

# Towards Tracking Opium Related Discussions in Social Media

Albert Park\* and Mike Conway

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

## Objective

We aim to develop an automated method to track opium related discussions that are made in the social media platform called *Reddit*. As a first step towards this goal, we use a keyword-based approach to track how often Reddit members discuss opium related issues.

## Introduction

In recent years, the use of social media has increased at an unprecedented rate. For example, the popular social media platform *Reddit* (<http://www.reddit.com>) had 83 billion page views from over 88,000 active sub-communities (subreddits) in 2015. Members of *Reddit* made over 73 million individual posts and over 725 million associated comments in the same year [1].

We use *Reddit* to track opium related discussions, because *Reddit* allows for throwaway and unidentifiable accounts that are suitable for stigmatized discussions that may not be appropriate for identifiable accounts. *Reddit* members exchange conversation via a forum like platform, and members who have achieved a certain status within the community are able to create new topically focused group called subreddits.

## Methods

First, we use a dataset archived by one of *Reddit* members who used *Reddit*'s official Application Programming Interface (API) to collect the data ([https://www.reddit.com/r/datasets/comments/3bxlg7/i\\_have\\_every\\_publicly\\_available\\_reddit\\_comment/](https://www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment/)). The dataset is comprised of 239,772 (including both active and inactive) subreddits, 13,213,173 unique user IDs, 114,320,798 posts, and 1,659,361,605 associated comments that are made from Oct of 2007 to May of 2015. Second, we identify 10 terms that are associated with opium. The terms are 'opium', 'opioid', 'morphine', 'opiate', 'hydrocodone', 'oxycodone', 'fentanyl', 'oxy', 'heroin', 'methadone'. Third, we preprocess the entire dataset, which includes structuring the data into monthly time frame, converting text to lower cases, and stemming keywords and text. Fourth, we employed a dictionary approach to count and extract timestamps, user IDs, posts, and comments containing opium related terms. Fifth, we normalized the frequency count by dividing the frequency count by the overall number of the respective variable for that period.

## Results

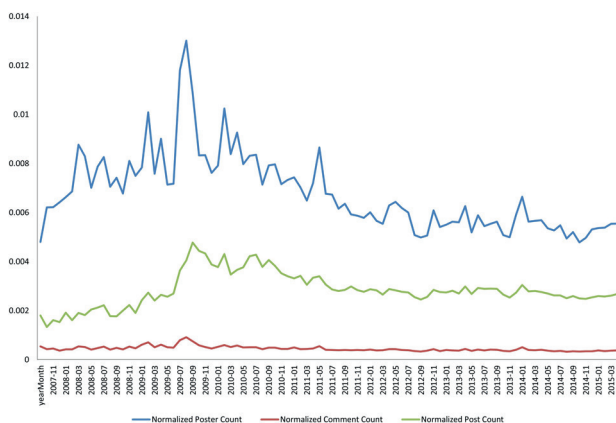
According to our dataset, *Reddit* members discuss opium related topics in social media. The normalized frequency count of posters shows that less than one percent members, on average, talk about opium related topics (Figure 1). Although the community as a whole does not frequently talk about opium related issues, this still amounts to more than 10,000 members in 2015 (Figure 2). Moreover, members of *Reddit* created a number of subreddits, such as 'oxycontin', 'opioid', 'heroin', 'oxycodon', that explicitly focus on opioids.

## Conclusions

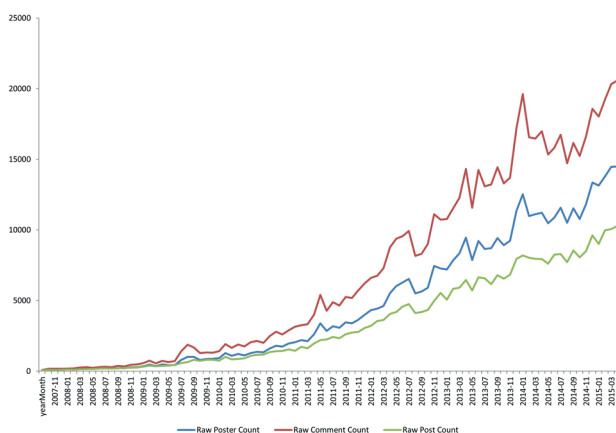
We present preliminary findings on developing an automated method to track opium related discussions in *Reddit*. Our initial results suggest that on the basis of our analysis of *Reddit*, members of the *Reddit* community discuss opium related issues in social media,

although the discussions are contributed by a small fraction of the members.

We provide several interesting directions to future work to better track opium related discussions in *Reddit*. First, the automated method needs to be further developed to employ more sophisticated methods like knowledge-based and corpus-based approaches to better extract opium related discussions. Second, the automated method needs to be thoroughly evaluated and measure precision, recall, accuracy, and F1-score of the system. Third, given how many members use social media to discuss these issues, it will be helpful to investigate the specifics of their discussions.



Line Graphs of normalized frequency counts for posters, comments, and posts that contained opium related terms



Line Graphs of raw frequency counts for posters, comments, and posts that contained opium related terms

## Keywords

Data mining; Surveillance system; Social media



## ISDS 2016 Conference Abstracts

### Acknowledgments

We restricted our analysis to publicly available discussion content and the University of Utah's Institutional Review Board (IRB) reviewed the study procedure and data (IRB 00076188).

Dr. Park was funded by National Institute for Health-National Library of Medicine T15 LM007124. Dr. Conway's contribution to this research was supported by National Library of Medicine of the National Institutes of Health under award numbers R00LM011393 & K99LM011393.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

### References

1. Reddit. Reddit in 2015 [Internet]. 2015. Available from: <http://www.redditblog.com/2015/12/reddit-in-2015.html>; Archived at: <http://www.webcitation.org/6eTHN0TFD>

### \*Albert Park

E-mail: [alpark7@uw.edu](mailto:alpark7@uw.edu)

