

Social Media Use by Government: From the Routine to the Critical

Andrea Kavanaugh* λ

Seungwon Yang* λ

Donald Shoemaker* +

*Virginia Tech (0902)

λ Department of: Computer Science,
+ Sociology, α Information Systems
(540) 231-1806

{kavan, fox, sheetz, seungwon,
tfw115, shoemaker}@vt.edu

Edward A. Fox* λ

Lin Tyz Li* ψ

Paul Natsev #

ψ Institute of Computing
University of Campinas
Campinas, SP Brazil

lintzyli@gmail.com

Steven Sheetz* α

Travis Whalen* +

Lexing Xie #

IBM Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532
(914) 784-7541

{natsev, xlx}@us.ibm.com

ABSTRACT

Social media (i.e., Twitter, Facebook, Flickr, YouTube) and other services with user-generated content have made a staggering amount of information (and misinformation) available. Government officials seek to leverage these resources to improve services and communication with citizens. Yet, the sheer volume of social data streams generates substantial noise that must be filtered. Nonetheless, potential exists to identify issues in real time, such that emergency management can monitor and respond to issues concerning public safety. By detecting meaningful patterns and trends in the stream of messages and information flow, events can be identified as spikes in activity, while meaning can be deciphered through changes in content. This paper presents findings from a pilot study we conducted between June and December 2010 with government officials in Arlington, Virginia (and the greater National Capitol Region around Washington, DC) with a view to understanding the use of social media by government officials as well as community organizations, businesses and the public. We are especially interested in understanding social media use in crisis situations (whether severe or fairly common, such as traffic or weather crises).

Categories and Subject Descriptors

K.4.2 [Computing Milieux]: Computers and Society – *social issues*.

General Terms

Algorithms, Management, Measurement, Experimentation, Human Factors.

Keywords

Digital government, crisis informatics, social media.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Digital Government '11, June 12–15, 2011, College Park, MD, USA.
Copyright 2011 ACM 1-58113-000-0/00/0010...\$10.00.

1. INTRODUCTION

Citizens are increasingly relying on social media for communication with their friends, work teams, social organizations, and government. The capabilities to facilitate interpersonal and group interaction provide new and unique opportunities for community leaders, elected officials, and government service providers to inform, and be informed by, the citizenry. Twitter, Facebook, Flickr, YouTube, and other services with user-generated content have made a staggering amount of information available.

Government officials seek to leverage these resources to improve services and communication with citizens, especially segments of the population that previously were underrepresented. Yet, the sheer volume of social data streams generates substantial noise that must be filtered. Nonetheless, potential exists to identify issues in real time, such that emergency management can monitor and respond to issues concerning public safety. By detecting meaningful patterns and trends in the stream of messages and information flow, events can be identified as spikes in activity, while meaning can be deciphered through changes in content.

Similarly, monitoring these patterns and themes over time could provide officials with insights into the perceptions and mood of the community that cannot be collected through traditional methods (e.g., phone or mail surveys) due to a variety of reasons. First, and perhaps most importantly to emergency management, no other method works in real time. Surveys require substantial time and effort prior to data collection, during the collection process, and for analyses of the results. Often taking months to complete effectively. Secondly, substantive costs are associated with these survey activities, making them especially difficult in light of reduced and shrinking budgets of governments at all levels. Finally, once completed a survey captures perceptions at a single point in time. Although it is possible to use surveys at intervals to monitor progress, it is not a common practice.

The depth and breadth of information flow is breathtaking, with Twitter generating an estimated 55 million tweets a day ("Twitter blog: Measuring tweets."), Flickr amassing more than 6000 photos each minute (Flickr), YouTube accumulating over 24 hours of video a minute, taking up more than 10% of all internet traffic (YouTube), and Facebook having more than 400 million active users, making it the most visited site on the Internet in the USA.

Mining a diverse real-time feed of social streams related to real-world events is needed to enable officials to make sense of the vast amount of information that is generated. In so doing, government should be able to act more effectively on matters both routine (e.g., ongoing issues of public concern) and critical (e.g., major weather or traffic disruption, public safety or rapid response). We can answer questions that cannot be addressed with the gather-and-report style of journalism involving traditional sources, such as: When and where are events currently happening? What are the different views of a given event? Which social media should government use to communicate most effectively with a diverse public? How should messages be formed and framed across social media to be effective? To what extent can messages in social networks be used to explain how influential messages form and spread? Who are the influential users in an online or local community? Is civic information, disseminated through social media as opposed to through the Web or email, more likely to reach some traditionally underrepresented groups, such as those with lower socio-economic status (SES) or younger voters? What role do social media play in the general mix of information sources for citizens to communicate about civic life, with each other and with government? How can social media be used to affect civic participation?

We seek to leverage technology to help government manage information and facilitate interaction in meaningful ways in order to achieve broader public participation than is possible through normal channels (e.g., public commenting at county board meetings). Deep analysis of social media streams can provide access to segments of the community that have not participated in traditional ways.

This pilot study was part of a larger investigation funded by NSF (IIS-0916733) to build a Crisis, Tragedy and Recovery Network (CTRnet) (<http://www.ctrnet.net>). In collaboration with Arlington Virginia County government, we conducted a six-month pilot study of how social media data analysis can be applied in Arlington and environs to improve services and communication with citizens. Our primary research objectives are to investigate the use and impact of social media and to identify and develop methods to effectively meet a variety of local government and community needs. Specifically, we have begun to:

- 1) leverage and further refine tools for collecting and correlating large amounts of public social media data relevant to Arlington County, VA and environs,
- 2) archive and curate collected social media data over a period of time into a digital library, including social media for crisis conditions, and
- 3) identify, research and implement applications of multimedia analytics and text mining for government services and communication.

To address these goals we conducted an exploratory study of social media use in Arlington, Virginia and environs. We crawled, collected, aggregated, and archived relevant social media data; conducted exploratory focus group interviews with key stakeholders in government and community leaders; and developed tools to analyze and render data more usable and meaningful for local governments and citizens.

