



Misinformation on Reddit

David Aminifard & Samuel Huang



Table of contents

01

Introduction

How did we get here?

02

Context

What work already exist
in this space?

03

Content

Our methods.

04

Conclusion

The results of our work
and our findings.



01

Introduction

How did we get here?

Progress



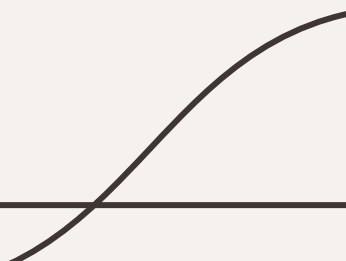
Last Quarter

K-Core Analysis of Twitter
data about the world
COVID-19 pandemic



This Quarter

Data Analysis on Reddit
data of the most
well-known and
controversial subreddits

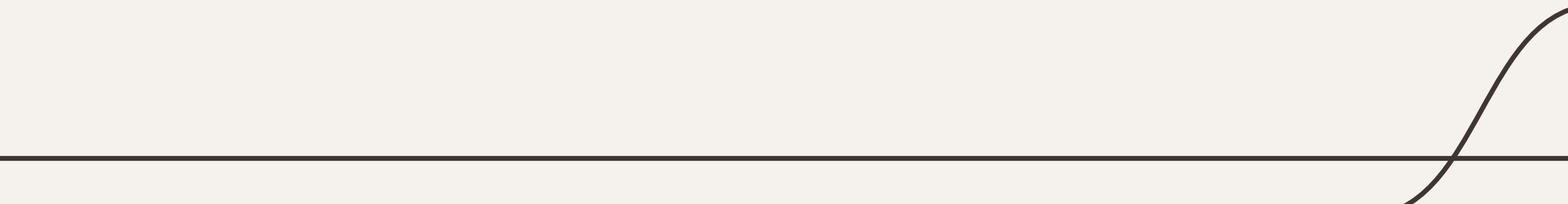




02

Context

What is the current status of the field?



Similar Project

"The Web Centipede: Understanding How Web Communities Influence Each Other Through the Lens of Mainstream and Alternative News Sources"

Authors

Nicolas Kourtellis, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn

Objective

The paper looks into the spread of misinformation on the Reddit Platform

Method

The paper compares the spread of mainstream and alternative news URLs on Reddit, Twitter, and 4chan.

Results

4chan and six selected Subreddits have a very high percentage of alternative news URLs compared to other communities

How does the spread of misinformation vary from subreddit to subreddit, and what is the relationship between these subreddits?

What can user activity tell us about trends within subreddits and users who commonly post misinformation?

Hypothesis

Subreddits leaning more towards sensitive topics, such as politics, news etc. are more likely to be spreading misinformation. In addition, these subreddits may share a lot of the same users.





03

Content

What did we actually do?

Our Project

Summary

- We identified the frequency of misinformation on selected Subreddits based on the presence of misinformation URLs
 - Our data primarily consists of Reddit submissions and their respective attributes
 - We gauged the presence of misinformation on Subreddits individually and analyzed the relationships between Subreddits based on their shared users.
-

