Putting the Public into Public Health Information Dissemination: Social Media and Health-related Web Pages

Robert Steele
Discipline of Health Informatics
The University of Sydney
Sydney, NSW, Australia
robert.steele@sydney.edu.au

Dan Dumbrell
Discipline of Health Informatics
The University of Sydney
Sydney, NSW, Australia
ddum7449@uni.sydney.edu.au

ABSTRACT

Public health information dissemination represents an interesting combination of broadcasting, sharing, and retrieving relevant health information. Social media-based public health information dissemination offers some particularly interesting characteristics, as individual users or members of the public actually carry out the actions that constitute the dissemination. These actions also may inherently provide novel evaluative information from a document computing perspective, providing information in relation to both documents and indeed the social media users or health consumers themselves. This paper discusses the novel aspects of social media-based public health information dissemination, including a comparison of its characteristics with search engine-based Web document retrieval. A preliminary analysis of a sample of public health advice tweets taken from a larger sample of over 4700 tweets sent by Australian health-related organization in February 2012 is described. Various preliminary measures are analyzed from this data to initially suggest possible characteristics of public health information dissemination and document evaluation in micro-blog-based systems based on this sample.

Categories and Subject Descriptors

D.3.3

General Terms

Documentation.

Keywords

Twitter, Web documents, Public Health

1. INTRODUCTION

The role of the Internet in enabling document retrieval and dissemination has affected vast change in the past 15-20 years. Relatively recently, searches for certain types of information were (and still are) usually done via search engines resulting in the ranked presentation of the algorithmically calculated most relevant Web documents. With the introduction and development of social media platforms however there has been some change in the discovery and retrieval aspects of Web documents.

Public health information dissemination involves communication of disease prevention and health promotion information through organized efforts. While a majority of Web-based public health information dissemination and retrieval has been done via search engines, it has been found that relevant public health documents were not always successfully located and disseminated via search engines due to the query behavior of the user [1].

The novel approach of utilizing social media for public health information dissemination and retrieval has recently been explored [2]. The rapid and widespread uptake of social networking sites (SNSs) such as Twitter allows for the sharing of public health-related information to result in more up-to-date information dissemination due to the instantaneous and 'push' nature of the application. Twitter, a widely used micro-blogging service, has characteristics that make it a useful tool for information dissemination and retrieval, such as instantaneous 'tweeting' (postings of 140-character limited updates), 're-tweeting' (forwarding of other users' tweets) and the ability to publicly interact with other users and their tweets. Unlike the case with search engines, documents in Twitter are discovered via embedded URLS in received tweets.

When dealing with public health information dissemination via Twitter, there are a number of objectives and goals that are beyond simply retrieving relevant documents that match queries based on information retrieval metrics. These co-occurring objectives include: the aim of widespread dissemination of quality health information; dissemination to targeted groups and individuals so that the information is reaching those it is relevant to; achieving 'push' communication with the ultimate goal of positively affecting the health behavior of recipients; involving users with cognate interests to interact and communicate; as well as providing up-to-the-minute information.

The purpose of this paper is to provide a preliminary overview of micro-blog-based public health information dissemination and its novel document retrieval and evaluation characteristics. A preliminary experiment to examine relationships between public health advice re-tweeting behavior and the nature and authority of the Web pages pointed to by embedded URLs in public health advice tweets is described and results presented.

2. BACKGROUND

URLs that are embedded in tweets represent Web links to documents that can provide the user with lengthier health-related information (usually summarized or suggested in the tweet itself). A recent study by Cui et al. found that from an analysis of one million tweets, 29.1% contained URLs [3]. However after further examination it was found that over half these URLs were 'spam'-

related. On the other hand, a study examining public health-related tweets by Australian health organizations found that a large majority of public health-related tweets included URLs and that they were also genuine (i.e. non-spam and contained appropriate information as described in the tweet) [4]. These characteristics of tweets embedding URLs for public health document dissemination and retrieval suggests the value in further research into social media-disseminated public health documents.