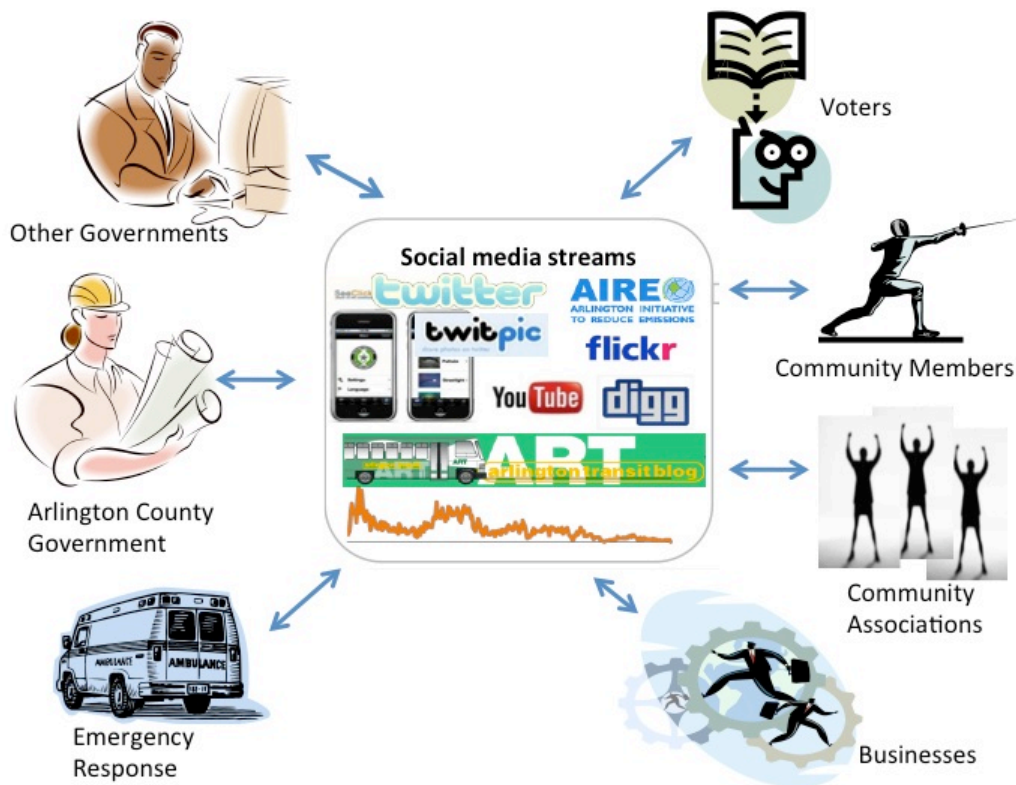


Figure 1. Social media streams to improve services and communication with citizens

Our target information sources include official county publicity portals ("Arlington County Blog Central; Arlington County Facebook Profile; Arlington County Flickr account; Arlington County News on Tweeter; Arlington County, VA Official Site; See-Click-Fix in Arlington County,"), blogs, news, community forums, as well as relevant postings by the public on social media sites such as Twitter, Facebook, YouTube, and Flickr. Applications of such analyses could include monitoring public opinion before and after large public events, monitoring planned or unplanned activities, identifying and categorizing important community issues over time and location, enhancing community recovery in response to crises or tragedies, and monitoring and tracking the development of long-running themes in civil life.

Our pilot study was funded by the Virginia Tech Center for Community Security and Resilience (CCSR) July-December 2010. The CCSR is a partnership among Virginia Tech, IBM, and Arlington County. Based on interests and needs demonstrated in a CCSR workshop with officials from Arlington County and the National Capital Region (NCR) (the area around Washington, D.C.), we planned the pilot study in collaboration with IBM and Arlington County government to explore social media applications that might improve community resilience in times of crises, as well as provide timely and complementary open sources of information for facilitating city, county, and community services. Further, we explored social media applications that might help agencies make sense of the deluge of information by providing meaningful consumable insights.

2. SOCIAL MEDIA AND GOVERNMENT

Social media are internet-based applications designed for use, development and diffusion through social interaction. Social media build on many of the same concepts and technologies of Web 2.0, most basically, the creation and exchange of user generated content (O'Reilly, 2007). There is much overlap between the two concepts and technologies in terms of examples, including blogs, wikis, ratings, and recommender systems; websites to share videos, music, pictures and podcasts; and social networking sites such as Facebook and MySpace. Broadly, Web 2.0 and social media are considered social software, i.e., software that enables people to rendezvous, connect, or collaborate through computer-mediated communication (boyd & Ellison, 2007; Lampe, Ellison, & Steinfield, 2006). This type of software has existed for years in the form of online bulletin board systems (BBS), listservs, forums, and newsgroups. More recently, however, blogs (Tepper, 2003) and microblogs (e.g., Twitter), RSS feeds, tagging systems (Furnas et al., 2006), and collaborative filters have made social software very popular.

2.1 Social Media Use by Citizens

Just as social media and just-in-time applications have changed the way Americans get information about current events or health information, these media are now providing new ways for citizens to interact with each other and with elected officials and government agencies. A national study conducted by Pew Internet & American Life in 2010 finds that almost a third (31%) of all online adults in the USA used social tools such as blogs, social networking sites, and online video as well as email and text alerts to keep informed about government activities (Smith, 2010).

Social media seem to have particular appeal for groups that have historically lagged in their use of other online government offerings -- in particular, minority Americans (Smith, 2010). Latinos and African Americans are just as likely as whites to use

these tools to keep up with government, and are much more likely to agree that government outreach using these channels makes government more accessible and helps people be more informed about what government agencies are doing. Findings from the national Pew study also show that 40% of adult Internet users have gone online for raw data about government spending and activities. This includes anyone who has done at least one of the following: looked online to see how federal stimulus money is being spent (23% of internet users have done this); read or downloaded the text of legislation (22%); visited a site such as data.gov that provides access to government data (16%); or looked online to see who is contributing to the campaigns of their elected officials (14%).

