AAAI Technical Report WS-13-04
When the City Meets the Citizen 2

Aggregating Community Information to Explore Social Connections

Kyungsik Han, Patrick Shih, John M. Carroll
Information Sciences and Technology, The Pennsylvania State University
{kuh178, pshih, jcarroll}@ist.psu.edu

Abstract

We introduce our ongoing work in community networks and present a system that aggregates geographical community news and events information from local news and social media. Our goal is to present community information according to its content and to make it more visible to local residents, thereby increasing community awareness. We also present a data analysis of how local residents are connected by analyzing the usage of their tweets and discuss future directions for our system.

Introduction

Research on community networks has attempted to assess the impact as well as the potential of computing resources when applied to geographic community goals, such as information dissemination, discussion, and collective activity (Schuler, 1996). Technologies such as Web 2.0 have suggested new community-oriented approaches to foster such networks for creating and distributing content.

Accordingly, investigating how technologies contribute to the strengthening of community identity, enhancement of residents' awareness, and promotion of participation in activities in the community environment has become an essential element of community networks research. Such examples include understanding the contribution of Internet connectivity in public spaces with respect to social interaction and participation (Hampton & Livio, 2009) and investigating the benefits of ubiquitous computing in community infrastructure, service supports, education, culture, recreation (Carroll & Rosson, 2008), etc.

Along with these theoretical insights, much research has been done on the development of systems that harness the benefits of supplementing community activities with technologies. One of the main goals of system development is to make community information more visible to local residents, which also leads them to be more aware of their community. One example that has gained much attention recently is the presentation of user-generated content from social media sites (e.g., Twitter) on large screens in a public space to deliver one's messages to a broad audience (Ryan, Hazlewood, & Makice, 2008).

Our group also investigated how technologies can be used to increase community awareness in a local area. We studied and observed the practices of different community organizations (e.g., nonprofit community groups, non-governmental social service providers, and most local levels of government) with regard to technology adoption and use and found that many local organizations had their own websites and provided information to local residents through this means (Carroll et al., 2011).

However, when we examined each website and service more closely, local community online interactions often failed to exceed the threshold of community perception (Carroll, Hoffman, Robinson, Han, & Rosson, 2013), because local information is accessed primarily, or perhaps exclusively, by local people. With the exception of the local news media websites, there are only few posts from other local community websites. There are even fewer comments, including comments on local news websites, when compared to other national news channels such as CNN.com. Therefore, unless there is systematic support, a great deal of community information will not be visible to residents. In that sense, we have begun to explore the potential of web-based feeds such as RSS and iCalendar, with the goal of surpassing the threshold of community perception regarding community information by building a system to achieve this goal (Carroll et al., 2011).

Although our system contributes to the increases in the visibility of community information, we have observed several limitations. First of all, aggregating community information does not necessarily entail people's reactions to content. For example, it is still hard to find people's comments to news articles. In addition, we have placed too much emphasis on providing information as much as possible and less on understanding its content. Lastly, the system does not leverage social media, such as Twitter or Facebook. Because so many people use these social platforms

to express their personal thoughts or opinions on certain topics (Bollen, Pepe, & Mao, 2009), it is important to consider them as critical additional channels for community information.

In this paper, we introduce our design extension to address these issues. Our design rationale is to extract a set of keywords from local news articles and to collect social media content (we use *Tweets* from Twitter) based on keywords. If we can find a link between local news articles and tweets, and present both resources in an integrated manner, tweets will be reframed as *comments* to the news by people. In this sense, community information will be represented by its content and enriched by aggregating two resources.

We explain how our system presents community information by a keyword (or popular point of reference) and has the potential to allow people to easily notice what is happening in their local community. Moreover, we present some of our findings based on the data we have collected during a one-week trial period. Because the system is still in its infancy, we also discuss its limitations and future direction.

Local Community Information Aggregator

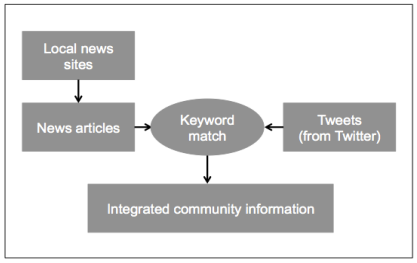

Figure 1. Overview of local community information aggregator.

System overview

Figure 1 represents an overview of the local community information aggregator. Regarding local news sites, our assumption is twofold. First, local news sites are the most active community news providers when compared to other organizations. Second, if several news media cover the same news topic, or if multiple news articles about the same topic are published, that topic may be considered important or popular.

Accordingly, we have used five local news sites to collect local news articles. We have designed the system to collect news articles from each local news site and to store them in our local database on an hourly basis. Each news item contains a set of metadata, including a title, description, source URL, and created date and time.

After a typical pre-processing step in NLP (e.g., stop word removal), the keywords are extracted from news titles and descriptions, and their TF-IDF (term frequency-inverse document frequency) scores are calculated. The TF-IDF score is a standard metric in Information Retrieval to measure the "importance" of a keyword. The TF-IDF value of a keyword *w* in an article *a* is positively correlated with the frequency of *w* within *a* but is negatively correlated with the frequency of *w* in the entire article collection.