Throughout the literature reviewed, evaluation of public health messaging in micro-blog applications such as Twitter (and more generally social media) has not been thoroughly explored. There have been various temporal estimation methods and models introduced for information retrieval and document rankings when compared to traditional media (i.e. newspaper articles) [5]. Document and data ranking (by Web search) has recently been explored when taking into consideration the social aspects of SNS like Twitter and Facebook [6]. The authors of the study proposed a ranking system based on the characteristics and communicative relationships between SNS application users as well as the actions these users performed on Web documents.

A form of socially-based Website review was introduced called tagging, where end-users place a content and quality label on a particular document on a topic of interest. These types of technologies [7] used on various Websites can also be seen on Twitter, whereby a user can perform various actions (e.g. re-tweeting and favouriting) to provide both dissemination and opinions/evaluation of the content of the document (or tweet).

3. SOCIAL MEDIA-BASED PUBLIC HEALTH INFORMATION DISSEMINATION

There are a number of inter-connected and diverse characteristics of social media-based public health information dissemination. In terms of dissemination, such systems could be considered to represent a distributed health information dissemination network with the network topology and routing depending upon the 'self-organizing' activities of the human members of the social media network. This self-organizing aspect of social media is present via users first manually choosing which accounts to follow, based on their preferences for what information and accounts are of interest and relevant to them, and also in choosing from various possible actions including forwarding or re-tweeting when receiving a health-related micro-blog post.

We identify novel aspects specific to social media-based public health information dissemination to include: user role in document dissemination, public review and evaluation, known and targeted recipients, impact of population values and user-initiated content.

User role in document dissemination: Users are actually forwarding and hence are the parties disseminating public health information or Web documents. When users receive information that they find interesting or useful, they may re-tweet the micro-blog post including any embedded link referring to a Web document.

Public review and document evaluation: Actions taken in relation to the public health information received from other accounts is also a source of evaluative information. By carrying out such actions as re-tweeting or favouriting, users are also passing a form of judgment and evaluating the type, quality and/or relevance of the tweet and any document referred to. For example, favouriting a document may be making an indication about the relevance of the public health information based on the needs of

the particular user, whereas re-tweeting may suggest that there is some value in that piece of public health information for followers to benefit from. These issues will be further explored in Section 5 via the preliminary experiment, analysis and results described.

Known and targeted recipients: Another significant characteristic of social media-based public health dissemination is that the recipients (at least the receiving accounts) of any given piece of public health information can be known. One of the challenges with mass media-based public health dissemination is that it can be hard to establish who has received a given message or how successfully targeted it is. Such capabilities in social media suggest the ability to analyze and optimize dissemination in detail across the population and also create software and analytic tools to measure and optimize public health information dissemination via social media.

Impact of population values: Social media-based systems also include an interplay between community interests and values, and what and how broadly information is disseminated. For example in [8] the issue of acute health risk indicated in tweets, and hence its possible relationship to feelings of social obligation, was identified as a characteristics of highly re-tweeted tweets.

User-initiated content: Varied types of users are also able to generate their own health information and micro-blog postings (whether this is accepted or not is based on various factors, such as user perception of the authority or interest of that piece of information). This example identifies another aspect of Twitter-based public health dissemination, whereby accepting and interacting with certain users and documents involves establishment of an informal 'network of human trust' as part of that information dissemination.

4. SOCIAL MEDIA vs SEARCH ENGINES

As stated, traditional forms of Internet-based access to public health information are often through the use of search engines. This could be broadly considered to be a 'pull'-based approach, where the user would discover and request information from various sources, and in this case public health Web documents. However with the introduction of SNS, the users of these services share information as the means for these documents to reach other individuals. While SNSs like Twitter support the sharing of short textual messages, they are very commonly utilized to direct others to Web pages and documents by providing a URL. Twitter is a good example of a SNS that incorporates a 'push'-based approach (see Table 1), where specified, relevant and up-to-the-minute health information is pushed to users. This may suggest the relevance of social media-based public health dissemination where health behavior modification is aimed for.

