Understanding the Evolution of Online Community Interactions

(Technical Paper)

Unification of Sub-communities within Geographic Communities

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree

Bachelor of Science in Computer Science, School of Engineering

Cory Junghoon Kim

Spring, 2020

Technical Project Team Members

Siddharth Nanda

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature  Date 4/29/20

Cory Junghoon Kim

Approved _____ Date _____

N. Rich Nguyen, Ph.D., Department of Computer Science

Approved _____ Date _____

Sean Ferguson, Ph.D., Department of Engineering and Society

I. Introduction

Understanding the dynamics of community interactions has always been sought to be understood, where various attempts like Actor Network Theory only serve to bring about technosocial insights. With the advent of modern computing, theoretical community models have been created and further analyzed. Researchers have already achieved software-implemented representations of communities consisting of social actors and interactions using multi-mode networks in the form of graph networks.

For my technical topic, evolutionary dynamics of online multiple communities are, from a theoretical and implementation perspective, explored. An example of where one might find a collection of said communities is on a website, conveniently founded by a University of Virginia alumni, *reddit.com*. Studying the evolutionary dynamics of online communities is highly applicable to better understanding how real communities and societies evolve and interact. Using an online medium is convenient in that the data is easy to access and parse, facilitating further analyses of intra-community and inter-community dynamics.

My STS topic aims to clarify the differences between Chinese and American community dynamics. More specifically, I scrutinize the degree unification of Chinese sub-communities that are a part of the same geographic communities. To clarify, a geographic community, which may be defined by partitioning using landmarks or generalized geographic regions, consists of several sub-communities. For example, Charlottesville as a whole could be viewed as a singular sub-community a part of the James River watershed. In the setting of Charlottesville, Virginia, sub-community separation has proven to be a major problem for several reasons, namely in those that negatively further problems faced by the greater geographic community.

II. Technical Topic: Understanding the Evolution of Online Community Interactions

Historically, to better understand the dynamics of online community interactions, researchers have created a variety of models, ranging from identification of social roles using graphs and decision trees (Buntain and Golbeck, 2014), and observing network exchange patterns in online communities via exponential random graph models (Faraj and Johnson, 2010). However, what all methods have yet to explore is theorizing and implementing a model for the *trajectory* of a provided online community. For my technical project, we are working to determine similarities between different online sub-communities and possibly predict the trajectory of said communities.

Past models have utilized node network representations of various actors in a community. However, since we will be collecting data on real online sub-communities from *reddit.com*, no extra work is needed with regards to determining the boundaries between sub-communities. This is because Reddit is structured in such a way that posts are organized by a specified topics, or in Reddit terms: *subreddits*. Users may subscribe, post, and comment on specific posts, or to other user comments. Using text and language analysis, which may be done via Natural Language Processing (NLP), Latent Dirichlet Allocation (LDA), and/or other models, we aim to quantify a provided sub-community's, or subreddit's, *values*.

The first steps, before even reaching the modeling and statistical analysis phase, is to create a data-collection tool to parse and collect user comments and activity from target subreddits. Provided that Reddit has a comprehensive and well-documented API, the data-collection tool will likely be written in the form of our own personalized API catered to our uses (and possibly to other users). The data-collection tool's main functionality will include functions

to efficiently receive comprehensive sets of comments, in the form of hierarchical tree structures, from an input subreddit and/or post. The use of tree structures is useful, in that Reddit's comment section itself may be represented as a tree, where multiple *n* users may reply to a singular comment, thus creating *n* child comments.

After the creation of a reliable and efficient data-collection API, we will then likely begin a preliminary analysis, to grasp a better understanding of how different subreddits may be differentiated. In other words, we will aim to grasp this understanding by determining the outstanding features of two subreddits in question. For this step, a K-means analysis will likely be used for feature separation. With better insights into quantifying different subreddits (which may be done by topic, sentiment, etc.), we will then determine the next steps and which models to utilize for our further analysis.

Creating a basic framework with which we can predict and quantify subreddit similarities may prove to be extremely valuable in studying how different communities interact with each other. In our proposed representations of communities in our implementation, we aim to quantify the unique values of each community. With a proper quantification of a set of provided communities' values, a groundwork is set for further analysis on how the introduction of actors in a network of communities causes change.

III. STS Thesis: Unification of Sub-communities within Geographic Communities

Geographic problems are well understood by the general population and well-explained through research and science. However, tackling said geographic problems has proven to be a more difficult task. The prime example of a high-profile geographic problem is stormwater management. It does not take an expert to understand that run-off water as a result of poor

drainage has the potential to distribute harmful chemicals and sediment to nearby rivers, thus in turn transferring these harmful substances deeper throughout a given watershed and connecting larger bodies of water. The primary reason geographic problems are relatively difficult to tackle is that cooperation between the existing relevant sub-communities within the geographic community is required. When taking the perspective of different stakeholders, the lack of cooperation among sub-communities is easily explained by the disproportionate weighing of burdens. My research question, through a comparative analysis of successful and unsuccessful examples, is: how can governing bodies better *environmental governance* to improve efficacy of policies that tackle geographic problems? As later discussed, the enforcement of policies and/or incentives alone is not enough; policymakers must consider the *non-human* actors involved. Provided that *human* actors alone do not sufficiently explain success vs. failure, environmental governance will be analyzed using Actor Network Theory (or ANT).

One significant geographic problem faced by Albemarle County in Virginia is an excess of watershed pollution stemming from the lack of sufficient stormwater management. According to the U.S. Environmental Protection Agency, 762 million pounds of sediment from polluted runoff from impervious surfaces enters the Chesapeake Bay every year from the Rivanna and James Rivers (Savage and Street, 2018). Tackling such a large geographic problem would require the cooperation between various key sub-communities within Albemarle County. However, encouraging such cooperation has proven to be a difficult task.

