

Open Social Comparison and Analysis for
Reddit

The Team

Anna DeBrunner, debrunac@mail.uc.edu

Charles Greenman, greenmcg@mail.uc.edu

Isaac Mathers, matherit@mail.uc.edu

Advisor: Dr. Raj Bhatnagar

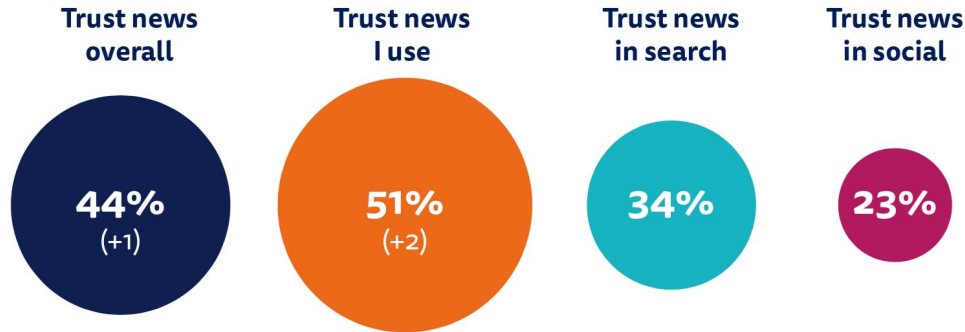
Abstract

The **exploitation of social media networks** came to light following the 2016 US Presidential election. Since then, the websites we use every day have become more aware of the issue, but disinformation has also become more pervasive. With another election cycle just around the corner, we must **arm ourselves with the correct tools and information** to ensure we do not fall prey to **false stories, targeted beliefs, and general discord**. With that in mind, the Open Social Comparison and Analysis for Reddit tool (OSCAR) **provides a more informed experience** on Reddit regarding its users and communities.

.

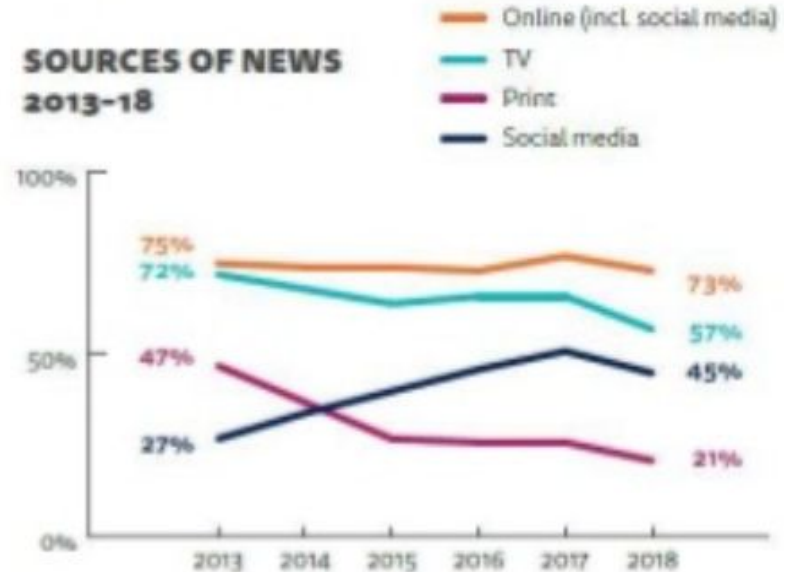
Background

PROPORTION THAT SAY THEY TRUST NEWS FROM EACH SOURCE – ALL MARKETS



From a study at the Reuters Institute:

<https://reutersinstitute.politics.ox.ac.uk/risj-review/trust-misinformation-and-declining-use-social-media-news-digital-news-report-2018>



Background

About one in four report sharing fabricated news – whether aware at the time or not

% of U.S. adults who say they ...

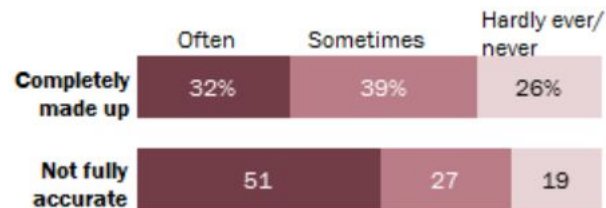


Source: Survey conducted Dec. 1-4, 2016.
"Many Americans Believe Fake News Is Sowing Confusion"

PEW RESEARCH CENTER

About a third say they often see made-up political news online; 51% say they see inaccurate news

% of U.S. adults who often/sometimes/hardly ever or never come across political news online that is ...



Source: Survey conducted Dec. 1-4, 2016.
"Many Americans Believe Fake News Is Sowing Confusion"

PEW RESEARCH CENTER

From a study at the Pew Research Center:

<https://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/>

Background

- Reddit is an American social news aggregation, web content rating, and discussion website.
- People that post, comment, and vote on Reddit must have an account. These people are referred to as **users**.
- Reddit is divided into communities/forums called **subreddits**. Each subreddit usually centered around a topic. Examples include Minecraft, Politics, and Fitness.
- **Submissions** are posted by users on subreddits.
- **Comments** are left on submissions by users.
- **Karma** is the amount of positive votes that a user has received.

