

Original Paper

# Measuring Health Information Dissemination and Identifying Target Interest Communities on Twitter: Methods Development and Case Study of the @SafetyMD Network

Venk Kandadai<sup>1\*</sup>, MPH; Haodong Yang<sup>2\*</sup>, MS; Ling Jiang<sup>2\*</sup>, MS; Christopher C Yang<sup>2\*</sup>, PhD; Linda Fleisher<sup>1\*</sup>, MPH, PhD; Flaura Koplin Winston<sup>1,3\*</sup>, MD, PhD

<sup>1</sup>Center for Injury Research and Prevention, Children's Hospital of Philadelphia, Philadelphia, PA, United States

<sup>2</sup>College of Computing and Informatics, Drexel University, Philadelphia, PA, United States

<sup>3</sup>Perelman School of Medicine, Department of Pediatrics, University of Pennsylvania, Philadelphia, PA, United States

\*all authors contributed equally

**Corresponding Author:**

Christopher C Yang, PhD

College of Computing and Informatics

Drexel University

3401 Market Street

Philadelphia, PA,

United States

Phone: 1 215 895 1631

Fax: 1 215 895 2492

Email: [chris.yang@drexel.edu](mailto:chris.yang@drexel.edu)

## Abstract

**Background:** Little is known about the ability of individual stakeholder groups to achieve health information dissemination goals through Twitter.

**Objective:** This study aimed to develop and apply methods for the systematic evaluation and optimization of health information dissemination by stakeholders through Twitter.

**Methods:** Tweet content from 1790 followers of @SafetyMD (July-November 2012) was examined. User emphasis, a new indicator of Twitter information dissemination, was defined and applied to retweets across two levels of retweeters originating from @SafetyMD. User interest clusters were identified based on principal component analysis (PCA) and hierarchical cluster analysis (HCA) of a random sample of 170 followers.

**Results:** User emphasis of keywords remained across levels but decreased by 9.5 percentage points. PCA and HCA identified 12 statistically unique clusters of followers within the @SafetyMD Twitter network.

**Conclusions:** This study is one of the first to develop methods for use by stakeholders to evaluate and optimize their use of Twitter to disseminate health information. Our new methods provide preliminary evidence that individual stakeholders can evaluate the effectiveness of health information dissemination and create content-specific clusters for more specific targeted messaging.

(*JMIR Res Protoc* 2016;5(2):e50) doi:[10.2196/resprot.4203](https://doi.org/10.2196/resprot.4203)

## KEYWORDS

Twitter; health information; dissemination; health communication; digital health

## Introduction

The Centers for Disease Control and Prevention (CDC) uses Twitter as its sole microblogging platform, actively encouraging its use to reach stakeholders with relevant health information [1] and to provide a framework for health information

dissemination best practices [2]. Web 2.0 and social media platforms like Twitter provide an opportunity to bridge the gap between innovation and dissemination by leveraging the viral spread of information across large networks of potential stakeholders. However, little is known about the ability of

individual stakeholder groups to achieve their health information dissemination goals through Twitter.

Twitter allows for easy communication and spread of information through networks. Methods to determine the effectiveness of this communication are well documented and widely accessible through an application programming interface (API). Twitter information spread begins when a user posts a short message (currently limited to 140 characters) referred to as a tweet. Users (the stakeholders to which the CDC refers), including influential users, can spread health information through Twitter by encouraging others to follow their tweets. In addition to receiving the information, followers can spread the information to their followers through retweeting (ie, echoing a message by another user). This is the benefit of social media networks: important information can be spread rapidly by sending a message on Twitter to a large and active following base.

An analysis of data from the Health Information and National Trends Survey (HINTS) revealed 23% of Internet users actively engaged in social networking sites [3], and a 2013 study showed 74% of online adults use social networking sites, suggesting a more than 2-fold growth of social network users in the past 5 years [4]. In particular, 35% of Internet users who use Twitter are young adults (aged 18-29 years) or elderly (65 years and older), and nearly 30% are from racial and ethnic minority populations [5]. This is the second potential benefit of Twitter: special groups that might be difficult to reach through other channels can be reached on Twitter, where they are already actively engaged.

