**Networking and Parallel Computation Final Project Report**

James Raboin

Ruben Arutyunov

Jack Morgan



**I. Introduction**

Our group decided to implement a Serial and Map/Reduce solution to judge whether particular Reddit slang words (henceforth known as “redditisms”) garner a net positive or net negative reaction from the Reddit community when they are used in comments.

**II. Background**

Reddit is a popular content aggregator where any user can submit any posts and comments at any time. As of 2017, it boasts 234 million unique monthly visitors and is the 8th most visited website in the world. The posts and comments are then rated/sorted according to their “karma score” which is the sum total of *Upvotes* (think of these like they are “likes” from Facebook) minus the total number of *Downvotes* (dislikes). These votes are only able to be cast once by each user for a given comment (so one user cannot downvote/upvote a single comment more than once).

Within the Reddit community, there are a collection of slang words (mostly acronyms) that have popped up over the last 12 years of the websites life (such as “FTFY” which stand for “Fixed that for you” and “TL;DR” which stands for “Too long; didn’t read”). Our goal was to find out if there was a correlation between the redditisms and their karma score and, if there was, which redditisms scored higher than the others.

**III. Datasets**

We scraped the dataset from reddit using the Praw python library. We selected a static number of posts to scrape (instead of our initial approach to pull the top posts at the current time) as this helped us test if we were pulling the right data. Within each post, our script searches through every comment and the corresponding child comments to see if there are any matches to the group of selected redditisms (listed below).

From there, we stored the comment’s karma and the redditism in a text file. After we scraped the data, we ran it through our serial and parallel implementations to find the average karma per redditism. More details about the specifics of each implementation are in the next sections.

**IV. Serial Implementation**

**V. Parallel Implementation**

**VI. Serial vs. Parallel Comparison**

**VII. Findings**

**VIII. Conclusion**