Goals

- Explore users, communities, and topics on Reddit to discover flags of potentially harmful users.
- Present information, comparisons, and recommendations to users to make them more informed about the users and content they are interacting with online.
- Implement data analysis algorithms and visualizations to summarize available information about a user in an easily digestible way.
- Implement graph analysis algorithms and visualizations to summarize available information about user networks in an easily digestible way.

Intellectual Merits

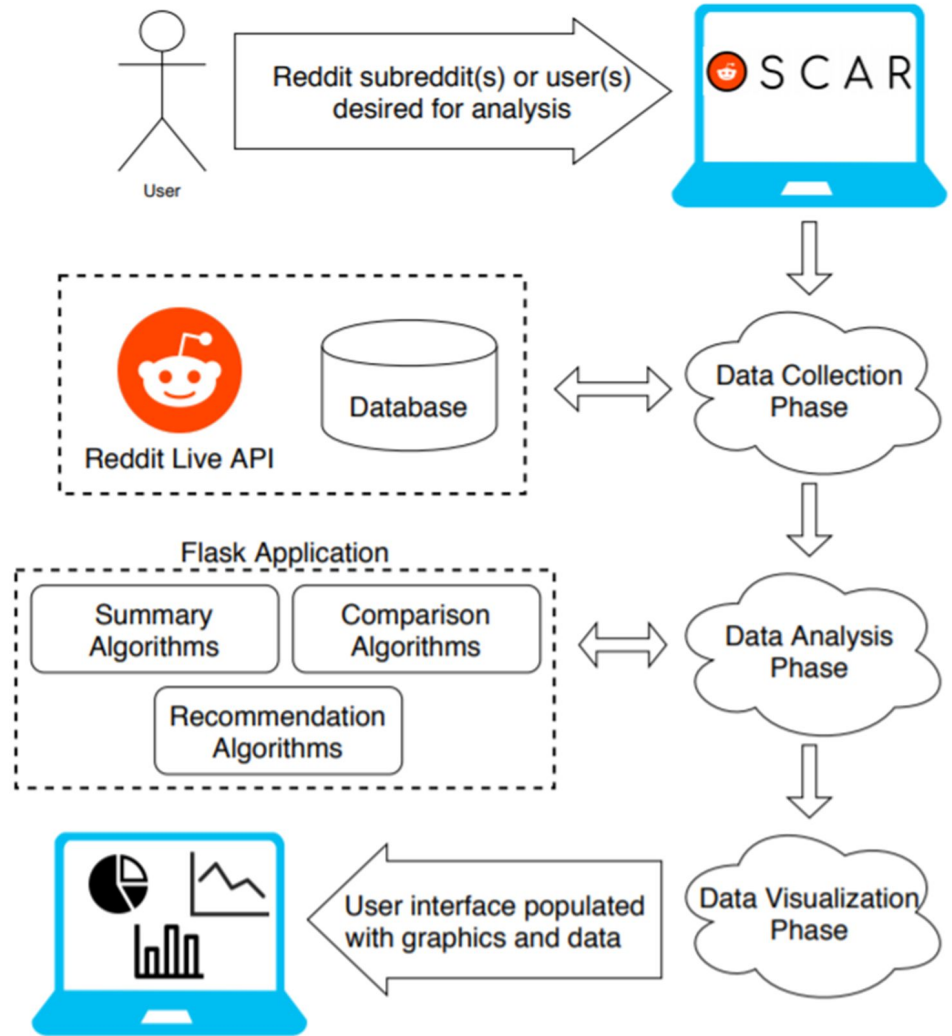
- The identification of misinformation via **analysis of social networks** rather than analysis of the content has been researched but has not been implemented in a way that can be easily used by consumers.
- We have taken research regarding this topic to create **our own implementation of a recommendation algorithm** that supplies a likelihood that a Reddit user could be a spreader of false information.
- This methodology could be improved and expanded to other social media websites to provide **a new technological solution to the problem of misinformation**.

