Mapping and Modeling Strategic Manipulation and Adversarial Propaganda in Social Media: Towards a tipping point/critical mass model

Position paper submitted to
Workshop on Mapping Ideas: Discovering and Information Landscapes
San Diego State University, San Diego, CA
June 29-30, 2011

submitted by
Daniel Z. Sui
Department of Geography & Center for Urban & Regional Analysis (CURA)
The Ohio State University
Contact info: 1036 Derby Hall, 154 North Oval Mall, OSU, Columbus, OH 43210
E-mail: sui.10@osu.edu; phone: (614) 688-5441; fax: (614) 292-2320

Due to the explosive growth of Web 2.0 technologies and related services/applications in recent years, we have witnessed the convergence of the technological networks that connect computers on the Internet and the social networks that have linked humans for millennia. The growing popularity of social media, such as Facebook, Twitter, LinkedIn, MySpace, del.icio.us, Flickr, YouTube, PatientsLikeMe, etc., is becoming a new global trend among millions of users worldwide, as so vividly demonstrated in the most recent anti-government protests in the middle-east.

However, just like any other major technological innovations throughout human history, social media has its dark side as well. Inflammatory messages posted on social networking sites by certain radical groups have triggered clashes of civilizations at the global level on the one hand whereas other seemingly innocent personal messages have caused divorce, jealousy, and loss of privacy at the individual level on the other hand. While the positive impacts of social media from local to global levels are well documented, the dark side of social media has not been sufficiently brought to light despite its potentially pernicious effects on social networking. One disturbing new development is the development of sites like Subvert and Profit (www.subvertandprofit.com), which claims to have access to "25,000 users who earn money by viewing, voting, fanning, rating, or posting assigned tasks" across social media sites. Related services can be found at fansandinvites.com, socioniks.com, and usocial.net. Beyond advertising and product promotion, there are also reports of strategic manipulation of social media for political gains – the so-called "astroturfing", as illustrated in "Training Tea Party Activists In Guerilla Internet Tactics" (www.astroturfwars.com).

Even within the great firewalls of China, we have witnessed the emergence of the so-called "Wang Luo Shui Jun," which can be directly translated as "Online Water Army" (e.g., http://shuijunwang.com or http://www.51shuijun.net). It represents full-time or part-time on-line "mercenaries" to help a specific company or a person to post articles, replies, and comments on online bulletin boarding system (BBS), social networks, and websites of major media. Usually hundreds, sometimes even thousands, of online "mercenaries" work together to help their customers to accomplish their desired goals. According to a recent CCTV report (http://news.cntv.cn/china/20101107/102619.shtml), online mercenaries in China help their customers using one of the following three tactics: 1. Promotion of a specific product, company, person or message; 2. Smear/slander the competitor or adversary or competitors' products or services; 3. Help delete negative or unfavorable posts or news articles. To fulfill their goal, the online "mercenaries" use different strategies ranging from word of mouth to viral marketing. Most online "mercenaries" work part-time and are paid around 5 US cents per post in BBS systems, and online social networks.

It is a general consensus among researchers, entrepreneurs, politicians, and Internet users that in recent years, the social web has quickly become a weapon of mass persuasion in personal, socio-economic, political-cultural as well as conventional warfare. And yet the rules of engagement of using social media for mass persuasion have been not very well understood. Intuitively, a critical mass is needed in order to shape the public opinion or perception about a phenomenon according to classic communication theories using the conventional medium such as word of mouth or printing press. The idea of a tipping point has been discussed by some bloggers as one of the key rules for using social media effectively. But so far there exist no methods to detect the tipping point in mass persuasion using social media. The lack of knowledge in this area has hindered our effort to make social computing more intelligent, much less for other higher social goals.

This paper reports the preliminary results on an ongoing project that aims to undertake such a challenge by developing new techniques and algorithms to more effectively detect strategic manipulation and adversarial propaganda via leveraging the specific abilities of massive numbers of human participants. More specifically, I will report preliminary findings related to following two aspects: 1). techniques for detecting and mapping strategic manipulation and adversarial propaganda in social media; 2). a tipping point/critical mass model of mass persuasion to better monitor the social trends using data harvested from social media.