In a 2009 online convenience sample survey conducted in the US by the American Red Cross, 75% of respondents reported they would use social media in crisis and civic-related situations (e.g., traffic jam, car crash, potential crime, or downed power lines). Nearly half would use social media to let others know they were safe in an emergency; 86% report they would use Facebook; 28% would use Twitter, and 11% would use a blog. Solutions that are (already) provided by the industry for public safety include call processing products and notification systems. For example, Plant CML offers call processing software that is used by 2/3 of all 911 centers in North America. They also provide notification systems, CAD & Mapping, data management and analysis, information management, and land mobile radio. These systems, however, are mostly based on phone communications and are not using the power of social media (<http://www.plantcml-eads.com/solutions-products/public-safety>).

Large public gathering events, such as parades or demonstrations, are examples of conditions of social convergence, that is, high-intensity events with large population density and heightened security needs. Before the event it is beneficial to monitor online discussions on national and global sources, such as YouTube and Twitter, as well as local sources, such as Arlington blog central ("Arlington County Blog Central,"), local Facebook pages, YouTube and Twitter posts ("Arlington County Facebook Profile; Arlington County News on Twitter; Gasbuddy: Find local gas prices,"), or Foursquare "check-ins" (or similar location-aware mobile media applications). This monitoring helps community leaders and the public to stay informed about the various perspectives, sentiments, feedback, and insights around an event or an issue of interest. Afterwards, if a security event has emerged (e.g., violence, vandalism), sometimes evidence will be posted on photo and video sites, which can help local officials to identify and track suspects as an event progresses. In epidemic propagation and prevention, on the other hand, the focus of information management is on early spotting of cases and managing public input, contributions, and feedback around issues like quarantine, vaccination, and distribution of sanitary advice (e.g., swine flu).

Research on the use of Twitter in crises has a short history, due to Twitter's short life. A form of micro-blogging using an open-source web framework called Ruby on Rails, Twitter is a free, short messaging service with some social networking features, established in 2006. Some of the most relevant work to ours has been done by Palen, Hughes, and colleagues (Hughes & Palen, 2009; Hughes, Palen, Sutton, Liu, & Vieweg, 2008) and by Zuckerman on the Moldovan election protests in Africa (Zuckerman, 2009). These studies specifically focus on the use of Twitter in a disaster or crisis situation. Hughes et al. (Hughes & Palen, 2009) report that Twitter use under duress and in crisis conditions of the two hurricane episodes of Ike and Gustav in

2008 is distinct from routine general Twitter communication behavior in two ways: 1) fewer tweets are sent as replies to other tweets; and 2) fewer URLs are included in the tweet posts. They surmise that this is because in a crisis, people need to broadcast information as widely as possible to as many people as possible at once (i.e., no need to reply to a specific individual) and people are less likely to go to a website for additional information during an emergency.

2.2 Social Media Use by Government

Twitter and other social sources have been effective in early event spotting (Opsahl, 2010; Sakaki, Okazaki, & Matsuo, 2010), the response time of which can be even faster than official sources (e.g., earthquake reporting). Such monitoring strategies also can be used for epidemic spotting and trending, where monitoring should be both distributed and spanning a longer period of time, such as the first case in each school district, resurgence of disease cases, and long-range planning for local management. In the case of continuous monitoring, social media can help measure the effectiveness of control measures and propaganda, e.g., if the public is embracing the vaccine distribution scheme, complaining about it, or helping authorities stay better informed about gaps or deficiencies in its execution.

We have been studying social media use and impact as part of an ongoing longitudinal investigation of Internet use and impact in XXX, YYY and environs since the early 1990s ("Social Media Sells," 2010). XXX is home to the main campus of XXX (which also has a small campus in northern Virginia near Arlington) and is home to the community computer network known as the YYY. XXX town government has won several awards for its rich mix of media to inform and communicate with citizens, including Twitter and Facebook since January 2009 as an additional channel for 'XXX Alerts' available by email or text message. The Communications Specialist in town government monitors Twitter (using TweetDeck) for relevant posts that would benefit from a reply (e.g., "the town does not have control over the old middle school in Blacksburg, that is the County's jurisdiction") or should be brought to the attention of town council as a citizen suggestion ("it would help to have a cross walk painted at this intersection; it's very busy").

While this was not the case for the town of XXX government, in the National Capitol Region, focus group participants noted that the public relations person for various government agencies was typically not familiar with nor comfortable with social media. This limitation makes it especially difficult for the public relations office to manage this channel of communication with the public.

From our preliminary study of social media use in XXX, we found that sometimes the person posting tweets or managing an organization's Facebook page was not the organization leadership. Instead, a college student or other young adult was often working in tandem on behalf of the organization to post announcements, updates or other information. Some other US communities, such as Catawba County, North Carolina and our partners at Arlington County, Virginia, are experimenting with monitoring Twitter and Facebook using a Web tool called Hootsuite, attempting to monitor social media communications and potentially to reduce workload and enhance responses at 911 centers (Andrew Opsahl, 2010).

3. STUDY METHODS

We collected and analyzed area-specific social media (SM) sources, and conducted focus group interviews with 24 county

officials (specifically, personnel from emergency management services, the police department, and volunteer leadership office), including a questionnaire about their social media use and community involvement. We were able to recruit 25 participants and organized them into three separate focus group sessions (lasting two hours each) held in November and December 2010 in Arlington. At the outset of each of the interview sessions, we asked participants to complete an online questionnaire. The questionnaire asked them about their use of social media and their involvement in the local community.

The focus group sessions consisted of two steps, beginning with the participants engaged in electronic brainstorming to generate a substantive number of ideas quickly, followed by their identifying categories that grouped the ideas by similarity.

