Spatiotemporal Diffusion of Hostility in Social Media: Toward an Interdisciplinary Collaboration

A position paper for Specialist Meeting on Representing Human Dynamics with Big Data, Social Media, and Social Networks in Hyperlocal Contexts, August 11-12 2015

Author: K. Hazel Kwon, School of Social and Behavioral Sciences, Arizona State University

Social media is a reservoir of human behavioral, attitudinal, and opinion data. While some pundits/scholars point out a risk of social media data misuses for the purpose of reinforcing surveillance and subsequently greater social control, academic communities have foreseen the benefits of leveraging social media data to enhance our understanding of social dynamics for the betterment of human life.

This position paper demonstrates two of my research teams' project activities that explore attitudinal/behavioral dynamics observable in social media. I intend to respond to the workshop's agendas by showing ways in which geospatial patterns of human dynamics could be meaningfully represented. In particular, this position paper highlights the advantages of interdisciplinary collaboration. It is apparent that the comprehensive understanding of human interactions in social media requires the spectrum of approaches ranging from social psychological to computational, and to mathematical perspectives. My research projects exemplify (1) how social science theories could lay the groundwork for organizing and interpreting social media data to meaningfully represent human behaviors, and (2) how computer science and mathematical models could offer the analytic frameworks to scale up the data-driven empirical investigation of social problems. Below, I will describe two of my research projects to discuss the role of social theories and innovative interdisciplinary collaboration in helping us insightfully analyze spatial social media data.

- **Project 1: Incorporating a mathematical diffusion model into social media-based social movement research.**

This project explores a transnational diffusion process of social movement ideas in social media by introducing a mathematical model. While the use of social media in promoting social movements has been widely discussed in literature, few studies have empirically examined the dynamic diffusion process along temporal and spatial dimensions, and even fewer have taken into account international relational factors that may interplay with transnational information diffusion processes. To address this gap, we adopted the sociological lens of transnational social movement and reviewed different notions of spatial proximity pertaining to ideology (the level of democracy), population (migration size), economy (bilateral import/export relations), and physical distance. We applied them to a partial differential equation (PDE)-based mathematical model called "diffusion-advection model," which allows for an examination of a spatial dimension without disregarding the temporal diffusion aspect. To validate the model, twitter activities during the Egyptian Revolution of 2011 were analyzed. Empirical results suggest that the spatial diffusion of information was most effectively explained when the data was fitted using a

democracy-based spatial arrangement. We also referred to the collective action theories in Sociology to classify different types of messages (e.g., improvised reporting, situation-verifying, and action-supportive messages) and observed the differences in diffusion patterns among them. By demonstrating the model validity with Egyptian revolution Twitter data, the project shows the potential to apply mathematical models to social scientific research.

- **Project 2: Incorporating a machine learning approach to automatically detect hostile messages during a social crisis in social media.**

While social media communication often contributes to collective sense making and consensus building, it occasionally serves as a platform through which prejudice is ventilated. Expressions of prejudice and intergroup hostility markedly increase especially when a society undergoes a collective threat such as unforeseen crisis events. In response to the call for understanding intergroup relations represented in social media at times of crisis, this project aims to develop a machine-learning model that identifies group-divisive, hostile messages. Building a computational model that reliably detects the cues of intergroup prejudice and conflict out of the vast array of social media messages may lay the groundwork for the advancement of a human-machine collaborative system for hostility reduction and effective social reporting for a crisis management. After preprocessing by using Google Translate, removing duplicates, and filtering only negative sentiments, we employed a knowledge discovery framework, including feature extraction (which was defined by referring to intergroup communication/psychology literature), label learning (which leveraged both human annotation and automation approach), and designing machine-learning models. The modeling so far achieves a high rate of precision and recall with the balanced sample sets. We are currently working on increasing precision/recall rates with the unbalanced sample distribution.

These two projects demonstrate a potential for interdisciplinary collaboration to accomplish a systematic social media analysis with the intent of social problem-solving. My team's future plan is to integrate the two projects for exploration of the hostility diffusion pattern in a localized context. Along with the automated hostility detection, a reliable geo-coding is also important. Human annotation of our dataset showed that, when we manually examine profiles and their recent tweets carefully, detecting where the Twitter user currently dwells seems plausible in the form of best 'guesswork.' The current team effort is focused around determining whether or not a computerized guesswork is adequately reliable in identifying locative cues, and up to which level –e.g., nation, state, or city –they can be reliably predicted. The computer-assisted identification of hostility and geo-locative cues may enable us to utilize a large-scale social media data for understanding how quickly (temporal diffusion rate) and how broadly (spatial diffusion rate) hostility is propagated under a crisis.