

Nathaniel Miller

Computer Science Honors Project Proposal

Title: Finding New Ways to Measure Social Polarization in Wikipedia

Faculty Sponsor: Shilad Sen

Description:

Utilizing the software library and dataset provided by Colin Welch's Honors Project, *Measuring Balkanization in Wikipedia* (http://digitalcommons.macalester.edu/mathcs_honors/20/), I propose an investigation of social polarization in Wikipedia and new ways to measure it. In order to do so, I will first research social polarization and current social polarization metrics. Additionally, in preparation for my later work I will make improvements to the provided software library as necessary. I will then develop new metrics for measuring social polarization and create algorithms to implement them efficiently. As the data set represents Wikipedia as a graph, I anticipate that this will involve large amounts of graph theory. Finally, I will use the metrics to analyze Wikipedia, possibly focusing on certain sections: language editions, periods of time, or articles within a certain group, for example.

Timeline:

Summer:

Research social polarization and current metrics.

September:

Continue researching social polarization and current metrics as necessary. Familiarize self with current software library and Wikipedia dataset. Begin developing new metrics of social polarization.

October:

Make improvements to software library as necessary. Continue developing new metrics. Begin to create algorithms to enable the efficient calculation of these metrics.

November:

Continue development of metrics and algorithms. Begin testing algorithms and metrics on Wikipedia data set. Write fall semester paper. Develop talk for department seminar.

December:

Continue implementing algorithms for metric calculation. Refine metrics if necessary. Continue to test algorithms and metrics.

January:

Refine algorithms and metrics as necessary. Continue analyzing Wikipedia and selected subgraphs with metrics. Complete progress report.

February:

Finish analyzing Wikipedia and selected subgraphs with metrics. Begin writing Honors paper. Begin preparing Honors presentation.

March:

Tweak analysis of Wikipedia and subgraphs as necessary. Finish paper by Spring break. Finish preparing presentation.