At the end, top-*k* keywords with the highest TF-IDF scores are used as a parameter, along with a geo-coordinate, and submitted to a Twitter search API. Twitter returns a set of tweets that are pertinent to the provided keyword and location. The call for this request is also operated on an hourly basis, and the results are stored in our local database. Because the data from local news media and tweets are already connected by "keywords," we are able to present this information together in a unique way.

Figure 2. Screenshots of the app. Tag clouds (left) and combined view of news items and tweets (right).

Mobile application

We are interested in presenting community information on a mobile screen, because of the idea for community members to access the aggregated content while on the go. We have created APIs to support the communication between server and mobile client, and built a simple mobile app that displays community information.

Figure 2 features an example of how community information is displayed on a mobile screen. The first page of the application shows tag clouds (Figure 2, left). Each tag represents a keyword that has been used in local news articles, and the size of the tag cloud represents the number of articles mentioning the keyword. Because each news article contains date and time information, tag clouds can also be regenerated by different time variances, such as the day, week, or month. If one of the keywords is clicked, the corresponding news article and tweets by local residents will be displayed (Figure 2, right). The news article is displayed at the top of the page, and tweets are displayed below the article.

Since there can be multiple articles that mention the same keyword *w*, one design issue is to decide which article should be displayed when *w* is clicked (e.g., oldest vs. latest article). By exploiting the TF-IDF scores computed for each keyword per article, in current implementation, we

by default show the article where the keyword has a higher TF-IDF score.

Figure 2 (right) illustrates community news about the calls to cancel annual drinking activities surrounding St. Patrick's Day (one week before Spring break), because several incidents, including excessive drinking, property damage and criminal behavior, have caused major problems in this community over the last several years. Because this was the first attempt to regulate the students' annual event, many news articles were published to provide a detailed description of the proposed regulations and reactions from students, local residents, and local businesses, in addition to interviews with university authorities, student organizations, and so on. Regarding social tweets, as seen in Figure 2 (right), people simply created short messages to express their opinions on the proposed regulations.

In summary, we believe that our design approach presents a unique way to represent and deliver community information by identifying popular keywords using the TF-IDF metric from local news media (which are not randomly chosen topics), representing popular local news coverage. Our approach also strives to combine local news articles and socially generated tweets based on keywords and location to present rich and more dynamic community information. As it is generally difficult to see people's comments on local news articles, tying tweets into comments and attaching them to news articles increases the visibility of people's reactions and creates a new venue with respect to accessing local community information.

Exploring Connection Among Users

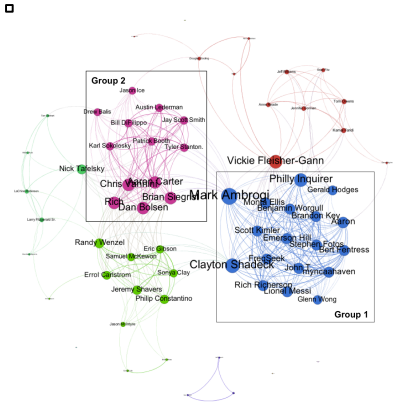

Figure 3. Social network analysis of individuals.

In recent years, we have witnessed a significant trend, in which people create and share community information by taking advantage of social media platforms. This aspect is also consistent with the amount of tweets that we have collected over a one-week period. Along with combining community information from two resources, we are also interested in exploring how people are socially connected with each other by analyzing data stored in our local database. In this section, we articulate our approach to exploring these social connections.

We realize that different entities, including individuals, organizations, groups, clubs, etc., create tweets every day. Because we are interested in connections among people, we have only considered tweets from individuals. Moreover, to reduce the size of data, we have excluded individuals who have tweeted only once. As a result, we have analyzed 416 tweets from 67 distinct users.

As described before, we use keywords from news articles to retrieve tweets from Twitter. Because we have also collected user information from each tweet, we are able to link users based on keywords. Our assumption here is that if the tweets from two users refer to the same topic, then the two users might have a similar interest. Taking it a step further, if there are multiple news articles with different topics that the two users have tweeted about, they might have more common interests or perhaps more connections. In this sense, it will be interesting to see how people are linked to each other based on the patterns of their tweets and to explore design implications based upon these results.

We ran a social network analysis tool, Gephi, to visualize the said connections (Bastian, Heymann, & Jacomy, 2009). As depicted in Figure 3, each node represents an individual Twitter user, and each edge represents the connection between individuals. If two Twitter users mention the same keyword at least once, an edge is created between them. The size of a node is correlated with the number of connections between one and the other, and the width of an edge reflects the degree of connection between two individuals.