One solution proposed by Albemarle County in 2014 is the imposition of a ‘rain tax’, which would tax residents based on area of owned impermeable surfaces that hold potential to contributing to stormwater runoff. Said taxes would be utilized to improve local stormwater infrastructure, which would then hopefully reduce the pollution caused by stormwater runoff.

The utilization of a rain tax would essentially force the individual sub-communities to abide by a strict pressure that would in turn effectively 'synchronize' the geographic community to better combat pollution from stormwater runoff. However, it should be noted that such a solution does not constitute a cooperation of sub-communities within a geographic community. In fact, this rain tax caused a controversy, where it pressured specific sub-communities like rural farmers, whose multi-acre properties had many impermeable surfaces. Communities of farmers within Albemarle County, who were affected by the rain tax, felt indirectly targeted, as they would suffer less profit with the presence of a direct tax burden, in comparison to city residents (Baars, 2018). When representing the communities, various environmental issues, and solutions as actors in a larger network, it is simple to see that the stormwater runoff is an actor that brings negative impacts to local communities. However, the rain tax, which is intended to act in opposition to stormwater pollution, also acts in opposition to rural farmers. This highlights a key difficulty in encouraging inter-geographic cooperation in that often times intended solutions often result in an uneven distribution of burden on the actors involved.

In a land farther west, much different culturally and politically in comparison to central Virginia, lies Eastern Asia. Although the geography itself is quite different, the same community interaction framework may be applied: there exist larger geographic communities that consist of multiple sub-communities. Recently and historically, Chinese geographic communities have demonstrated outstanding sub-community coordination towards solving larger geographic issues.

Recently, China has already met its emission-reduction goals for the year 2020. Reaching such a goal, provided its rapidly-emerging economy and substantially large geographic communities, is an impressive achievement. More specifically, China's coal-fired power plant emissions of sulphur dioxide, nitrogen oxide, and particulate matter dropped by 65%, 60%, and

72% respectively (Tang et al, 2019). In comparison of scale to the stormwater pollution issue in Albemarle County, combating a nationwide issue that involves *many* geographic communities is extremely impressive. China's top-down approach to achieve its emission goals ahead of schedule is fascinating in consideration of its provided incentives to the many sub-communities within the larger geographic communities. China's excess air pollution has posed as a public health risk, not only for Chinese residents within cities, but all inhabitants of surrounding geographic communities. In this case, the haze is a harmful actor to many inhabitants of China, especially to those in cities. Researchers in China first determined the main three types of units where enhanced ultra-low emission compliance would best reduce pollution. To achieve these goals, simply forcing strict regulations on various sub-communities would be insufficient, especially considering the prevalence of manufacturing in China. To beat its emission goals far ahead of schedule, China utilized "strict emission supervision system, effective economic incentive mechanisms, and [nationwide applications and upgrades of ultra-low emission] technology" (Nannan, 2019). Further complicating policymaking, it was found that the highest contributor of pollution often varied widely between different models, and on top of that, often changed seasonally (Tao et al, 2016).

There exist many more examples as to how China's top-down approach has expedited cooperation between sub-communities to tackle larger, overarching geographic problems. A key factor in tackling decades worth of air pollution was the addition of incentives between stakeholders. In the previous case, economic incentives were provided to coal-fired units in Western China, namely in Beijing, concurrently to the introduction of ultra-low emission standards (Nannan 2019). To say China's handling of regulations and incentives was challenging is an understatement — to introduce such strict regulations whilst maintaining incentives to all

stakeholders within the geographic communities is a herculean task. Such a sheer difference in successful environmental governance between cities like Beijing and Albemarle County cannot be explained by *human actors* alone. It is clear that, to understand environmental governance, we must consider the *non-human actors* involved in the creation and execution of these environmental policies/incentives. How does the environmental governance differ in Beijing versus Albemarle County? What non-human actors are introduced and are already present? For future study, further literature regarding the policymaking process in China and artifacts containing the societal impact in Beijing must be studied. With the comparative analysis of the human and non-human actors involved in Beijing and Albemarle County, I hope to gain better insights of how sub-community interactions can be better synchronized via the introduction of economic and societal incentives to improve environmental governance.

IV. Bibliography

Baars, S. (2018, April 4). *County controversy: Farmers say rain tax targets rural areas*. Retrieved from https://www.c-ville.com/rain-tax/.

Buntain C., Golbeck J. *Identifying social roles in reddit using network structure*. WWW '14 Companion (2014). doi:10.1145/2567948.2579231

Faraj S., Johnson S. *Network Exchange Patterns in Online Communities*. Organization Science (2010). doi:10.1287/orsc.1100.0600

Nannan, Z. *China Meets Ultra-low Emissions in Advance of the 2020 Goal*. Chinese Academy of Sciences (2019). Savage, R and Street, W. (2018, February 4). *The Crozet Gazette, To the Editor: Yes "Rain Tax"*. Retrieved from https://www.crozetgazette.com/2018/02/04/to-the-editor-yes-rain-tax/

Tang, L., Qu, J., Mi, Z. et al. *Substantial emission reductions from Chinese power plants after the introduction of ultra-low emissions standards*. Nat Energy (2019) doi:10.1038/s41560-019-0468-1

Tao J., Zhang L., Zhang R. et al. *Uncertainty assessment of source attribution of PM2.5 and its water-soluble organic carbon content using different biomass burning tracers in positive matrix factorization analysis — a case study in Beijing, China*. ScienceDirect (2016). doi:10.1016/j.scitotenv.2015.11.057