Broader Impacts

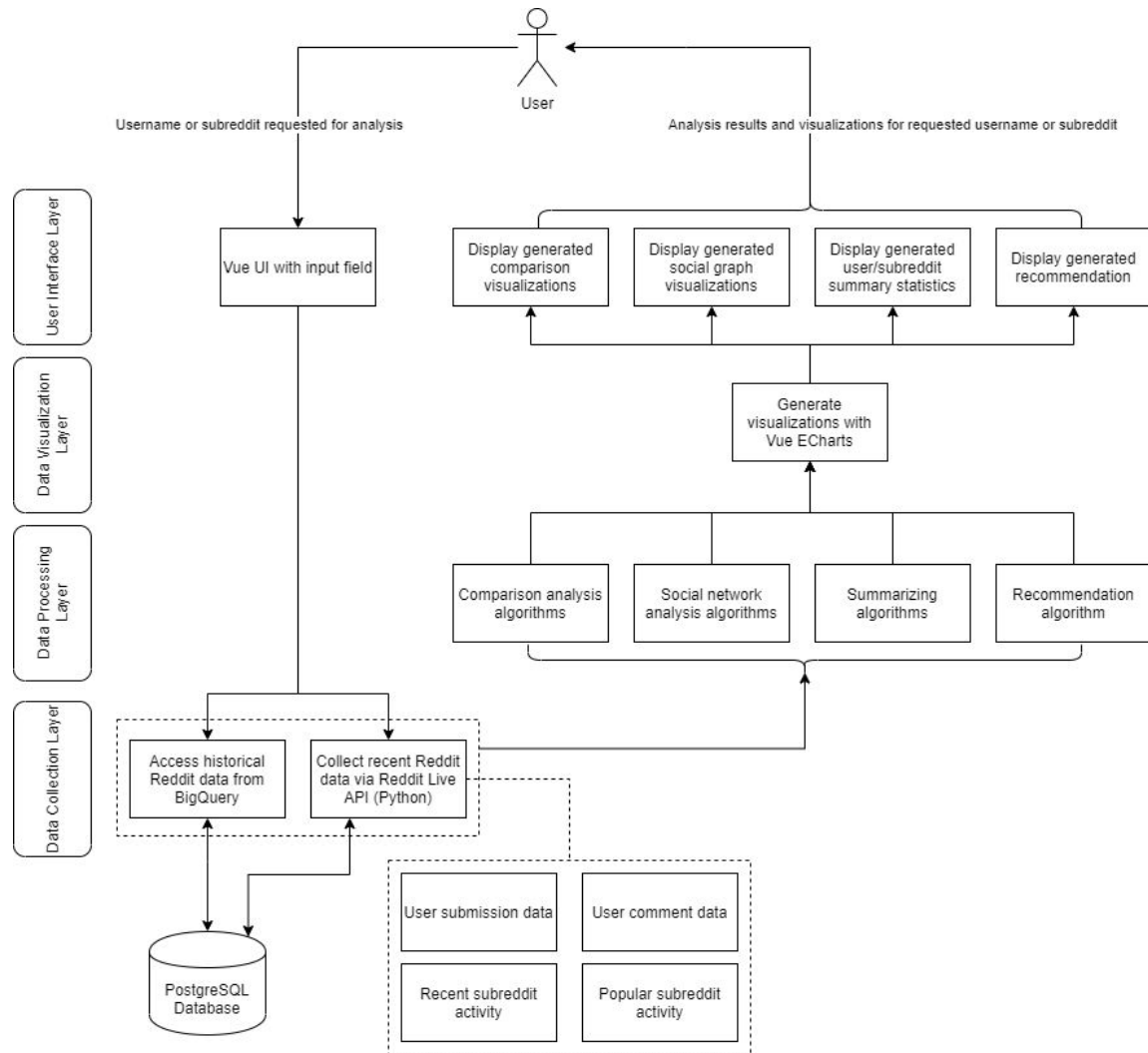
- Allows users to educate themselves about who they are interacting with online and determine if these users might have hidden intentions.
- Gives users an indication as to whether or not they should be particularly wary of information posted by a certain user.
- Provides an alternative, analytical way to view users and communities on Reddit.
- Helps fight the intentional and unintentional spread of misinformation, allowing us to live in a more fair and honest society.

Project Design

High Level Diagram



Low Level Diagram



Technologies

- **Reddit Live API** - We created a Python library that uses the Reddit PRAW library to obtain live data from Reddit. We clean that data so it appears in a format that can be used by our summary and analysis algorithms that also live in this library.



Technologies

- **SocksCatch Algorithm** - We implemented this algorithm based on research done at Normandie Université by Zaher Rabah Yamak. The premise of the algorithm is to determine if a person might be part of a sock puppet organization based on their activity. It calculates percentiles three factors: delay, action count, and relationships with other users in the community (subreddit). Based on the percentile the user falls into for each factor, their activity may be deemed normal or abnormal.

Technologies

- **PostgreSQL Database** - We created a PostgreSQL database that contains an aggregated, historic sample of recent Reddit activity. This data is meant to be representative of “typical” Reddit activity around the time that the data was collected.



Technologies

- **Python Flask service** - We created a Python Flask service in order to pass information from the Reddit Live API back to the user interface. The user interface can query this service via HTTP requests for baseline data or results from our algorithms including SocksCatch.



Flask

Technologies

- **VueJS User Interface** - We created a user interface to allow easy access to the insights we are calculating on the back end. The user will be able to submit a user or subreddit they want to analyze and then will be able to view and interact with the output of the analysis. During analysis, the user interface will query the Python Flask application for data and results.



Challenges and Future Work

Challenges

- Limited budget and resources, particularly processing power.
- Reddit API request limitations
- Implementation of recommendation methodology
- Shortened timeline and limited communication due to COVID-19

Opportunities for Further Development

- Comparison capabilities, both to the backend and as a new tab in the user interface
- Analysis of a single submission post
- More research and development into the recommendation methodology
- Scaling up and deploying the application for easier use by consumers.

Thank you!