Abstract:

Introduction:

* Main findings:

Background

* Twitter Introduction
  + What is it?
    - Twitter is a social media platform based on the sharing of short posts. It was originally founded as a microblogging organization, but has expanded into the world of social media. It serves as a way to facilitate discussion and share information and opinions to millions of users.
  + How does it work?
    - Users can share short posts known as “tweets” which other users can react to by “liking” the post or by further sharing the information by “retweeting.” Users can also “follow” other users to keep up-to-date with any information they share.
      * If we use mentions/hashtags in our analysis we can include more about these notions
  + Who uses it?
    - Politicians, celebrities, and everyday people

Datasets

* Where did we get the dataset
  + This data set was obtained from Kaggle and is a collection of information about detecting hate speech with context on Twitter. The dataset contains a network of around 100k users with about 5k out of these users labeled as “hateful users”. Additional information of users such as their network-related activity and content-related information are included in the dataset as well.
* Why did we choose it?
  + As our original plan was to crawl data from Quora but out of security protocol issues we have to abandon our original data source and found this dataset on Kaggle. Our idea is to study the network-related influences of users in the online community from different aspects and fortunately, this dataset matches our interest perfectly.

Methods:

* Building of network
  + Creating edge list and applying attributes such as follower/followee number
* Statistical
* Network theory
* Comparing hateful users and normal users

Statistical Analysis:

* General Characterization
  + Hate-neighbors
  + Normal-neighbors
    - Plot the ratio of (hate neighbors/total neighbors)
  + Statuses count(number of (tweets|follower|followees|favorites) a user has.)
    - To measure influences**(plot)**
  + Followers count**(plot)**
  + Followees count
  + Hashtag numbers
  + Tweet numbers
  + Retweet numbers
    - An ratio plot?
  + Listed count (number of lists a user is in.)
    - To measure influences**(plot)**
  + Retweet numbers
  + Lists
  + Quote numbers
  + mentions
  + Status length(time permitting)
  + suspended(?)
  + Conclude the behavior of hateful users
* Correlation plots we need:
  + Followers count vs Statuses count
  + Follower count vs Retweet count
  + Follower count vs Followee count
  + Mention count vs Retweet count

Influence Estimation:

* Centralities
* Page Rank
* Identifying influencers
* Identifying hubs (time permitting)
* Subgroups (time permitting)
* Small-world (time-permitting)
* strongly connected component
  + average shortest path length
  + diameter
* strongly connected subgraph
  + average shortest path length
  + diameter
* Correlation plots we need (?)
* Take-aways

Related work:

* Opinion manipulation(?)

Conclusion:

(Acknowledgments)

(References)

Presentation:

* Mostly figures
  + Explain our findings