Methods



Reddit API

Used Python Reddit
API Wrapper



Subreddit Selection

Used a relevance
framework to



Data Collection

Got 800-1000 top
posts per Subreddit



Data Wrangling

Analyzed every URL
for misinformation



Data Analytics

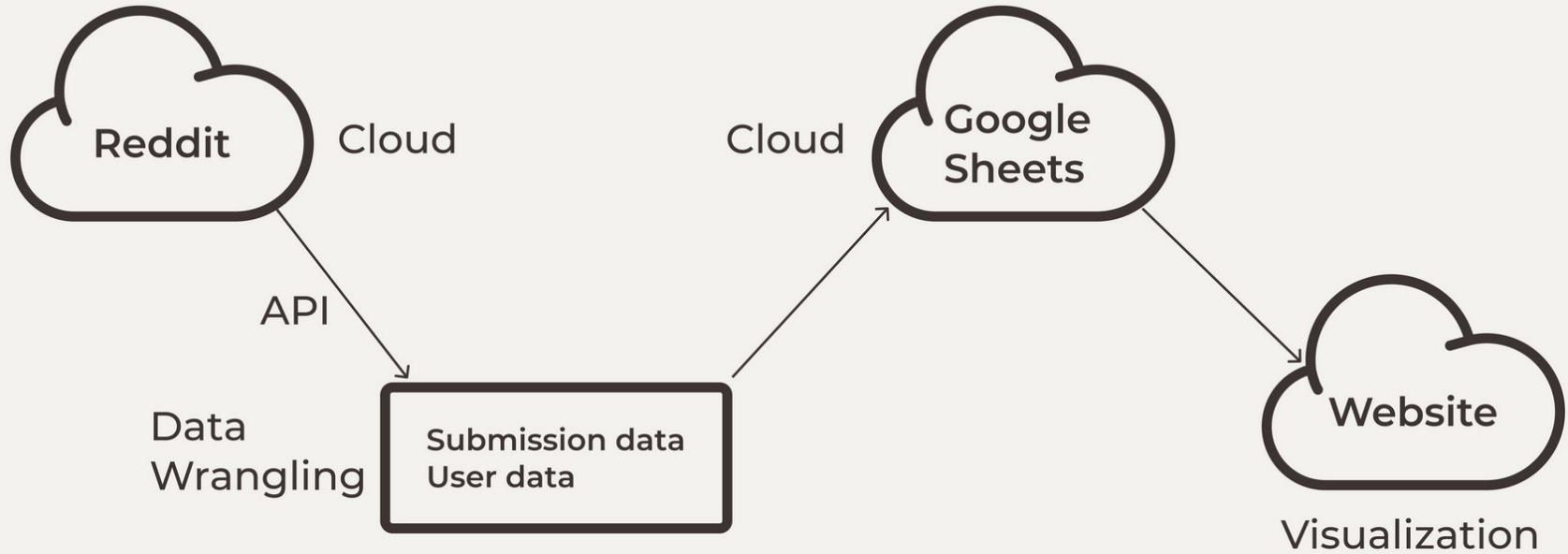
Used data viz and
modeling methods



Iterations

Added new
Subreddits based
misinfo-posters

Data Collection Flow





04

Conclusions

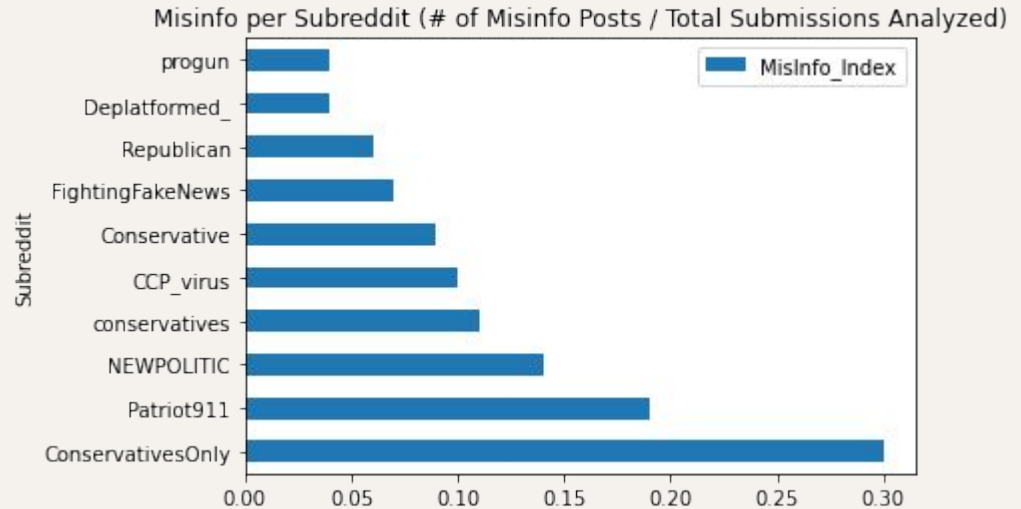
The results of our work and our findings.