The criterion used to determine the resulting documents seen by users varies between the two systems. Search engines base their results on the user's query and identify the most relevant and authoritative pages based on those query words. In Twitter however, discovered documents are based on the type of account the user chooses to follow, and hence the quality and content of the public health information is dependent on the tweeting users being followed. This inherently includes some form of manual evaluation being done by the mass of users (see Section 5) implicit in the sharing and dissemination activities.

Social media is well suited to providing up-to-the-minute information. Up-to-the-minute information is well-suited to public health for various reasons and scenarios, such as epidemic outbreaks, health warnings, natural disasters and environmental information updates to name a few.

Another of the major differences between social media and search engine systems is the level of interaction with a public health Web document or information. Due to the open environment of Twitter, there are many mechanisms for peer feedback.

Table 1. Comparison of Web-based public health information dissemination systems

	Social Media	Search Engines
Mode	'Push'	'Pull'
Temporal	Most recent	Various times
Ranking/ evaluation	Manual/human selected	Algorithmic
Interaction	Community and peer-based	Individual-based
Document description	Manual by micro-blog poster	Automated/ anchor text-related
Documents disseminated	Changing rapidly	Relatively constant for a given query
Documents received	Relevent to a topic & topic personally selected	Relevent to a query
Web coverage	Limited to documents shared	More comprehensive

5. PRELIMINARY ANALYSIS OF A SAMPLE OF PUBLIC HEALTH RE-TWEETS

In the dissemination of public health advice via Twitter, there are a number of ways in which the broad populations of users inherently provide some form of evaluation of that information.

5.1 Preliminary Analysis of Sample

As part of a broader study described in [4], all tweets by health-related organizations in Australia meeting threshold criteria of at least 150 followers per account and having been sufficiently recently active, were collected and manually analyzed for the month of February 2012. There were 114 identified health-related organization accounts meeting these criteria, and these accounts produced 4787 tweets during that month.

These 4787 tweets were also categorized as being of various 'types' including for example: public health advice, organizational news, advertising, fundraising, conference and event, amongst various others. In this paper we limit our consideration to just the sub-set of these tweets that were public health advice tweets.

Across the whole 114 identified health-related organisations, there were a total of 772 (out of the overall total of 4787 tweets) public health advice tweets (re-tweeted and non re-tweeted) sent in February 2012. A total of 359 of these public health advice tweets were found to have been subsequently re-tweeted at least once. The identified Twitter accounts were also categorized into three sectors – government (16), for-profit (FP) (27) and not-for-profit (NFP) (71) organizations.

For this paper we were interested to gain preliminary insights into the relationships between public health information re-tweeting behaviour, and the Web documents pointed to by embedded URLs within the public health advice tweets, and also other preliminary characterizations of public health social media usage in terms of this sample.

Of the 359 public health advice tweets sent in February 2012, that were re-tweeted, 329 of these re-tweeted tweets contained a URL (91.6%). Notably, this is substantially higher than the proportion of re-tweeted tweets in general which contain embedded URLs (56.7%) [9].

All 359 tweets were ordered in terms of number of times they were re-tweeted. The most re-tweeted was re-tweeted 40 times, the top 25 tweets were re-tweeted seven or more times, the next 94 tweets were re-tweeted between three and six times inclusive and the remaining 240 were re-tweeted one or two times.

The Web pages indicated by the embedded URLs were manually inspected to determine the source of their content. The tweets that either did not include an embedded URL or for which this link was no longer functioning were excluded from this analysis. In the most highly re-tweeted group (seven or more times) there was some evidence of high authority for the information source of the linked-to Web pages: 40% were from leading NFPs such as the Australian Red Cross, Cancer Council NSW, the Heart Foundation etc.; 25% from government departments and 20% based on input/ content from Professors.

