

David Aminifard, Samuel Huang

DSC 180A, Section A10

12/05/21

Project Proposal Statement

As social media has grown in popularity, namely Reddit, its use for rapidly sharing information based on categories or topics (subreddits) has had massive implications for how people are usually exposed to information and the quality of the information they interact with. While Reddit has its benefits, e.g. providing instant access to - nearly - real time, categorized information, it has possibly played a role in worsening divisions and the spread of misinformation.

We are aiming to analyze the prevalence of misinformation in certain subreddits and the overlap of users between different subreddits. This is similar to our current project in that we are still analyzing the nature of misinformation in online social media platforms. However, this is different because Reddit fundamentally has a different structure than Twitter. Reddit exists in subgroups called “subreddits” which are typically built around a group of liked-minded people who are interested in a specific topic. Therefore, when studying the misinformation of a topic like COVID-19, it would be more about the subgroups that spread misinformation instead of influential individuals.

We decided to analyze Reddit because we thought it would be interesting to see similarities and differences between the spread of misinformation on Twitter compared to Reddit. While, many times, Twitter is compared to the public square, we believe

Reddit is more comparable to how social interactions occur in real life. Reddit exists in groups of like minded people similar to how humans typically group with people that they consider similar to themselves. Furthermore, on Reddit we can analyze how misinformation relates to certain topics. For example, if we are on a subreddit about Unix, we assume that it would be unlikely that we would find many posts that contain misinformation. However, if we are on a subreddit about politics, it would be more likely that we would find posts that contain misinformation.

So, in short, our project would be to download data using the Python Reddit API Wrapper or PRAW and the data will be used to assess the influence of misinformation given certain Subreddits and to showcase relationships among users who comment in those Subreddits. The PRAW documentation can be found [here](#), where already publicly available subreddit and user data is made accessible, which we will use for our analysis. PRAW's source code (GitHub repo) is available [here](#), it is evidently a popular library used for accessing data on the Reddit platform

We plan to create a website that shows the relationships between subreddits and the spread of misinformation. Also, the website will contain a ranking of subreddits based on the frequency of misinformation per subreddit.