Nodes are clustered into several groups, and people in the same group have tweeted on the same topics. For example, we found that people in Group 1 are grouped by the keywords "Michigan" and "Basketball," which are related to the recent basketball game between Penn State and Michigan. The game has excited many people, as it was Penn State's first win over a top-five team since 2001. Moreover, we found that the tweets from Group 2 are based on the keywords "NCAA" and "Freeh." This points to the ongoing debate surrounding the penalty to Penn State by the NCAA because of the Jerry Sandusky child sex abuse scandal that garnered national media attention in 2012.

Within a group, we found that the degree of connections varies by individuals. A thicker edge between two individuals shows that they have tweeted several times about the same topics. Perhaps this indicates that these people might have more motivation to express their thoughts or share related information with others. There are also a few individuals (e.g., Mark Ambrogi or Vickie Fleisher-Gann in Figure 3) who have some connections to people in other groups. It is interesting to note that the two individuals

have a number of edges to people in other groups, but each edge is weaker than those of people within the same group. This might indicate that the two individuals are broadly interested in different news in their community.

Discussion and Conclusion

In this paper, we have introduced our design of an innovative community information aggregating system, which leverages local news articles and user-generated Twitter messages from social media. The system has great potential to make community information more visible and to allow people to access popular topics in their community. Because we realize that online interactions in a local community are very sparse and only few posts or comments are created (except a few metropolitan cities), we believe our mechanism, which utilizes tweets as comments attached to news articles, presents people's reactions and provides a rich amount of community information in a unique way. Because the current version of the mobile application focuses exclusively on the provision of information, we will extend the system's functionalities, such as adding an interface that will allow people to add comments or engage in discussions from the mobile application.

Along with the design investigation, we also present our analysis of how users are connected based on their use of tweets. Our analysis indicates that there are groups of individuals who have common interests, and some within a group have stronger connections than others. There are also some individuals who have multiple connections to people in the same group as well as in other groups. It will be interesting to investigate how people actually perceive this feature and whether it makes them feel connected in the sense of knowing that other people have similar interests. The findings also support the creation of a recommendation feature designed for recommending individuals who share similar interests in the community.

We can extend our system further by addressing two issues. First, there are some mismatches among the tweets. Even if they were searched using the same keyword, we found that some described different news or events due to ambiguous meaning of words Although this could be resolved by building a better algorithm to identify the degree of relevance, we are more interested in handling it heuristically. More specifically, we could utilize users' inputs by adding an interface to enable users to flag relevant or irrelevant tweets. This would lead us to further explore how people perceive community information in different ways by analyzing their inputs of relevancy rating. Second, our system only uses keywords from local news media, which might undermine the opportunity to introduce other interesting community news or events from additional resources. We could consider other possibilities, such as using top ten searched keywords in the U.S. (e.g., from Google search), and review how local people tweet about them and if our system is capable of helping people find others who share common interests. We could also design the system to ask users to define a keyword and then display results that are relevant to their local community.

Overall, although our system is still in its infancy, we believe that it provides a unique lens to assist in better understanding people's awareness of and their social connections within a local community, which also complies with the goal of strengthening community networks. Regarding activities from social media, there are a number of research efforts to explore the patterns of tweet usage; however, most of these focused on major national events, such as elections, major incidents or disasters (Vieweg, Hughes, Starbird, & Palen, 2010; Yardi & boyd, 2010). Significantly less attention has been given to everyday community news or smaller events. Using our system, we would like to further explore every day community activities and how technology can be leveraged to better understand the connections between local people and their community. Our system could also be used in other geographical communities and offers the potential to present a rich amount of community information to local residents.

References

Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. In *Proc. ICWSM '09*.

Carroll, J.M., Rosson M.B. (2008). Theorizing mobility in community networks, *International Journal of Human-Computer Studies*, 66(12), pages 944-962.

Carroll, J.M., Horning, M., Hoffman, B., Ganoe, C., Robinson, H., Rosson, M.B. (2011). Visions, Participation and Engagement in New Community Information Infrastructure, *The Journal of Community Informatics*, 7(3).

Carroll, J.M., Hoffman, B., Robinson, H., Han, K., Rosson, M.B. (2013). *Hyperlocality and Suprathresholding in Community Network Designs*, In *Proc. CHI '13 Workshop on Human Computer Interaction In Third Places*.

Hampton K., Livio O., and Goulet L.S. (2010). The Social Life of Wireless Urban Spaces: Internet use, Social Networks, and the Public Realm, *Journal of Communication*, 60(4), pages 701-722.

J. Bollen, A. Pepe, and H. Mao. (2009). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena, In *Proc. ICWSM '09*.

Ryan, W., Hazlewood, W.R., Makice K. (2008). Twitterspace: A Co-developed Display using Twitter to Enhance Community Awareness, In *Proc. PDC '08*, pages 230-233.

Schuler, D., New community networks: Wired for change, Reading, MA: Addison-Wesley, 1996.

Vieweg, S., Hughes, A.L., Starbird, K., Palen, L. (2010). Microblogging during two natural hazards events: what twitter may contribute to situational awareness, In *Proc. CHI '10*, pages 1079-1088.

Yardi, S., boyd, d. (2010). Tweeting from the Town Square: Measuring Geographic Local Networks, In *Proc. ICWSM '10.*