The second set of tweets considered were those re-tweeted three to six times inclusive. In this case only 31% of Web pages indicated by these tweets were from NFP organizations, a slightly lower percentage indicated government site pages (23%), a much lower percentage were from Professors (1.5%) and a small percentage (6%) had as their source international journal articles.

The third set, re-tweeted only once or twice showed again a decrease in NFP Web pages (26%), a decrease in government content pages (18%) and a decrease in Professors as sources of the Web page content to 1%. Interestingly there was an increased and relatively large proportion of pages in this group that used as their information source international journal articles (20%).

The above shows high authority sources such as government, well-recognized NFPs or Professors, as the sources for the content of indicated Web pages, being more prevalent for more highly re-tweeted tweets. On the other hand it also shows high authority sources such as international journal articles being more prevalent amongst lower re-tweeted tweets, counter to the possible overall trend of the embedded URLs within more highly re-tweeted tweets indicating pages with more authoritative information sources.

As an alternative analysis, to gain a simple measure of the 'authority' or 'quality' of the Web pages indicated by embedded URLs, the Google PageRank of each of the indicated Web pages (and of their domain name resulting from removal of the directory path from the end) were retrieved via the Google toolbar. Various averages of these PageRanks were then calculated. Tweets from the re-tweet set were excluded in this analysis if they did not include a URL, had a URL but it was no longer functioning or no PageRank was available for that URL.

For the most re-tweeted group (re-tweeted seven or more times) the average PageRank for the indicated Web pages was 4.25, for the set of tweets re-tweeted three or more times, the average was 3.9 and for the set of all tweets re-tweeted at least once the average was 3.38. In relation to the domain names for the top group of re-tweeted tweets the average domain PageRank was 6.52, for tweets re-tweeted three or more times, the average domain PageRank was 6.39 and for the set of all re-tweeted tweets the average domain PageRank was 6.05.

While these averages suggest that a higher PageRank was present on average for the URLs in more re-tweeted tweets, there was a low positive correlation between PageRank and re-tweet count with only $r=0.16$, with significance value of $p=0.01$.

In considering other factors with possible correlation with high re-tweeting, interestingly the number of followers of an account did not show a strong correlation with number of times re-tweeted with r= 0.243 and significance value p=0.000. Also there was a moderately strong negative correlation between total number of tweets sent from an account and PageRank of embedded URLs, r= -0.614 with p=0.000.

Other analysis of the sample data provides some preliminary insight into those doing the re-tweeting and hence carrying out the possible evaluative actions. In general, accounts of individuals were the most active re-tweeters: individual accounts re-tweeting numbers outnumbered government accounts re-tweeting for example by a ratio of ten-to-one. This suggests the large role of individuals in providing the document evaluative information. Finally, if a user refers to another on Twitter via the @ symbol, this is considered a 'mention'. There were only 51 instances of mentions in our sample of 359 re-tweeted tweets.

5.2 Discussion
The low r value for the correlation between PageRank and re-tweet number may be a result of various other factors impacting re-tweeting such as the actual semantic content of tweets, PageRank not being an appropriate measure of the quality or authority of public health information or possibly there being little linear relationship between page authority and re-tweet number. Previous work has also shown a not very high correlation between PageRank and the quality of health information [10].

The low r value of account follower number in relation to number of re-tweets is interesting in that it suggests that it is not just number of individuals receiving a tweet that drives numbers of re-tweets, which might be naively hypothesized, but it suggests that characteristics intrinsic to the content of the tweet may be more important in affecting the level of re-tweeting. The negative correlation between number of tweets and PageRank of embedded URLs suggests accounts sending many tweets are not tweeting URLs with high authority as indicated by PageRank.

Previous research has found that re-tweets contained a significantly larger percentage of embedded URLs when compared to regular tweets (i.e. URLs have a strong relationship with re-tweetability) [7], (56.7% and 19.0% respectively) [8]. From our sample, it may possibly be hypothesized that in relation to public health advice tweets the inclusion of URLs is more common and of even greater importance to achieve dissemination.

