

Project: Understanding Social Interactions on Reddit

Abhishek Yadav, CS12B032

Tapan Chugh, CS12B055

Pranav P. Nair, CS12B062

February 22, 2015

Reddit is a social networking website where users share images and text information. Other users can then upvote, downvote, post comments and resubmit their content. These votes then determine the position of the content on the site. This content is organized by areas of interests called subreddits.

We aim to understand interactions among various users at different time intervals. Users who resubmit the same content in one area, are more likely to resubmit the similar content in other areas. This way we can find users having similar interests.

By making a graph between posts and users, we wish to understand the clustering and ranking on it.

A network can also be made among users who have submitted the same content. We want to look at indentifying hierarchies in networks. We can do gross clustering of a network, even if there are many violations of similarity within the cluster. Then repeatedly cluster within a cluster.

By doing this kinds of analysis, we can understand social interactions between users on reddit.

1 Datasets

We will be using the data set available for reddit on SNAP
<http://snap.stanford.edu/data/web-Reddit.html>