There is little information, however, about the effectiveness of Twitter in spreading health information disseminated by credible, nongovernmental individual or group stakeholders. Such individuals and groups are usually part of a network of professionals with whom they work closely, and the network is a crucial channel to spread credible materials in the field. Exploring the effectiveness of the information dissemination in such networks is significant for enhancing the promotion of valuable news and findings. So far, few techniques are available for optimizing dissemination of credible health information by linking user interests with content. In an effort to develop tools to improve this dissemination, the authors report on a methodology examining a case study of an existing Twitter network, @SafetyMD. This study aimed to develop and apply

methods of (1) measuring the continued emphasis of health information themes as they spread through two levels of followers of @SafetyMD and (2) identifying targeted interest groups among the followers of @SafetyMD. The goal of this study was to advance methodology for the systematic evaluation and optimization of health information dissemination by stakeholders through Twitter.

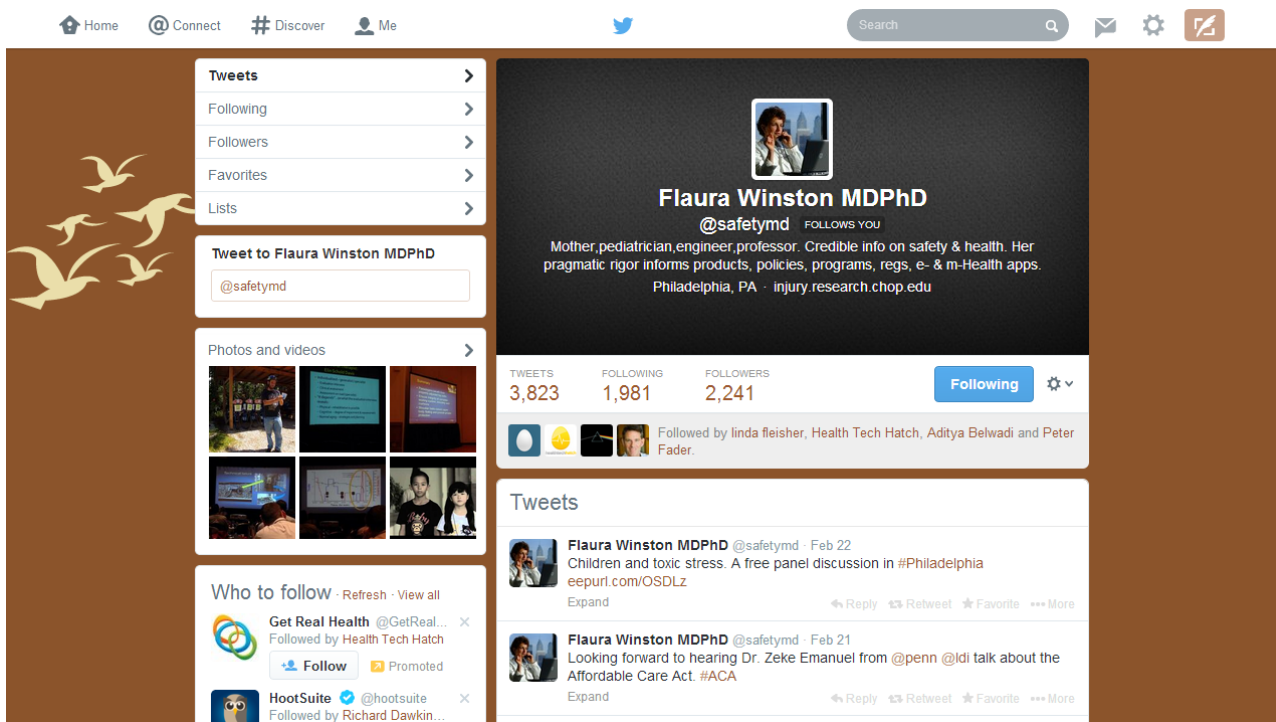
## Methods

### The @SafetyMD Network

The Center for Injury Research and Prevention (CIRP) at the Children's Hospital of Philadelphia has a research-to-action-to-impact outreach strategy that relies on a large network of child and adolescent professionals interested in injury prevention and treatment. Outreach professionals within the center work closely with scientists and engineers to translate evidence into credible messages and materials delivered mostly on the Web. One of the channels for disseminating this information is the official CIRP Twitter handle, @SafetyMD, launched in 2011 and led by the center's scientific director.