Using individual computers with group support software that we developed, the focus group participants anonymously generated and entered ideas, beliefs, issues, or concepts, in the form of short sentences or phrases that they feel were important to the situation. We provided them with a set of framing questions we developed to cue participants to begin entering ideas. Figure 2 shows the framing questions we used in the focus groups we conducted. The ideas participants generate are shared with all team members as

- What are the missions and objectives of your organization?
- What are you trying to accomplish using social media?
 - Do you feel you are currently accomplishing this goal effectively with social media? (if yes, why?)
 - If not, what do you need [to know? – to do? -- in order] to use social media more effectively?
- What concerns do you have about using social media?
- What difficulties do you have about using social media?
- What information would you like to have about how your organization uses social media?
- What information would you like to have about how social media is being used in your community?
- Is there anything else you would like to know about social media that would be helpful?

Figure 2. Framing Questions for Focus Group Interviews

they are generated, allowing ideas generated by one person to be expanded by others or to cue others to generate related ideas. Team members then worked together with the facilitator to create and name the meaning units or categories that organize their ideas by similarity.

We collected social media in the form of official posts and public comment data from the Arlington County Facebook page, twitter feeds from local civic organizations, YouTube videos, and crawls and searches of local web pages. We used different twitter analytical tools, such as '104 kit' (<http://www.104kit.com>) and the Archivist (<http://archivist.visitmix.com>) to collect tweets from 34 local organizations, including Arlington government, that were civic in nature (rather than commercial or residential).

We performed semantic analyses on the twitter data to identify popular topics and to characterize followers by their profile data; we conducted simple frequency counts to calculate the number of 'followers' and 'followers of followers' of a given organization. We used visualization software 'wordle' (<http://www.wordle.net>) to represent the results of the twitter analyses as tag clouds in order to be able quickly and easily to make sense of large amounts of data. For the YouTube video collections, we used Perl script to search all YouTube videos for the tags or video title 'Arlington

County' and represented the search results in a tag cloud indicating the most frequent tags in the image collection.

4. RESULTS

Our findings from the pilot study are based on the focus group interviews and participant questionnaires (N=25), the development of tools to analyze social media data we collected. The results fall into three main areas:

- 1) local government uses social media without knowing its costs and benefits, or who their actual audience is, who in their organization should monitor communications, how and when they should be responding, and what effect their social media communications have on the public;
- 2) new tools are needed to help government and citizens make sense of the overwhelming amount of data that is being generated, to model the flow of information, and to identify patterns over time; and
- 3) digital libraries are needed to archive and curate generated content, especially for crisis and social convergence situations, but also for analyses that cover longer time frames.

4.1 Focus Group Questionnaire

The 25 focus group participants completed an online-questionnaire at the outset of the focus group interview session. Of this sample, 15 (60%) were female and 10 (40%) were male. The majority (84%) was white, non-Hispanic. Sixty-four percent were married and 92% were employed on a full-time basis. It is reasonable to characterize the interviewees as community leaders, as they reported being very active in their community being well informed about local news and politics. Seventy-six percent reported that they kept up with local news daily.

Most respondents reported having ideas for improving their community at least once a month (76%) and that they frequently got together with others who were also well informed about local issues. Thirty-six percent reported that they worked to bring about change in their community on a daily basis. Slightly less than half of the sample (48%) reported that they either posted comments online, posted pictures or video online, or blogged about a political or social issue in the past year.

The overwhelming majority (80%) of respondents reported having a profile on at least one type of social media website (social networking, blog or microblog, photo/video collections, place-based applications, or other). All of these profile-users maintained a profile on a social networking site, with many having profiles on multiple types of social media sites.

Respondents used social networking sites more frequently than other types of social media sites. Fifty-six percent of the sample used social networking sites on a daily basis, and 76% used these sites at least once a week. Place-based applications were the least used type of site. Of the 5 individuals who used these applications, none used these sites more than once a month. Most respondents accessed these social media sites via personal computer (96%) and many used their cell-phones as well (68%).

Social media use was fairly well distributed across types of social media sites, with the exception of place-based application (social networking sites 56%, blog or microblog 44%, and photo/video collection 40%). All in all, 64% reported using social media sites to communicate with other members of their organization, with several respondents utilizing multiple types. Fifty-two percent reported using social media sites for such purposes at least once a week.

The sample was generally satisfied (88%) with current emergency response efforts in their community. All respondents felt that the county government should contact citizens by way of phone call or text message during a crisis. Eighty-four percent felt that social networking sites should also be utilized for this purpose, and 72% felt that blogs or microblogs should be as well. Of the sample, 56% reported that they were at least somewhat likely to use one or more types of social media to contact family members during a crisis. However, only 24% were likely to report a crisis to local government agencies via social media sites. The majority of respondents still report that talking to others in person or by telephone is the most important source of local information.

4.2 Focus Groups: Information Factors

In the electronic brainstorming step of the interviews, focus group participants identified 23 categories of factors related to 1) the organization and 2) the information exchanged between the organization and community (Figure 3).

Information factors include issues related to the quality and quantity of information generated through SM. They also include the tone of and types of communications in which government desires to participate, including outreach, feedback, and two-way communications. Additional types of information that can be obtained from some SM channels, e.g., detecting the locale of emerging events, are of substantial interests for emergency management and policing functions. Finally, the security of technology used to provide SM capabilities and new tools needed to meet legal obligations for saving public records comprise a set of technology issues that contribute to the information factors. Lastly which existing SM tools should be utilized remains a substantial question across the focus groups.

Together the factors identified by the participants describe a broad range of interests and concerns of the Arlington County government in relation to their use of SM. Each of these categories also contains a set of ideas from the electronic brainstorming that further clarify the intentions of the participants about the meaning of the categories.

