Opioid Surveillance using Social Media: How URLs are shared among Reddit members

Albert Park* and Mike Conway

Biomedical Informatics, University of Utah, Salt Lake City, UT, USA

Objective

We aim to understand (1) the frequency of URL sharing and (2) types of shared URLs among opioid related discussions that take place in the social media platform called *Reddit*.

Introduction

Nearly 100 people per day die from opioid overdose in the United States. Further, prescription opioid abuse is assumed to be responsible for a 15-year increase in opioid overdose deaths[1]. However, with increasing use of social media comes increasing opportunity to seek and share information. For instance, 80% of Internet users obtain health information online[2], including popular social interaction sites like Reddit (http://www.reddit.com), which had more than 82.5 billion page views in 2015[3]. In Reddit, members often share information, and include URLs to supplement the information. Understanding the frequency of URL sharing and types of shared URLs can improve our knowledge of information seeking/sharing behaviors as well as domains of shared information on social media. Such knowledge has the potential to provide opportunities to improve public health surveillance practice. We use Reddit to track opioid related discussions and then investigate types of shared URLs among Reddit members in those discussions.

Methods

First, we use a dataset[4]—made available on Reddit—that has been used in several informatics studies[5,6]. The dataset is comprised of 13,213,173 unique member IDs, 114,320,798 posts, and 1,659,361,605 associated comments that are made on 239,772 (including active and inactive) subreddits (i.e., sub-communities) from October 2007 to May 2015. Second, we identified 9 terms that are associated with opioids. The terms are 'opioid', 'opium', 'morphine', 'opiate', 'hydrocodone', 'oxycodone', 'fentanyl', 'heroin', and 'methadone'. Third, we preprocessed the entire dataset (i.e., converting text to lower cases and removing punctuation) and extracted discussions with opioid terms and their metadata (e.g., user ID, post ID) via a lexicon-based approach. Fourth, we extracted URLs using Python from these discussions, categorized the URLs by domain, and then visualized the results in a bubble chart[7].
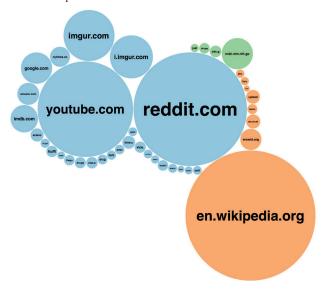
Results

We extracted 1,121,187 posts/comments that were made by 328,179 unique member IDs from 8,892 subreddits. Of the 1,121,187 posts/comments, 82,639 posts/comments contained URLs (7.37%), and these posts consisted of 272,551 individual URLs and 138,206 unique URLs. The types of shared URLs in these opioid related discussions are summarized in Figure 1. The color and size represent the type and size respectively of shared URLs. The '.com' is in blue; '.org' is in orange; and '.gov' is in green.

Conclusions

We present preliminary findings concerning the types of shared URLs in opioid-related discussions among Reddit members. Our initial results suggest that Reddit members openly discuss opioid related issues and URL sharing is a part of information sharing. Although members share many URLs from reliable information

sources (e.g., 'ncbi.nlm.nih.gov', 'wikipedia.org, 'nytimes.com', 'sciencedirect.com'), further investigation is needed concerning many of the '.com' URLs, which have the potential to contain high and/or low quality information (e.g., 'youtube.com', 'reddit.com', 'google. com', 'amazon.com') to fully understand information seeking/ sharing behaviors on social media and to identify opportunities, such as misinformation dissemination for improving public health surveillance practice.

Keywords

Social media; Surveillance system; Data mining; Opioid

References

1. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control D of UIP. Drug overdose deaths in the United States continue to increase in 2015. Archived at: http://www.webcitation.org/6rl2f65C2

2. Fox S. Health topics: 80% of Internet users look for health information online. Pew Internet & American Life Project. 2011. Archived at: http://www.webcitation.org/6gbUGjFS5

3. Reddit. Reddit in 2015. 2015. Archived at: http://www.webcitation.org/6eTHN0TFD

4. Reddit_Member. I have every publicly available Reddit comment for research. ~ 1.7 billion comments @ 250 GB compressed. Any interest in this?. 2015. Archived at: http://www.webcitation.org/6kgAuNxDE

5. Park A, Conway M. Longitudinal Changes in Psychological States in Online Health Community Members: Understanding the Long-Term Effects of Participating in an Online Depression Community. J Med Internet Res. 2017 Mar 20;19(3):e71.

6. Park A, Conway M. Towards Tracking Opium Related Discussions in Social Media. Online J Public Health Inform. 2017 May 2;9(1):e73.

7. Bostock M, Ogievetsky V, Heer J. D³: Data-Driven Documents. IEEE Trans Vis Comput Graph. 2011 Dec;17(12):2301–9.

*Albert Park
E-mail: alpark1216@gmail.com