Content for @SafetyMD tweets and retweets is based on empirical research conducted within CIRP, new science from the injury research community at large, news, policy decisions, and advances in injury prevention strategies developed by industry. Followers of @SafetyMD include fellow physicians, nonprofit organizations, corporations, journalists, policymakers, and other influencers (eg, users from governmental or corporate entities) as well as researchers, entrepreneurs, and the general community. At the time of this analysis, @SafetyMD had 1790 followers (now 2550), and 458 tweets had been composed with an original-tweet-to-retweet ratio of 18 to 1 (indicating an emphasis on original content versus aggregated content). Given the purpose of the @SafetyMD handle, the relative diversity of its followers (in comparison to similarly sized research groups), and its level of activity, the authors concluded that @SafetyMD represented a typical academic/public health nongovernmental research entity that the CDC referred to in its Twitter strategy for disseminating health information. As such, @SafetyMD provided a convenient and generalizable platform to pilot test methods for measuring health information dissemination on Twitter. See [Figure 1](#) for a screen shot of @SafetyMD's current Twitter page.

**Figure 1.** The @SafetyMD Twitter Page. Note: Screen shot was taken on February 16, 2014 (after this analysis had been conducted).



## Data Extraction and Organization of the @SafetyMD Network

This protocol was exempt from review by the Children's Hospital of Philadelphia's Institutional Review Board. Based on Twitter protocols [6], we used the representational state transfer (REST) API version 1.0 provided by Twitter to acquire data. Specifically, we used the Status method to collect tweets posted by @SafetyMD and the Search method to collect retweets with keywords "RT @username."

Between July and November 2012, available profile information (eg, handle name, total number of followers, number followed by, total number of original tweets, user-provided Twitter profile descriptions, and tweet/retweet text content) was obtained from the 1790 @SafetyMD followers. Using the REST API's *GET statuses/user\_timeline* method, a custom PHP-crawler program was developed to interface with the API, extract JavaScript Object Notation (JSON) data from followers, and store this data in a MySQL database for accessible tabular data formats. The resulting database included tweets from @SafetyMD's network: those who follow @SafetyMD, those who retweet @SafetyMD, and those who retweet content from followers of @SafetyMD. Text content from @SafetyMD tweets was queried to identify the most frequent repeated words after the content was processed to remove stop words such as "a," "and," and "or." [7]. In order to pilot test and evaluate the methodology, we extracted important keywords from the tweets to investigate how key information disseminated through the network. We employed term frequency, commonly used in information retrieval to measure a terms' significance in the corpus, to identify as keywords terms repeated at least 10 times.

## Objective 1: Dissemination

This objective aimed to develop and apply methods for measuring the continued emphasis of health information themes as they spread through two levels of followers of @SafetyMD. A measure of user emphasis was defined as the proportion of total words in retweets that contained @SafetyMD keywords. As a baseline, we also measured the proportion of @SafetyMD keywords in the cumulative words of all @SafetyMD tweets. For this analysis, the @SafetyMD network was limited to those that were active retweeters and organized into three levels. In order to isolate mutually exclusive Twitter users, we defined Level 0 as @SafetyMD (n=458 original tweets), Level 1 as users who retweeted content from @SafetyMD (n=112 users, n=252 retweets), and Level 2 as those users who retweeted content from Level 1 and not directly from Level 0 (n=2356 users, n=4508 retweets).

## Objective 2: Targeted Interest Communities

In order to develop more targeted dissemination strategies, this objective examined whether @SafetyMD followers clustered into content-relevant groups. To pilot test this approach, a 10.00% random sample of @SafetyMD followers (n=179) was chosen and further limited to those who retweeted keywords from @SafetyMD (final sample analyzed n=170). For this sample, each user's 50 most recent tweets (inclusive of original tweets, retweets from @SafetyMD, and retweets from others) were selected for content analysis, and available user profile information was extracted and linked to the content analysis. The range of each user's 50 most recent tweets was set at 0 to 128 days from data extraction (this does not factor in each Twitter user's account lifetime, which could fall within the 0-128 day range).

Unique words were identified from the sample of 8500 tweets by a custom-written Java program that removed handle names,

URLs, punctuation symbols, and stop words. Words that included the hashtag symbol “#” as a prefix were kept because they represented a grouping of similar messages and topics on Twitter. This process yielded 1027 unique words. A word importance metric was created by computing the term frequency-inverse document frequency (TF-IDF) value using a Weka (Waikato Environment for Knowledge Analysis) filter class [8] for each word in each of the 8500 individual tweets. This processing generated a large matrix (170 followers × 1027 unique words) that required further data reduction through a 2-step process.