4.3 Focus Groups: Organization Factors

The organization factors that focus group participants identified include policies, legal issues, costs, and training (Figure 3). The organization requires that policies be adopted to provide the environment needed for employees to achieve work objectives. Management buy-in is essential if benefits are to be realized and costs are to be controlled. To utilize SM effectively the activities and roles implemented are institutionalized through Human Resources (HR) developing job descriptions and ensuring related types of communication are managed effectively. There are attempts to control information and to communicate the government's opinions and actions that members of the public would want to be aware of.

Organizations also seek to define the types of information to be shared and the manner in which it is shared. The participants perceive the substantive legal issues related to maintaining government transparency, often through the Freedom of Information Act (FOIA), as important considerations of using SM. For example, should tweets by a government employee be part of the public record? What about tweets by a government employee that are related to their non-work life? The individual government employee would need to set up two different identities in Twitter, in order to separate professional and private roles.

Costs are always important to organizations, and government budgets have been squeezed due to reduced receipts resulting from slowing economic activity and increased use of government services. Yet the participants perceive that the potential exists for achieving efficiencies using SM and the potential return on investments should be evaluated. Complicating this calculation is the value placed on reaching previously uninvolved constituents and the most interested participants. One of the costs of adopting SM is the training of the employees that will conduct the activities. In addition, the public must be educated to understand how the government will interact with them and what expectations for interaction are appropriate.

Some of the Arlington County focus group participants said that they need social media aggregation tools. In general, dashboard services that accept search keywords and phrases help monitor information from multiple social media, such as trackur (<http://www.trackur.com/social-media-monitoring>) and Netvibes

(<http://www.netvibes.com/>). But these tools are designed to support businesses not government or citizens, so they are not optimal for civic needs. Having geo-mapping features would be very useful for the needs of cities and communities, which are not currently enabled in dashboard tools.

Some emerging applications allow citizens to contribute geo-tagged photos and video to a community database. For example, MIT's Mobile Media Experience Laboratory has developed a place-based application called Locast for this purpose (<http://www.locast.mit.edu>). The video analytic software IBM has developed will help to organize and cluster images of similar content or location. This would make it easier for users to find content of interest and to contribute to ongoing information exchange regarding a particular issue related to a specific place (e.g., building a new school).

- ❖ **Information Factors**
 - Communications
 - Community Outreach (emergency, crime/traffic alerts, 24/7 level of service, recruitment)
 - Feedback (from community to organization, social trends, locale, fast spreading ideas)
 - Population Reached (misses traditional/older population or can't afford technology)
 - One Way vs Two Way (pushing out vs creating dialogue, effort/costs different)
 - Tone (Government presents just the facts, not stories, not press release, listen then educate)
 - Information
 - Quality of Content (accuracy, facts of situation, un-vetted information, misinformation)
 - Quantity of messages (how to be heard, from 1 to 10 to 1000s, overwhelming, loss of control)
 - Personal Level (information overload, ability to write complete thoughts, nuances of face-to-face lost)
 - Technology
 - Security (network exposed to world)
 - Technology and Equipment (cost of technology and maintenance, cost savings, training)
 - Social Media (SM) Outlets (knowing audience/expertise, users expect transparency, so many outlets)
 - Public Record/FOIA (are SM public record, tools needed to save, outdated polices)
- ❖ **Organization Factors**
 - Policy
 - Management Buy-In (unknown expectations, under valued, need to set culture)
 - Control Issues (how much to control, what we can control, telling how/what to think/do)
 - Human Resource (HR) Components (job descriptions, evaluation, expertise, dialogue, positive and negative)
 - SM Communications Policy (what not to do/say, right people to make SOP, moving target)
 - Professional Level (privacy concerns, devices owned by county, investigative purposes)
 - Legal Issues
 - Data Maintenance (FOIA data maintenance and related costs)
 - Owing Vs Using Someone Else (official outlet versus imposter, use in investigations.
 - Public Record/FOIA (are SM public record, tools needed to save, outdated polices)
 - Costs
 - Resource Issues (SM adds to previously full time job, other duties, limit 24/7 expectation)
 - ROI/cost to value (how to measure value, who are we reaching, enough received messages)
 - Training
 - Education (tools to manage, learning from each other, train constituents where to go)
 - Training (best practices for dividing duties, case studies, understanding management's concerns)
 - Other (educate nonusers, establish boundaries)

Figure 3. Simple taxonomy of categories identified by focus group participants

Some participants in the focus groups also indicated that recent or projected budget cuts could erode 15 years of community outreach; the County wants to understand how to use technology to maintain and sustain established communications with citizens.

The neighborhood/civic associations have been key to the community outreach in the past, but not all neighborhoods have homeowners associations. Such residential neighborhoods are

usually characterized as having a lower socio-economic population; with budget cuts it is harder for government to sustain routine outreach and communication with these areas of the city. Social media may be particularly helpful for outreach to such households and neighborhoods, especially in combination with cell phones.

Preliminary evidence from a national study by Pew Internet & American Life (Smith, Verba, Brady, & Schlozman, 2009) indicates that the use of social media for civic purposes is not as strongly correlated with education and income as traditional internet use (web browsing and email). This may be because opinions leaders (i.e., influential individuals) exist at all social strata, and they may convey information to members of their social circles not only face to face, but also by cell phone. Cell phone ownership permeates all social strata and exceeds computer ownership among lower SES groups.

The cell phone is essentially a pocket computer. For lower SES groups it is likely to be the only computer they are using. While we were not able to study cell phone use among lower SES groups in this pilot study, we will be investigating cell phone use for civic purposes among different demographic groups in future research. We plan to investigate the possible use of cell phones to address information needs, and their connection to social media, especially text messaging and image sharing among lower socio-economic populations.

4.4 Tools for Analyzing Social Media Data

In order to study the pattern of communication and the information communicated using social media, we collected publicly available data from Twitter. We identified 34 civic organizations, some of which are government agencies, in the NCR that were tweeting; we collected and analyzed their tweets for 30 days between September and October 2010.