Part 1: Detecting and mapping strategic manipulation and adversarial propaganda

Toward the goal of efficiently detecting campaigns of strategic manipulation in social media, we have focused on detecting coordinated campaigns that rely on "free text" posts, like those found on blogs, comments, forum postings, short status updates (like on Twitter and Facebook), and so on. For our purposes, a *campaign* is an ad-hoc collection of users and their posts bound together by some common objective, e.g., promoting a product, criticizing a politician, inserting disinformation into an online discussion. Some campaigns may be organic and natural outgrowths of popular sentiment; others may be strategically organized. Our approach is to first identify and extract candidate campaigns from the massive scale of the real-time social web and then efficiently and successfully track the evolution of campaigns over time as users join, campaigns merge, and campaigns disperse. While there has been some progress in detecting isolated instances of long-form fake reviews (e.g., to promote books on Amazon) and in manipulating recommender systems, there is a significant need for new methods to support large-scale detection of coordinated campaigns. First, systems like Facebook and Twitter are extremely large (on the order of 100s of millions of unique users) placing huge demands on detected coordinated postings, especially considering the inherent lack of context in short posts. Second, these services support a high-rate of communication (e.g., new tweets inserted into the system; Twitter alone supports on the order of 80 million tweets per day) so the discovered campaigns may become stale quickly, resulting in the need to re-identify all campaigns at regular intervals (potentially incurring the high cost of community detection, which can be $O(n^3)$ in the number of users. The bursty nature of user communication demands a campaign discovery approach that can capture these highly-temporal based clusters. Third, campaigns may evolve at different rates, with some evolving over several minutes, while others taking days or even months. Since campaigns are inherently ad-hoc (without unique community identifiers), the formation, growth and dispersal of campaigns must be carefully managed for meaningful analysis.

Part 2: Modeling the tipping point in mass persuasion

The second focus of my paper will be on a generalized approach for modeling online campaigns and the dynamics of campaigns that reach a "critical mass" and attract a large and growing following. Some campaigns may impact only a small collection of targets (e.g., via targeted advertising or through direct payment as in the "Army of Water") whereas other campaigns may go "viral" and impact a much larger audience beyond the initially targeted group, such as the most recent Tweets in Tunisia about their government corruption as revealed by Wikileaks .

To take on this challenge, our approach builds on and extends theories of self-organized criticality that have been developed to study dynamical systems in the physical and natural sciences.   In this project, we try to initiate a study of the connection between self-organized criticality and mass persuasion, exploring the characteristics of critical points in these systems (the "tipping points"). To illustrate, we can consider a simple model in which users U are targets of a set of campaigns C. The campaigns are targeted using a set of information capsules I_c generated by S sources of different reputation (corresponding to strategic manipulation, e.g., a company employing the "Army of Water"). The number of capsules used to promote a campaign is proportional to the expense the promoter is ready to bear. Every user has a threshold θ_c for each campaign depending upon the impact the campaign can have. For example, users might have a higher threshold to believe a campaign about a Presidential election while a lower threshold to accept a campaign for a soft drink. We consider a simple user behavior model in which targets may be "activated" by a promotion when their threshold for a particular campaign is exceeded, but can occasionally change their opinion after making a decision. We model the decision criteria for a user to accept a campaign by a probability function $f_i = f(i_c) \, x \, f(c_a)$, where $f(i_c) \rightarrow [0, 1]$, is a monotonically increasing function with the number of information capsules i_c for a particular campaign as the parameter, and $f(c_a) \rightarrow [0,1]$ is a monotonically decreasing function with the age of the campaign c_a as the parameter. In a simulation of 25 campaigns over 10,000 users, we observe that under this model the impact of a campaign falls exponentially once reaching its peak, and that, the distribution of campaign sizes ranges in size, but it is not obvious if such a simple model effectively captures the dynamics of a self-organized critical system. We will explore under what scenarios such persuasion models display self-organized criticality by varying the "compartments" associated with the model, by considering different threshold models, by varying the underlying network topologies, and so forth. Under what scenarios do tipping points naturally occur (for example, leading to massive adoption posting of an "Army of Water" campaign, and under what scenarios is the system inherently resistant to such tipping points? How do the developed models correspond to the actual campaigns identified as part of the previous objective? What are the characteristics of campaigns that achieve critical mass? By linking these models to data-driven evidence of the first objective, we can rigorously explore the connections between mass persuasion and self-organized criticality.

This project is an extension of my long-term research interests in GIS as media and mapping social media using GIS. As part of this project, we will identify concrete instances of strategic manipulation and adversarial propaganda which can be used as part of a common dataset for mapping and modeling social media. We are in the process of developing feature extraction, feature weighting, and machine learning based approaches for automatically detecting the tipping points of mass persuasion in social networks. Our ultimate goal is produce a set social media mapping and modeling toolkit for automatically analyzing and visualizing the tipping points of mass persuasion that can be eventually linked to Google Maps/Earth. Insights gained by such an effort can then be used to design and develop more intelligent systems for collective & participatory computing.