First, principal component analysis (PCA) [9] was conducted to explain the variance-covariance structure of linear combinations of TF-IDF values. After excluding components with eigenvalues less than 1, a second step involved hierarchical cluster analysis (HCA) [10] on the components to further segment the sample of @SafetyMD followers. Similarity was calculated based on affinities among the components extracted

from PCA. See Figure 2 for the formula calculating the minimal distance.

In this analysis,  $x_{ij}$  = the TF-IDF value of term  $j$  in follower  $i$ . After calculating the distance between the new cluster and other clusters, HCA combined any two closest clusters recursively until the algorithm merged all the variables into one final cluster. The furthest neighbor (complete linkage) cluster method was used [10]. As an additional exploratory step, common interests from followers from each cluster were obtained by reading available Twitter profile descriptions and extracting common words and themes.

### Statistical Analysis

Descriptive statistics, including frequencies and proportions, were computed as appropriate. A scree plot and tabulated eigenvalues from the PCA and a dendrogram from the HCA were generated to identify unique clusters. All aggregate analyses were performed using R version 3.0.1 and SPSS version 19 (IBM Corp).

Figure 2. Formula calculating minimal distance.

$$\text{Similarity}(i, j) = \frac{x_{i1}x_{j1} + x_{i2}x_{j2} + \dots + x_{ip}x_{jp}}{\sqrt{x_{i1}^2 + x_{i2}^2 + \dots + x_{ip}^2} \sqrt{x_{j1}^2 + x_{j2}^2 + \dots + x_{jp}^2}}$$

## Results

### Description of the @SafetyMD Network and Keywords

@SafetyMD had 1790 followers by the end of the 5-month study period and had composed 458 original tweets and retweets. The 1790 followers had a cumulative following base of 10,866,958 followers. The 458 @SafetyMD tweets and retweets contained 6538 words, and the stop word filter removed 2560 words (39.16%) of these words. A total of 31 keywords (words repeated at least 10 times) were generated from @SafetyMD’s 458 tweets, and these keywords were repeated 785 times (Table 1).

### Dissemination Across the Network

Of the 6538 words from the 458 tweets generated by @SafetyMD (Level 0), the keywords reflected a user emphasis, or proportion of @SafetyMD words in tweets that were keywords, of 12.01%. Within each subsequent level, the user emphasis remained and decreased: Level 1 contained 6.10% of @SafetyMD’s keywords among its 3711 retweeted words; Level 2, 2.50% of 60,795 retweeted words. All 31 keywords were represented in each level at least once.

The results depicted a possible dilution effect when retweeting from one level to the subsequent level. The @SafetyMD dissemination strategy aimed to use the viral nature of Twitter to spread evidence-based injury prevention information. While there was evidence of dissemination, only 2.50% of the content reached a second stage of spread. The change in keywords

throughout the diffusion process demonstrated how users at different levels in the constructed retweet network could serve as a proxy filter. They pass on the core information posted by @SafetyMD while shifting the focus by disseminating other types of information such as social events and social behavior.

### Targeted Interest Communities

The random sample was selected of 170 @SafetyMD followers who had composed 1,073,770 tweets and had a combined following base of 2,066,980 users. The 50 most recent tweets from these followers had a total word count of 35,602. Of these words, 9.60% were keywords, and all 31 @SafetyMD keywords were represented at least once among the 8500 tweets. PCA revealed 129 unique components from the data. Figure 3 depicts the scree plot generated from the 170×1027 matrix.

In an attempt to further classify the 170 followers, HCA generated 12 unique clusters from the 129 components (the dendrogram could not be presented due to its extremely large size but is available upon request). Table 2 reveals that the clusters shared common interests based on available Twitter profile descriptions provided by the users.

Nearly 60.0% of the followers were grouped in cluster 2 with common interests shared around driving education/safety and pediatric health and the words “drive,” “new,” “help,” “driver,” and “safe” had the five largest TF-IDF values. Although the HCA algorithm was able to differentiate clusters 5, 7, and 8, common interests from available Twitter profile descriptions were not available and could not be determined.