We analyzed the tweets as well as the biographical information posted as profiles of the organizations' followers using Natural Language Toolkit, tag clouds, and graphs. Figure 4 shows the number of followers for the 34 civic organizations.

If we look into the number of followers of these followers, however, we see the extensibility of the communication chain radiating out beyond the organization originating tweets (Figure 5). Further analyses show us which words are used most commonly in the tweets or bios during this period. The predominance of various words (most common words appear larger in a tag cloud) provides a quick overview of what is being said or characterized (in the case of followers' bios).

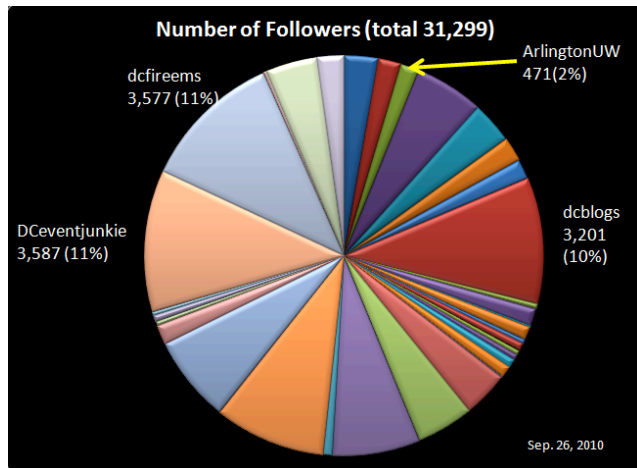


Figure 4. Number of followers for 34 NCR civic organizations

For the 34 civic organizations that were tweeting during the September – October 2010 period, we see there are a total of about 31,000 'direct' followers (i.e., people who subscribe to the RSS feed that carries each organization's twitter posts). What is

interesting to note is that the 'direct' followers are themselves being 'followed' by other people – what we refer to as 'followers of followers' (Figure 5). The number of followers of followers for these same organizations is over 67 million.

For an organization such as Arlington Unwired (Arlington UW), that disseminates announcements about local events, shown with an arrow in Figure 4, there were 471 followers on the date we captured these data (September 26, 2010). We can see from the analysis of the number of Arlington UW followers' followers (Figure 5) there are over 8 million followers. This is not to say that a tweet from Arlington UW will go beyond the 471 direct followers; however, if there is a crisis in the Arlington area (such as a major catastrophe or extreme violence) it is very likely that the indirect followers will retweet (forward along the same twitter post) regarding such a catastrophe to their own set of followers (i.e., over 8 million followers). In this way, we can see the potential reach of a critical piece of information being disseminated throughout a community way beyond the direct Twitter followers to a larger population of followers' followers.

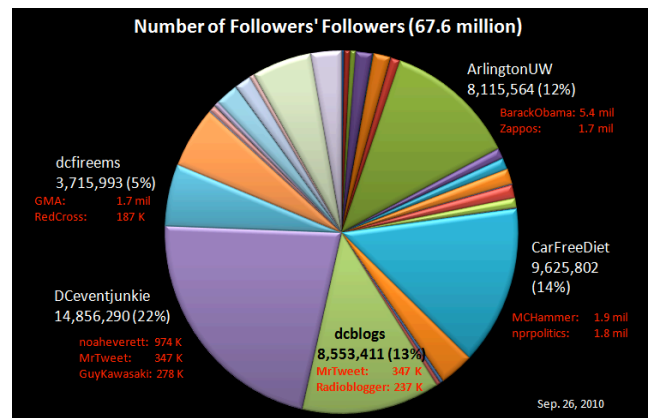


Figure 5. Followers of Organizations' Followers

It is also important to note that among the followers of Arlington UW is 'Barak Obama' – and the number of followers of 'Barak Obama' is over 5 million.

In order to get a sense of who are the followers of these 34 civic organizations in the National Capitol Region, we collected the publicly available biographical profile information that followers list on their own twitter accounts. Figure 6 shows in a tag cloud the predominant descriptors that followers' use in their profiles.



Figure 6. Profile biographies of Arlington UW followers

This kind of analysis and visualization provides quick overview of the type of individuals and their interests who are following a given organization. Followers of tweeting organizations or individuals are likely to be tweeting themselves. We analyzed and visualized in a tag cloud the 20 recent tweets of the followers of Arlington Unwired (UW) at the time of the data capture (September 26, 2010) shown in Figure 7.

By looking at the recent tweets of followers, we see a kind of ‘mood’ and ‘buzz’ among users. The large ‘RT’ stands for re-tweet, meaning that this is the most common term appearing in the twitter posts for these users. The organization knows from this analysis that many of their posts are going well beyond their immediate (direct) followers.

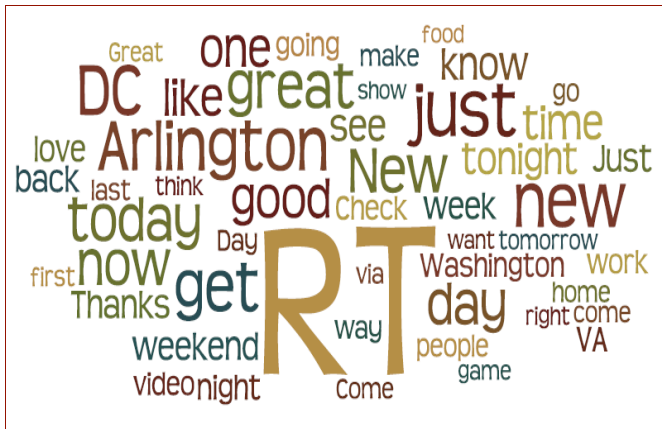


Figure 7. Twenty recent tweets by followers of Arlington UW

The purpose of these analytical and visualization tools, as noted earlier, is to allow government and citizens to see quickly and easily the big picture of the information and communication flows that interest them.