6. FUTURE RESEARCH
While public health social media networks may potentially create powerful 'self-organized' dissemination networks the effectiveness of these dissemination networks for public health needs to be further investigated.

Much further work is required to determine how effective re-tweeting and evaluative actions are in identifying and rewarding quality or relevance of public health Web pages or micro-blog posts. An immediate future step would be a more sophisticated measure of authority or quality of health information provided.

Related questions include: does such dissemination and evaluation implement a "wisdom of the crowd" behavior?; does the self-organized nature effectively route documents to individuals who are better qualified to evaluate these documents?

There are a far broader range of future research questions also. How effectively are public health tweets reaching target audiences? Given information on who are the social media recipients is available, related to this, what can be determined about a user from their connection network and postings? How well are users able to identify the relevant accounts to follow? Finally, in relation to healthcare, are users actually changing their health behavior as a result of receiving such information?

7. CONCLUSION
This paper has considered the novel characteristics of social media-based public health information dissemination. A preliminary study has considered the relationship between re-tweeting frequency and the authority or quality of the Web pages pointed-to by tweets and other preliminary dissemination characteristics. Future related directions of research are also identified.

8. REFERENCES
[1] Yang, C.C., Winston, F., Zarro, M.A. and Kassam-Adams, N. 2011. A Study of User Queries Leading to a Health Information Website:AfterTheInjury.org. In *Proceedings of iConference* (Seattle, USA, February 08 – 11, 2011). iConference'11. 267-272.

[2] Paul, M.J. and Dredze, M. 2011. You Are What You Tweet: Analyzing Twitter for Public Health. In *Proceedings of the 5th International Conference on Weblogs and Social Media* (Spain, July 17 – 21, 2011). ICWSM'11. AAAI, 265-272.

[3] Cui, A., Zhang, M., Liu, Y. and Ma, S. 2011. Are the URLs really popular in microblog messages? In *Proceedings of IEEE International Conference on Cloud Computing and Intelligence Systems* (Beijing, China, September 15 – 17, 2011). CCIS'11. IEEE, 1-5.

[4] Dumbrell, D. and Steele, R. 2013. Twitter and Health in the Australian Context: What Type of Information are Health-related Organizations Tweeting? In *Proceedings of the 46th Hawaii International Conference on System Sciences* (Hawaii, USA, January 07 – 10, 2013). HICSS'13.

[5] Efron, M. and Golovchinsky, G. 2011. Estimation methods for ranking recent information. In *Proc, of the 34th Intl. Special Interest Group on Information Retrieval* (Beijing, China, July 24 – 28, 2011). SIGIR'11, 495-504.

[6] Khodaei, A. and Shahabi, C. 2012. Social-Textual Search and Ranking. In *Proceedings of CrowdSearch Workshop* (Lyon, France, April 17, 2012). CrowdSearch'12. 3-8.

[7] Miguel, A.M., Karampiperis, P., Kukurikos, A., Karkaletsis, V., Villarroel, D. and Leis, A. 2011. Applying Semantic Web technologies to improve the retrieval, credibility and use of health-related web resources. *Health Informatics Journal*, 17, 2 (June, 2011), 95-115.

[8] Dumbrell, D. and Steele, R. 2012. What are the Characteristics of Highly Disseminated Public Health-related Tweets? In *Proceedings of the Australian Computer-Human Interaction Conference* November 26 – 30, OZCHI'12.

[9] Zarrella, D. Science of Retweets. http://www.slideshare.net/danzarrella/the-science-of-re-tweets, 2009. Date Accessed – September 15 2011.

[10] Griffiths, K., Tang, T., Hawking, D. and Christensen, H. 2005. Automated Assessment of the Quality of Depression Websites. Journal of Medical Internet Research, 7(5).