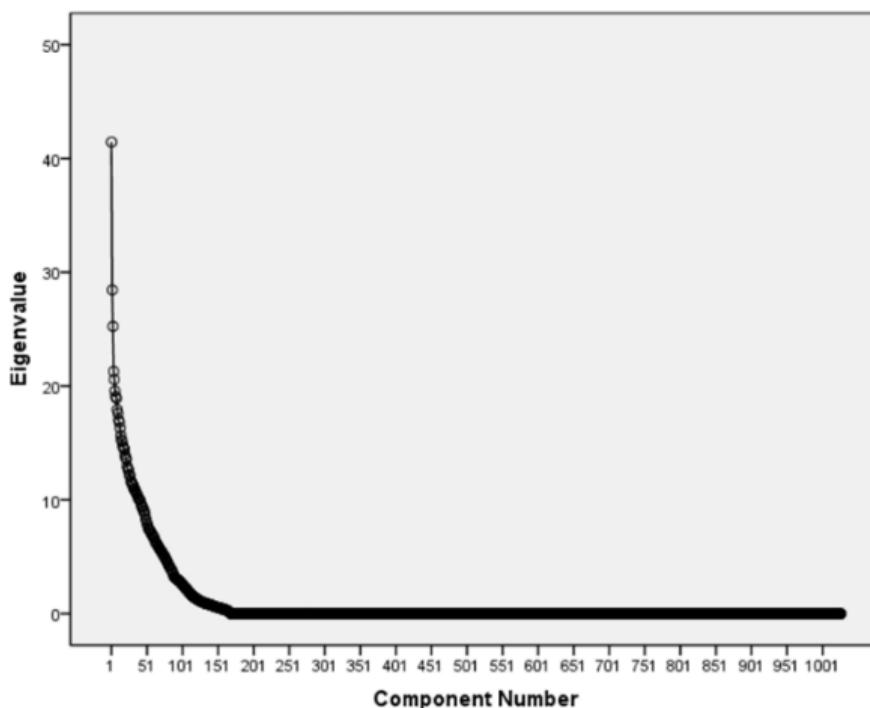
**Table 1.** @SafetyMD keywords.

Rank	Keyword	Frequency
1	#teendriving2012	112
2	teen	73
3	safety	51
4	driving	47
5	#safety2012	40
6	teens	32
7	driver	29
8	drivers	25
9	research	23
9	Safe	23
11	#roadsafety	21
12	#AMIA2012	20
12	#CelebrateMyDrive	20
12	CHOP	20
12	parents	20
16	crash	19
16	crashes	19
16	risk	19
19	chat	17
19	injury	17
21	car	16
22	study	15
23	drive	14
24	child	13
24	concussion	13
24	seat	13
27	#OHSU10X10	12
27	children	12
29	kids	10
29	passenger	10
29	positive	10

**Table 2.** Distribution of common interests across the 170 @SafetyMD followers grouped into 12 clusters.

Cluster	Followers, n	Common interests
1	5	Physician blogger, medical journalist
2	101	Driving school, drive training, driving safety, injury prevention, child health, pediatric physician
3	15	Moms, reporters, business owners
4	5	Spread of health-related information through social network
5	4	Could not be determined
6	8	Health research and health services, especially for driving safety
7	5	Could not be determined
8	6	Could not be determined
9	6	Physicians and medical research, especially child injury prevention
10	7	Medical education, driving education and training
11	4	Pediatricians and moms
12	4	Healthcare professionals, road safety professionals

**Figure 3.** Scree plot of extracted components and their corresponding eigenvalues. Components with Eigenvalues <1 were excluded revealing 129 components.



## Discussion

### Principal Findings

As a popular microblogging platform, Twitter enables users to disseminate information to a large audience. Users can be selective in deciding whether to retweet a tweet, and this is a natural filtering process. Only those who consider the information as valuable and credible would pass along the message and make it accessible to peers with common interests. Although the information could be diluted throughout the information dissemination process, Twitter makes it easy and fast for information to reach more relevant people than the original author’s immediate network. This study is one of the

first to develop methods for use by stakeholders to evaluate and optimize their use of Twitter for disseminating health information. Our newly developed methods provide preliminary evidence that individual stakeholders can evaluate the effectiveness of health information dissemination and create content-specific clusters for more specific targeted messaging through Twitter’s direct messaging function.

