Thesis Project Portfolio

Predicting Comment Popularity within Online Communities Using Multiclass Classification

(Technical Report)

Unification of Sub-communities within Geographic Communities

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science in Computer Science, School of Engineering

Cory Junghoon Kim

Spring, 2020

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

Inter-community and intra-community dynamics have long been loosely understood. Various attempts, by both technical and sociological fields of study, have been developed in hopes of understanding such community dynamics. Community structure has been understood, where society often follows predictable structural trends over time. However, interactions between partitions and subgroups within and between communities have proven to be much more difficult to understand. Studying evolutionary community dynamics is a difficult task, often being limited to the historical data, with an insurmountable number of variables to consider.

The classification of a group as a community is often best described as an abstraction. For example, one might consider a city a community. However, within said city, there may exist many different neighborhoods with different ideologies, each with complex relationships with one another. These neighborhoods each could be considered their own community, all smaller groups of a larger geographic community. This holds true at a larger scale, where a city could be a part of an even larger geographic community, consisting of several cities.

My STS thesis digs deeper in the structural dynamics of communities and why it is applicable. It takes a closer look at problems faced by large geographic regions, or communities, and how these problems evolve and manifest in relation to the interactions of the regions' sub-communities. Considering the smaller partitions of large geographic communities is important, as daunting geographic problems faced by these communities are nearly impossible to tackle without the synchronization and cooperation of all sub-communities within said larger community. Examples of this can be found worldwide, ranging from the stormwater management problems in Albemarle county to the emissions standards being enforced throughout China.

While my STS analysis dives into the existing historical literature and sources from a more sociological point of view, my technical topic takes a more head-on approach. With the advent of modern computing, large communities consisting of smaller sub-communities may be simulated and analyzed. Researchers have gone as far as representing communities with social actors and programmed interactions in the form of networks. However, my technical thesis takes advantage of a more realistic, but also digital, representation of communities: online discussion boards.

Provided the boom of the internet and its widespread influence today, many communities have gone digital. Many new communities have been formed, connecting physically-separated individuals online based on shared interests. Simulating evolutionary community dynamics with a realistic model is a problem that is far from solved. However, we must start somewhere. My technical study aims to develop better ways of understanding online communities, through understanding large stores of conveniently accessible data on internet discussion boards. Using this data, models are trained to predict comment popularity based on certain metrics, ranging from body text to simple metadata. These prediction models aid in understanding these online communities, as the ability to predict popularity within communities translates well to understanding what these communities value.