Proportion of misinformation per Subreddit

Insight

More political -> more misinformation

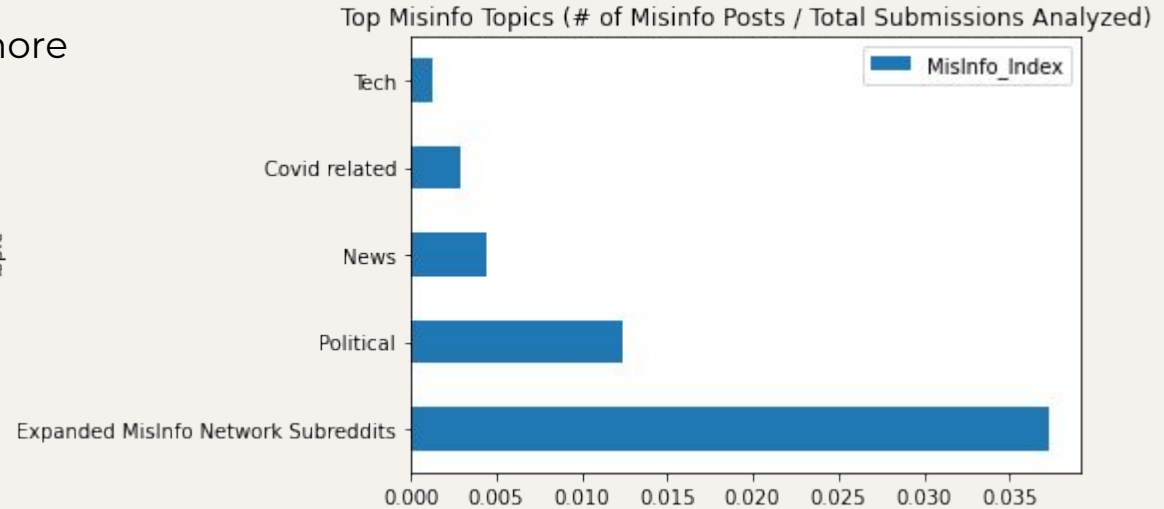


Top Topics

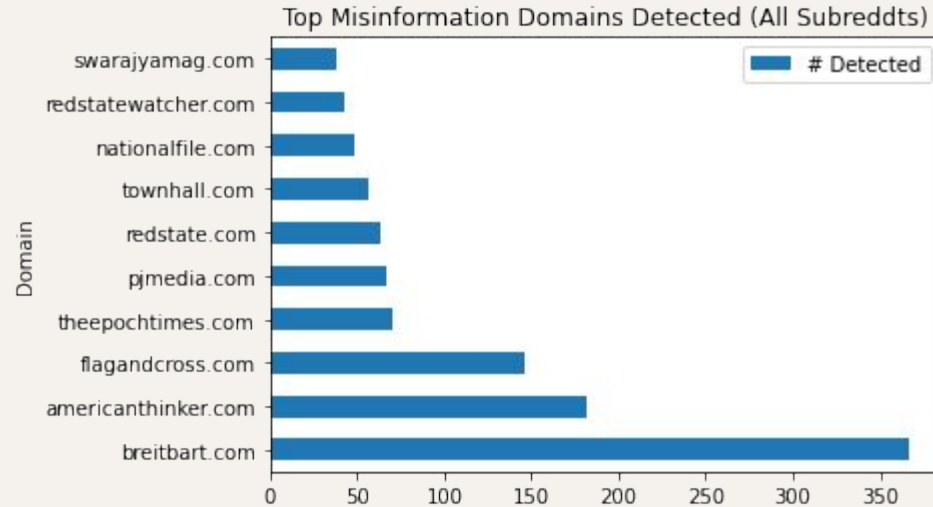
Insight

News & politics -> more misinfo

Topic



Top Domains

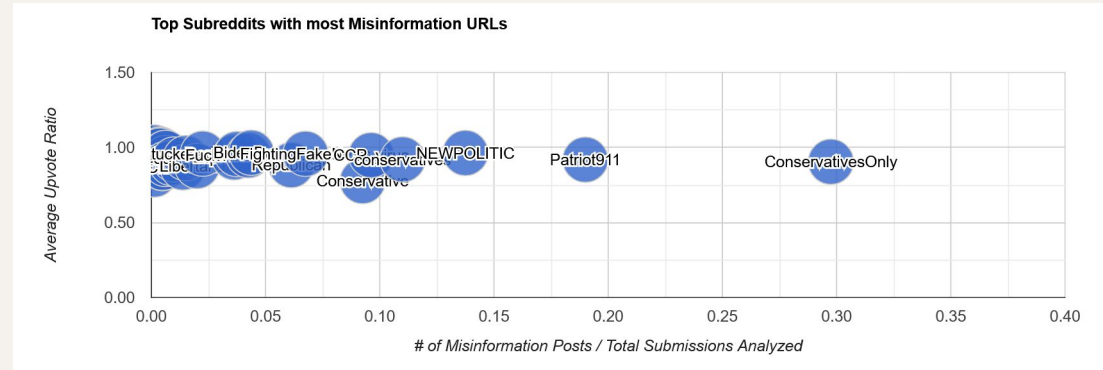


Upvote Ratio versus Misinformation index

Insights

Average Upvote Ratio remains consistent despite proportion of misinformation.

