A computational social science approach to agenda setting in chronic crisis situations: The California Drought Case Study

Sorin Adam Matei, Robert Kulzic, Valeria Chapman, Kartik Ariyur, Nihkil Sangwan, Venkatesh Merwarde, and Nicole Ningning Kong

Keywords: social media, crisis communication, agenda setting, mass media, drought

Crisis situations generate a large amount of information, fueled by and fueling at the same time, public interest. The manner in which the public agendas and media messages intersect becomes of paramount importance. A complex interlocked network of one to many (broadcast, traditional print), many-to-many (computer mediated), face to face (interpersonal) conversations has emerged in the last two decades. Understanding how messages and information travel and how they affect public opinion and institutional response to crisis situations is a research challenge that needs to be met. The problem becomes even more complex and the answers urgent as established media producers (television networks, newspapers, magazines, freelance reporters, new agencies) have become major actors on social media, the boundaries between sources of destination of media content becoming increasingly blurred. Fortunately, because social media tag each unit of content according to its source and its trajectory in the consumption space, we can easily determine its individual path. A fascinating research question, particularly relevant in crisis situations, is if content units are more or less likely to start with a traditional media, social (read, non-professional, consumer), or more generic computer mediated sources. In terms already consecrated in mass communication research, who sets the public opinion agenda in terms of crisis? Who and in what way starts the news cycle and what role do social media consumers play vis-à-vis traditional media channels? Emerging research suggested that a hybrid mass-personal media space has emerged, with social media taking the lead. However, the answers are not definitive, as quite the opposite has also been found, namely, that mass media is an important driving force in the process of content creating and dissemination. In other words, social media can be at times used as a message amplifier by traditional media.

The intersection between interpersonal, computer mediated, social media, or mass media (traditional) conversations presents not only challenges, but also opportunities. In one way or another, most information is transferred in a digital format, which makes its trajectory visible. The massive amounts of digital content created in crisis situations can be mined with computational social science approaches. These allow fast and efficient identification of the sources, destinations, and direction of information flows. Once identified, the effect and precedence of some information flows relative to others and of possible agenda setting effects can be more easily studied and specific hypotheses about the sources of agenda setting process can be proposed.

Our case study examines how Google searches, social media (Twitter status updates) and mass media satisfied the public information needs during the California drought of 2013-1015. It also highlights how a complex social scientific computational approach can offer the means to study important social scientific research questions.

Research context

In the 2014, California got 49 % of the historical average precipitation, making it impossible to sustain all needs. The Governor of California declared a State of Emergency on January 17, 2014. The

water year ending in September 2014 proved to be the third driest on record. (citation
http://ww2.kqed.org/science/2014/09/30/californias-water-year-ends-as-third-driest-on-record/)

Tensions over water usage between various regions of California and different segments of the population have been brought to the surface by the drought. Intense debates carried in old and new media forums compelled policymakers to respond. All these interactions were transacted in the digital communication sphere, from newspaper comment sections, to social media, or public Google searches for answers to the questions generated by the crisis. A chain of recursive mediated causes and effects emerged and explaining the communication processes behind them became very important. For an effect to be explained, a theoretical framework should be offered. Agenda setting theory is very appropriate in this context, since it offers explanatory mechanisms for the manner in which media prioritizes messages, emphasizes their importance, or selects some messages at the expense of others to ultimately influence the manner in which individuals make decisions. The theory needs, however, to be enhanced with a more sophisticated research strategy, which takes into account that media can take many forms, from traditional journalistic messages, to social media conversations, or to individual information seeking and consumption such as those supported by search engines.

We chose the California drought and water crisis of 2013-2015 as an example for how computational social science can be employed in an applied social-mass communication agenda setting research project. Media coverage, social media conversations and dissemination of news, facts and opinion, and the policy interventions overlapped and affected each other. In this context, it becomes important to determine what came first. Did social media pressure lead to traditional media coverage of the crisis? Or did traditional media create the social media discussions? Furthermore, how do individual information seeking strategies, such as use of search engines, preceded or succeeded social media or traditional media interventions in the debate?