As an indicator of dissemination, the new metric *user emphasis* was defined representing the proportion of Twitter content that included keywords used by the stakeholder. The case study network, @SafetyMD, demonstrated a persistent but decreasing user emphasis of original @SafetyMD content (keywords) as the information spread through two levels of followers via

retweeting. By Level 2 (the followers of @SafetyMD's followers), user emphasis indicated that the use of @SafetyMD keywords had persisted but dropped by 9.5 percentage points. A second method created follower clusters based on content tweeted. The data reduction methods were able to differentiate 12 unique clusters of followers of @SafetyMD.

While multiple studies have conducted health information content analyses across Twitter networks [11-17], few have developed systematic methods to measure health information spread and leverage an existing Twitter network in order to differentiate interest groups. In a particular study utilizing NodeXL methods, Smith and colleagues [18] were able to identify clusters of content and proximal-based groups based on hashtags or selected words. In a study outside of the health sector examining the dissemination of anti-Islamic extremism on a popular Twitter handle, Blanquart and Cook [19] concluded that message dilution was a common phenomenon without the use of hashtags and embedded URLs in original tweets to magnify the messages. Our methods extend those reported previously by examining health information dissemination through a specific Twitter user network, identifying metrics for health information dissemination and leveraging an existing organization's Twitter strategy to effectively reach targeted groups.

More than 10 years ago, Berwick [20] argued that health care leaders lag in translating successful scientific innovations into practice and provided 7 recommendations to accelerate the diffusion of innovations. More recently, Glasgow and colleagues [21] suggested that traditional implementation and dissemination strategies recommended by Berwick (such as getting packaged or messaged information to influencers via traditional networking and partnership) yield labor-intensive and cost-inefficient results. Kreuter and Bernhardt [22] extend these ideas and recommend that health care entities establish systematic evaluation measures to successfully disseminate evidence-based public health programs. The authors stress the need for more pragmatic methodologies for efficient dissemination of health care innovations. Our methods directly respond to this recommendation by providing tools for stakeholders to conduct systematic evaluation on their social media interventions.

Given the wide popularity of Twitter (nearly 232 million active users worldwide [23]), the CDC endorses its use as an opportunity to reach new audiences and bridge dissemination gaps [1]. However, health messages on Twitter can be lost in the large volume of content (more than 5000 tweets composed each second and nearly 300 billion cumulative tweets [23]). Our research supports the potential for Twitter to disseminate health information; however, Twitter communication strategies may need to be optimized.

## Limitations

Our results are not without limitations. The newly proposed indicator, user emphasis, only takes into consideration retweets across levels of followers within an existing network and does not account for original status updates that may reference @SafetyMD keywords. In addition, user emphasis may change over time based on new followers and the content that is shared. Also, the current calculation of user emphasis was limited to retweets resulting in a potential conservative estimate because it did not consider other ways that followers interacted with @SafetyMD (eg, through mentions and conversations). The list of the 31 @SafetyMD keywords was generated directly from @SafetyMD tweets and was not further compared to common words used in the broader health information environment. Therefore, we were not able to classify these keywords as original @SafetyMD content versus content influenced by the general health information environment on Twitter.

Our random selection of 170 @SafetyMD followers and their 50 most recent tweets may not necessarily represent content that best describes their information interests or needs, which might change over time. In particular, the range of recent tweet content was within 0 to 128 days of data extraction and may not have taken into account a user's most current Twitter behavior. Future studies should examine a larger sample and the evaluation should be rolling over time to look at trends in interests. Also, the user-provided Twitter profile description might not have been current or complete (as it is part of the registration process and limited in length). Future studies might consider use of surveys to evaluate user interests.

It is clear from the HCA that the overwhelming majority of users were grouped in cluster 2; less than 10% were spread among the remaining 11 clusters. In addition, the common interests of many of the clusters are quite similar to each other, which may imply that available Twitter profile information is not a reliable and valid tool to describe mathematically unique groups on Twitter (given the inconsistency of available profile information). Future studies are needed to validate the relative uniqueness of these clusters and describe the clusters generated from this method using rigorous methodologies that may not involve using available profile information. Finally, these methods will only be useful as long as Twitter continues to share the data.