4.5 Analysis of Facebook Comments

Arlington County government has maintained a Facebook page since early 2010 (<http://www.facebook.com/ArlingtonVA>). The page had roughly 4500 fans at the end of September 2010 (by February 2011, there were close to 4900 fans). We analyzed a two-month period (August -September, 2010) of posts by the County and responses (comments) from the public by conducting a simple content analysis by topic. There were a total of 112 posts; the top 10 most frequent topics are shown in Figure 8.

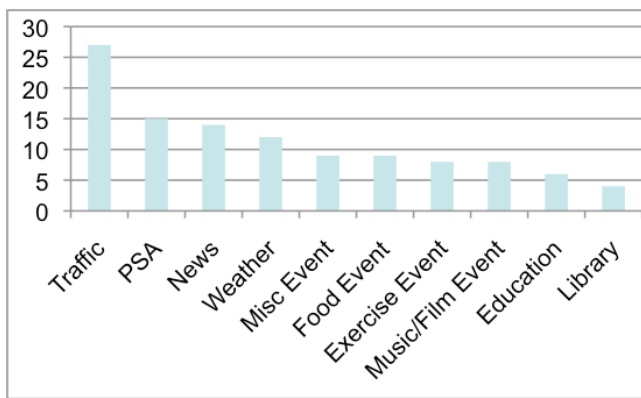


Figure 8. Facebook Topics Arlington County

The most common posts by the County on the Facebook page were about traffic (e.g., conditions, closures, metro outages), followed by public service announcements (PSA). News (shout-outs, updates, and other County announcements) and weather related posts (National Weather Service and Arlington Weather Service advisories) were followed by various events (good, bikings, walking, music or film) in terms of frequency of posts. There were only a few posts related to education (Arlington County School District) and library services (e.g., closures, speakers, special activities) during this two month period.

There were a total of 824 public comments to the County posts during this two-month period. Half of the comments pertained to about a fifth (19%) of the County posts (the top 21 posts by the County). Figure 9 shows the distribution of the bulk of the comments on the same top 10 County posts seen in Figure 8.

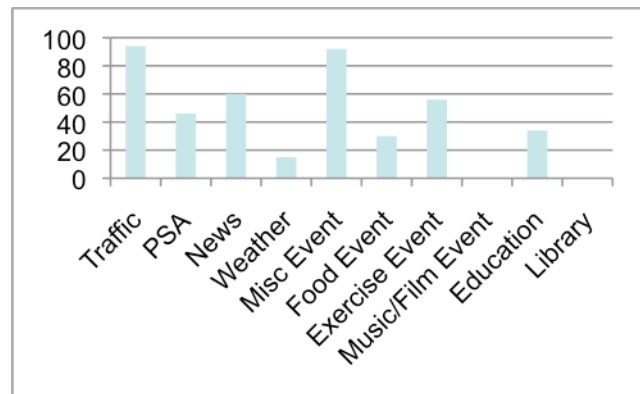


Figure 9. Public Comments by Arlington Facebook Topic

The comments are predominantly related to traffic and miscellaneous events (that is, events that do not fall into the other ‘event’ categories shown, such as food, exercise, music and film). Exercise events (bikes, walks) and News announcements generated the next most frequent number of comments from the public.

Almost all the comments were highly consistent with the social media policy of the County (e.g., no profanity or off topic comments) and were overwhelmingly positive in tone, including many “Likes” hits.

Lastly, we collected videos in YouTube pertaining to Arlington, Virginia and conducted a tag analysis of the video collection using image software developed by IBM. We performed a search using Perl script and the phrase ‘Arlington County;’ this produced about 1800 videos from YouTube. We then developed two types of tag clouds generated using video titles and video tags (see Figure 10).



Figure 10. Tag Cloud of Arlington YouTube Videos

The tag cloud as a visualization quickly and easily represents the frequency with which different terms appear in the search thereby providing a snapshot of what is in the large dispersed collection. The more frequently a term appears in the image collection, the larger it appears in the tag cloud. The cloud visualization also provides an indication of the importance of various civic issues to members of the community. The recurring civic themes revealed in the video analysis can be further explicated in the six categories shown in Table 1.

Table 1. Tag Cloud Categories for Arlington Videos

Law enforcement	Police, cops, officer, courthouse, robbery, accident, ACPD, surveillance
Transportation	Metro, street, boulevard, highway accident, parking, transit
Social issues	Environment, diversity, community, city, neighborhood, accountability
Economic development	Growth, sustainability, development, bank, private, local
Political	Government, elections, agencies, department
Communication	Media, ABC, NBC, CBS, television, news, network, bilingual, NoVAPJ, Spanish

The further clustering of video tags and video titles as shown in Table 1 allows government and other users to make sense more easily of the interests and needs of the community as expressed in the YouTube collection at any given point.

5. DISCUSSION AND IMPLICATIONS

The pilot study was intended to advance technologies and systems for social media analysis relating to both routine day-to-day civil life and critical incidents or emergencies. The results begin to address a combination of technical and social science challenges; on the technical side, these include:

1) recognizing relevant information accurately and in a timely manner, especially short content from micro-blogging sites (e.g., Twitter); the limited information in a tweet (i.e., less than 140 characters) makes it difficult to identify its meaning and context

which may lead to incorrect classification and misleading analysis of tweet data;

2) alerting government officials to the analyzed information from multiple social media sources; due to the massive volume of the social media data stream, it is a challenge to quickly analyze the collected information from different sources and to make a decision based on the analysis; and

3) visualizing the current and past status of incoming information and the analysis of it; simple yet informative visualization design is essential in making-sense of the data presented. We support the sense-making process by incorporating interaction methods with the visualization to deal with the large amount of data.

On the social science side, our pilot study results build on social network analysis and social and political participation research on the use and of social media. We also seek to contribute to crisis informatics research and an understanding of the use of social media in crisis situations, including more mundane crises, such as weather or traffic problems, and in social convergence situations, such as crowds, rallies and other large gatherings which are not unusual in the National Capitol Region.

