

Determinants of Social Media Engagement

An Analysis of Reddit Posts

Christopher Messier

General Assembly

Washington, D.C.

November 1, 2017

Overview

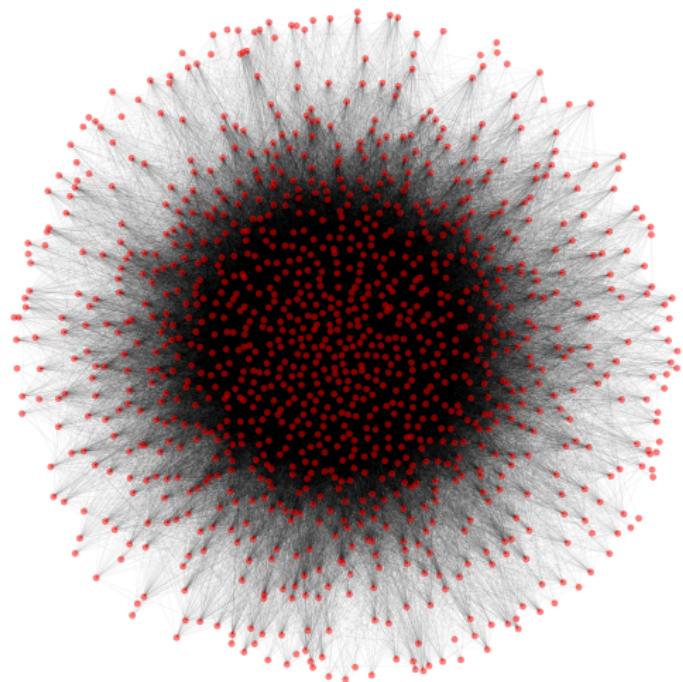
Determinants of
Social Media
Engagement
Messier

Methodology

1 Methodology

Graph of Subreddits by User Cross Posting

Determinants of
Social Media
Engagement
Messier
Methodology



Data was collected using a proprietary python web scraper

- Using r/random to generate a random sample of subreddits
- Up to 2,500 most recent posts on each subreddit
-

Omitted Data:

- Posts from "AutoModerator" bot
- Non-english titles

Modelling:

- Random Forest implemented using Apache Spark

Understanding the Data

Determinants of

Social Media

Engagement

Messier

Methodology

- TFIDF Vectorization of the title

Engagement for a post defined as:

- The number of comments on a post,
- Weighted?

Profits Per Capita

Determinants of

Social Media

Engagement

Messier

Methodology

Distribution of Comments

Determinants of
Social Media
Engagement
Messier

Methodology

Thank You