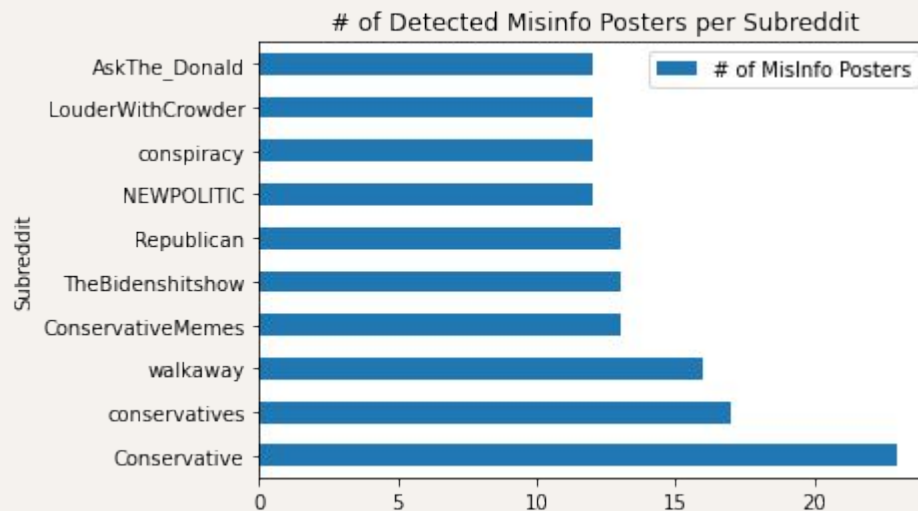
Implies that most subreddits are ideologically homogeneous.



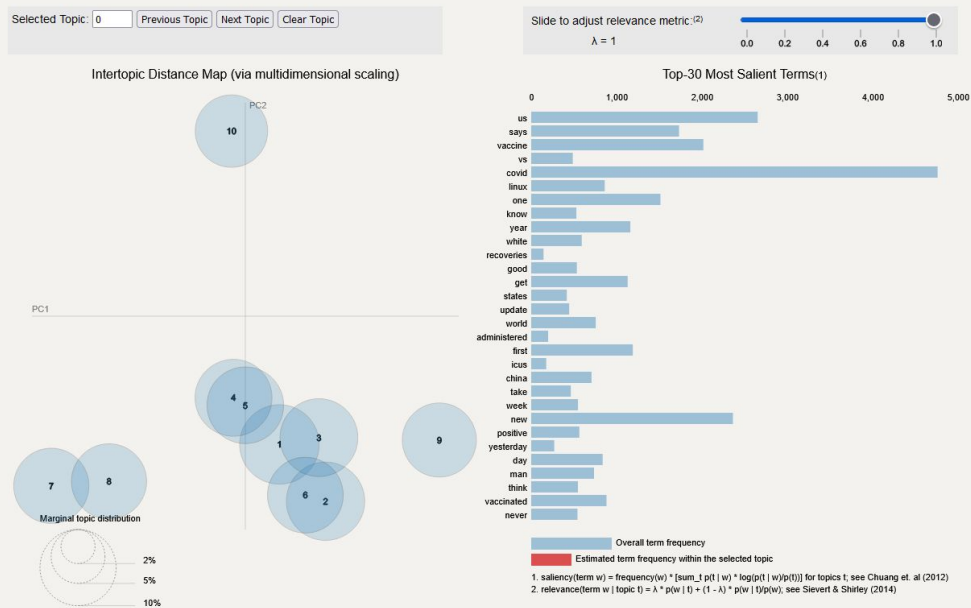
Subreddits with most misinfo posters

Insight

Implies which Subreddits may be more foundational within the misinformation network.



Topic Model Visualization



Undirected Graph Visualization