Method

To answer these questions, we collected information both about the observable severity of the drought and about the human reactions to it, as observed in mediated spaces, for 02/2013-09/2015. The physical impact of the drought was measured by standardized soil moisture index readings from San Joaquin River Valley, the area of California most impacted by the drought, both by severity and economic impact. The standardized soil moisture index is calculated using soil moisture data from NASA's Land Data Assimilation Series. The mediated impact of the drought was assessed looking at volume of Google searches related to the drought and at Twitter activity and media coverage of the California water crisis. We collected data using a theoretically and pre-tested query algorithm that uses as a core strings "California water" and "California drought". From Google Trends we retrieved volume of searches in California for "California Water" and "California Drought." Using the Twitter API we retrieved the query ((climate OR drought OR water) (california OR ca)) OR cawater OR cadrought OR saveourwater. It is more complex because it includes several hashtags, specific to Twitter. The query was limited in space to central California tweets. A total of 186,149 tweets were retrieved. The site of the newspaper Modesto Bee, which is he most important regional newspaper covering the area most impacted by the crisis, the San Joaquin River Valley, was crawled for articles containing the string "California water." 1853 articles were retrieved.

Newspaper and Twitter data was cleaned for relevance. Non-relevant items were identified by hand coding a random sample of the collected items and then using a Multinomial Naive Bayesian algorithm to extend the classification made in the hand coded items to the entire data set. This avoided including in the dataset tweets related to water heater troubles or similar tweets.

Analysis, results, and conclusion

In order to ascertain the relationship among the four trends (soil moisture index, Google Search, Twitter, and Modesto Bee coverage) a mix of time-series analysis approaches were used. An analysis of association and temporal precedence between geological data (soil moisture index) and the media data series indicates that only Twitter data is significantly associated with a decline in water supply. Twitter activity reacted in just under a month. Vector Auto-Regression analysis was used to determine which media series, if any, influences the outcomes of the other media. Following a diagnostic test to determine which order of VAR was appropriate, an examination of the VAR results for statistically significant coefficients provides evidence that Google searches influenced both Twitter activity and the number of newspaper articles. A one standard deviation increase in Google Trends from the previous week results in almost two-thirds of a standard deviation increase in the number of Tweets. There is an almost equally strong relationship between the number of articles published on the drought and the previous week's Google search activity.

In conclusion, the core chain of interactions in the conversation about the California drought crisis takes the following path. Twitter conversations emerge as early as three weeks after drought conditions start worsening and increase in strength with a lag of tree weeks thereafter. Twitter conversations are further influenced and feed on by Google searches. As people search more about the drought, more Twitting activity is generated. Finally, media coverage comes last, increasing in intensity after Twitter and Google search behaviors intensify.

Our study indicates that the agenda-setting role of the media has waned, at least in this situation of chronic crisis. Media now follows rather than leads the conversation about the crisis, the leaders being social media (Twitter) supported by Google search.
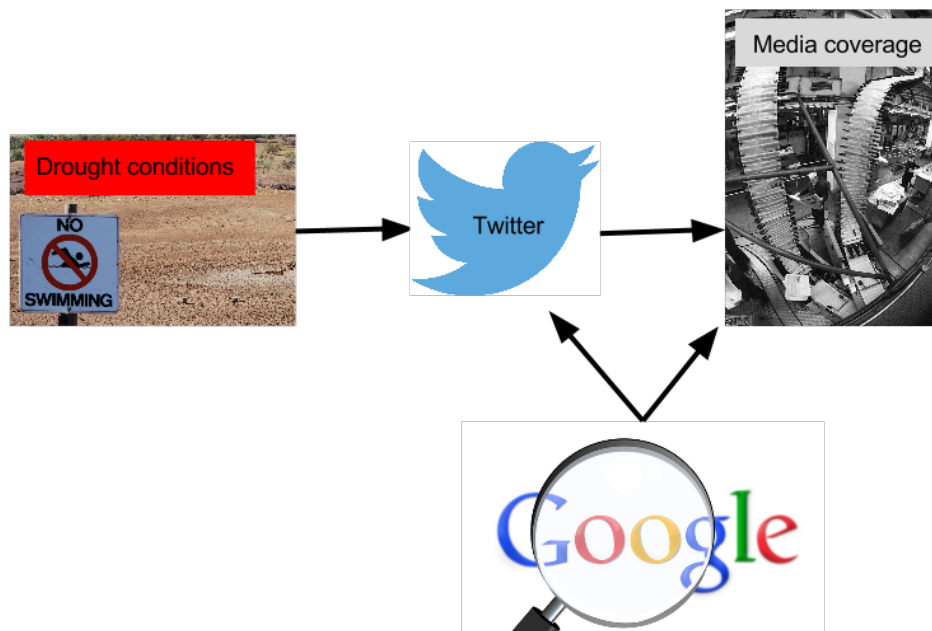

Figure 1. Chain of media effects in the California Drought Crisis

3 – 3 pages, Sorin Adam Matei et. al