In future research we seek to address the benefits and limitations of using digital libraries (DLs) for this work, specifically: 1) having a common platform for collaboration among project team members and contributors; 2) adapting DL mechanisms for searching and browsing as well as collecting and archiving streamed resources; and 3) developing communities of interest.

Our social media data analyses are intended to help government and citizens of Arlington County and the National Capital Region (NCR) know how and where to get useful information and critical communication in the event of a crisis or social convergence condition. Our tools should help government and citizens monitor and make sense of the diversity of voices and information that enrich the quality of life in their communities. Tools we are developing will be available in open source for government and citizens to help them find information clustered by topic or place and to further contribute, discuss, and interact with each other.

We focus on Arlington and environs (around Washington, DC) as our test case in order to analyze information, its use and impact related to local, state, national, and international events – since it has close connections to the US capitol. Our tailored digital library should serve as an advanced information system, with a broad range of services that allow us to build a self-sustaining collection that can be suited to a particular community. To ensure scalability, and to move toward sustainability, the digital library resources we develop are being made available to the public so that interested members of the community will continuously update it for broader use.

By mining content and services covering multiple media types (i.e., text, audio, image, and video) we can develop tools to recognize events and alert government, citizens, and community groups to see quickly the ‘big picture’ through visualizations of social media activity and content and changes in both over time. The intent is to enable proactive responses, as routine problems or crises start to loom, as events unfold, as individuals and groups respond, and as plans (short or long-term) are made for improved services and communication. Such capabilities are relevant to a broad range of governments throughout the US and globally.

Given the efficiency of communication coupled with the potential to reach many constituents personally and quickly provided by

social media, it seems clear that governments should seek to understand and to leverage these communication channels.

6. ACKNOWLEDGMENTS

We are grateful for support for this research from the Virginia Tech Center for Community Security and Resilience (CCSR) and National Science Foundation (IIS-0916733). We would like to thank the government officials and representatives and the citizens of Arlington, Virginia who participated in the six-month pilot study.

7. REFERENCES

- [1] Arlington County Blog Central. Retrieved Dec. 14, 2010, from <http://www.arlingtonva.us/departments/Communications/page70455.aspx>
- [2] Arlington County Blog Central. Retrieved June 4, 2010, from <http://www.arlingtonva.us/departments/Communications/page70455.aspx>
- [3] Arlington County Facebook Profile. Retrieved May 14, 2010, from <http://www.facebook.com/ArlingtonVA>
- [4] Arlington County Flickr account. Retrieved Dec. 14, 2010, from <http://www.flickr.com/photos/arlingtonva>
- [5] Arlington County News on Tweeter. Retrieved Dec. 14, 2010, from <http://twitter.com/arlingtoncounty>
- [6] Arlington County, VA Official Site. Retrieved Dec. 14, 2010, from <http://www.arlingtonva.us/default.aspx>
- [7] boyd, d., & Ellison, N. (2007). Social network sites: Definition, history and scholarship. . *Journal of Computer-Mediated Communication*, 13(1), article 11.
- [8] CTRnet: Community, Tragedy, and Recovery network. from <http://www.ctrnet.net/>
- [9] Flickr. from <http://www.flickr.com>
- [10] Furnas, G. W., Fake, C., von Ahn, L., Shachter, J., Golder, S., Fox, K., et al. (2006). Why do tagging system work? . Paper presented at the Computer-Human Interaction.
- [11] Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. Paper presented at the Information Systems for Crisis Response and Management, Gothenburg, Sweden.
- [12] Hughes, A. L., Palen, L., Sutton, J., Liu, S., B., & Vieweg, S. (2008). "Site-seeing" in disaster: An examination of online social convergence. Paper presented at the International Conference on Information Systems for Crisis Response and Management (ISCRAM), Washington, DC (May).
- [13] Lampe, C., Ellison, N., & Steinfield, C. (2006). A face(book) in the crowd: social searching vs. social browsing. Paper presented at the Conference on Computer Supported Cooperative Work, Banff, Alberta Canada.
- [14] O'Reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies*, 65(1st Quarter), 17-37.
- [15] Opsahl, A. (2010). County Experiments with Monitoring Social Media in Emergencies. *Emergency Management* Retrieved October 22, 2010, from <http://www.emergencymgmt.com/disaster/NC-County-Monitoring-Social-Media-Emergency.html>
- [16] Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors, WWW '10: Proceedings of the 19th international conference on World wide web (pp. 851-860). Raleigh, NC, USA: ACM.
- [17] See-Click-Fix in Arlington County. Retrieved June 2, 2010, from http://www.seeclickfix.com/arlington_2
- [18] Smarter transportation: 10 social media tools to navigate your city. from [http:// mashable.com/2009/09/29/social-media-transportation](http://mashable.com/2009/09/29/social-media-transportation)
- [19] Smith, A. (2010). Government Online: Pew Internet & American Life Project.
- [20] Smith, A., Verba, S., Brady, H., & Schlozman, K. (2009). The Internet and Civic Engagement: Pew Internet & American Life Project.
- [21] Social Media Sells. (2010). Ogilvy & Mather Retrieved November 1, 2010, from <http://www.ogilvy.com/News/Press-Releases/November-2010-Social-Media-Sells.aspx>
- [22] Tepper, M. (2003). The rise of social software. *netWorker*, 7(3), 18-23.
- [23] Twitter blog: Measuring tweets. Retrieved November 4, 2010, from <http://blog.twitter.com/2010/02/measuring-tweets.html>
- [24] YouTube. from <http://www.youtube.com>
- [25] Zuckerman, E. (2009, April 13, 2009). Studying Twitter and the Moldovan protests. My heart's in Accra Retrieved January 6, 2009, from <http://www.ethanzuckerman.com/blog/2009/04/13/studying-twitter-and-the-moldovan-protests/>