. Furthermore, during the production of the system, the team found it necessary to create a new fully automated similarity metric that would ultimately work between both new articles and Tweets.

The similarity metric ultimately became the biggest development during the Spring of 2019. This algorithm iteratively creates a new network model of the "documents" using the mutual information between the words and documents as a similarity measurement, clusters the documents, combines the clustered documents, and repeats until a convergence condition is met. The mutual information between documents is found by finding a joint distribution over the documents using a Blahut-Arimoto algorithm. Simultaneously, words are filtered from the documents using the conditional mutual information. This method shows promise and resulted in clusters of articles that matched intuition. The code for this algorithm can be found on GitHub repositories under the Polarization of Information organization.

Presentations and Posters

An updated paper and poster for the The Polarization of Information project has been presented at the UH Manoa College of Engineering Spring 496 Poster Session. Along with the poster, a brief slide show presentation of the results of the project as of May 2019 was made.

Remaining Balance and Intent

The team did not spend any of the \$6,000 balance rewarded and unfortunately does not plan on continuing this research in the near future and thus does not intend to spend the money in any way.