## Conclusions

This study aimed to develop and test a set of methods to (1) measure health information spread across an existing Twitter network and (2) leverage an existing Twitter network to identify target interest groups in an effort to provide tools for organizations that use Twitter to communicate health information. The results from @SafetyMD case study provide preliminary evidence of systematic yet simple tools that can be used to effectively leverage an existing Twitter network used to promote credible health information.

## Acknowledgments

This project was funded by a grant to Flaura Winston from the Pennsylvania Department of Health. The Department specifically disclaims responsibility for any analyses, interpretations, or conclusions. The authors thank Twitter and their well-documented API and the followers of @SafetyMD. We also acknowledge the outreach professionals at CIRP at the Children's Hospital of Philadelphia for the review of this article and their efforts in translating research into action for child injury prevention.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Peer-reviewed final report to funding agency (Pennsylvania Department of Health).

[\[PDF File \(Adobe PDF File\), 931KB - resprot\\_v5i2e50\\_app1.pdf\]](#)

## References

1. Social media at CDC: Twitter. Atlanta GA: Centers for Disease Control and Prevention URL: <http://www.cdc.gov/socialmedia/tools/twitter.html> [accessed 2015-01-05] [WebCite Cache ID 6VMEhtb9w]
2. Twitter guidelines and best practices. Atlanta GA: Centers for Disease Control and Prevention URL: <http://www.cdc.gov/socialmedia/tools/guidelines/twitter.html> [accessed 2015-01-05] [WebCite Cache ID 6VMEjovIR]
3. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. *J Med Internet Res* 2009;11(4):e48 [FREE Full text] [doi: [10.2196/jmir.1249](https://doi.org/10.2196/jmir.1249)] [Medline: [19945947](https://pubmed.ncbi.nlm.nih.gov/19945947/)]
4. Social Networking Fact Sheet. Washington DC: Pew Internet and American Life Project URL: <http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet> [accessed 2015-01-05] [WebCite Cache ID 6VMEIY90S]
5. Brenner J, Smith A. 72% of online adults are social networking site users. Washington DC: Pew Internet and American Life Project URL: <http://www.pewinternet.org/2013/08/05/72-of-online-adults-are-social-networking-site-users> [accessed 2015-01-05] [WebCite Cache ID 6VMEnATpt]
6. Twitter Developers. Twitter Developers Web site URL: <https://dev.twitter.com/> [accessed 2015-01-05] [WebCite Cache ID 6VMEpusaQ]
7. Scott JG. *Social Network Analysis: a Handbook*. London UK: SAGE Publications; 2000.
8. Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: An update. *SIGKDD Explorations* 2009;11(1):10-18.
9. Jolliffe IT. *Principal Component Analysis*. New York NY: Springer; 2002.
10. Knoke D, Yang S. *Social Network Analysis (Quantitative Applications in the Social Sciences)*. London UK: Sage Publications, Inc; 2007.
11. Shive M, Bhatt M, Cantino A, Kvedar J, Jethwani K. Perspectives on acne: what Twitter can teach health care providers. *JAMA Dermatol* 2013 May;149(5):621-622. [doi: [10.1001/jamadermatol.2013.248](https://doi.org/10.1001/jamadermatol.2013.248)] [Medline: [23677100](https://pubmed.ncbi.nlm.nih.gov/23677100/)]
12. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: Twitter and antibiotics. *Am J Infect Control* 2010 Apr;38(3):182-188 [FREE Full text] [doi: [10.1016/j.ajic.2009.11.004](https://doi.org/10.1016/j.ajic.2009.11.004)] [Medline: [20347636](https://pubmed.ncbi.nlm.nih.gov/20347636/)]
13. Myslin M, Zhu S, Chapman W, Conway M. Using Twitter to examine smoking behavior and perceptions of emerging tobacco products. *J Med Internet Res* 2013;15(8):e174 [FREE Full text] [doi: [10.2196/jmir.2534](https://doi.org/10.2196/jmir.2534)] [Medline: [23989137](https://pubmed.ncbi.nlm.nih.gov/23989137/)]
14. Lee JY, Sundar SS. To tweet or to retweet? That is the question for health professionals on Twitter. *Health Commun* 2013;28(5):509-524. [doi: [10.1080/10410236.2012.700391](https://doi.org/10.1080/10410236.2012.700391)] [Medline: [22873787](https://pubmed.ncbi.nlm.nih.gov/22873787/)]
15. Heavilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain via Twitter. *J Dent Res* 2011 Sep;90(9):1047-1051 [FREE Full text] [doi: [10.1177/0022034511415273](https://doi.org/10.1177/0022034511415273)] [Medline: [21768306](https://pubmed.ncbi.nlm.nih.gov/21768306/)]
16. Harshavardhan A, Gandhe A, Lazarus R, Yu SH, Liu B. Predicting flu trends using Twitter data. *Computer Communications Workshops* 2011:702-707. [doi: [10.1109/INFCOMW.2011.5928903](https://doi.org/10.1109/INFCOMW.2011.5928903)]
17. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of Twitter among local health departments: an analysis of information sharing, engagement, and action. *J Med Internet Res* 2013;15(8):e177 [FREE Full text] [doi: [10.2196/jmir.2775](https://doi.org/10.2196/jmir.2775)] [Medline: [23958635](https://pubmed.ncbi.nlm.nih.gov/23958635/)]
18. Smith MA, Rainie L, Shneiderman B, Himelboim I. Mapping Twitter topic networks: From polarized crowds to community clusters. Washington DC: Pew Internet and American Life Project URL: <http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/> [accessed 2015-01-05] [WebCite Cache ID 6VMErjTEN]
19. Blanquart G, Cook DM. Twitter influence and cumulative perceptions of extremist support: A case study of Geert Wilders. 2013 Presented at: Proceedings of the 4th Australian Counter Terrorism Conference; 2013; Perth, Australia p. 1-11.
20. Berwick DM. Disseminating innovations in health care. *JAMA* 2003 Apr 16;289(15):1969-1975. [doi: [10.1001/jama.289.15.1969](https://doi.org/10.1001/jama.289.15.1969)] [Medline: [12697800](https://pubmed.ncbi.nlm.nih.gov/12697800/)]



21. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. *Int J Med Inform* 2014 Jul;83(7):e1-e11. [doi: [10.1016/j.ijmedinf.2013.07.002](https://doi.org/10.1016/j.ijmedinf.2013.07.002)] [Medline: [23910896](https://pubmed.ncbi.nlm.nih.gov/23910896/)]
22. Kreuter MW, Bernhardt JM. Reframing the dissemination challenge: a marketing and distribution perspective. *Am J Public Health* 2009 Dec;99(12):2123-2127. [doi: [10.2105/AJPH.2008.155218](https://doi.org/10.2105/AJPH.2008.155218)] [Medline: [19833993](https://pubmed.ncbi.nlm.nih.gov/19833993/)]
23. Ahmad I. 30+ of the most amazing Twitter statistics.: *Social Media Today* URL: <http://www.socialmediatoday.com/content/30-most-amazing-twitter-statistics-infographic> [accessed 2015-01-05] [WebCite Cache ID 6VMEtNG21]

## Abbreviations

**API:** application program interface

**CDC:** Centers for Disease Control and Prevention

**CIRP:** Center for Injury Research and Prevention

**HCA:** hierarchical cluster analysis

**HINTS:** Health Information and National Trends Survey

**PCA:** principal component analysis

**TF-IDF:** term frequency-inverse document frequency

*Edited by G Eysenbach; submitted 05.01.15; peer-reviewed by J Lee; comments to author 08.04.15; revised version received 11.12.15; accepted 04.01.16; published 05.05.16*

*Please cite as:*

*Kandadai V, Yang H, Jiang L, Yang CC, Fleisher L, Winston FK*

*Measuring Health Information Dissemination and Identifying Target Interest Communities on Twitter: Methods Development and Case Study of the @SafetyMD Network*

*JMIR Res Protoc 2016;5(2):e50*

*URL: <http://www.researchprotocols.org/2016/2/e50/>*

*doi: [10.2196/resprot.4203](https://doi.org/10.2196/resprot.4203)*

*PMID: [27151100](https://pubmed.ncbi.nlm.nih.gov/27151100/)*

©Venk Kandadai, Haodong Yang, Ling Jiang, Christopher C. Yang, Linda Fleisher, Flora Koplin Winston. Originally published in *JMIR Research Protocols* (<http://www.researchprotocols.org>), 05.05.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Research Protocols*, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.researchprotocols.org>, as well as this copyright and license information must be included.