

## Project Proposal

Yu Xuan Liu, Roy Tu, Nikhil Unni

## 1 Project Description

The project would be a Pagerank application. Just like how you can measure popularity of websites through how sites link to each other, I thought it would be interesting to analyze how popular Twitter celebrities are based on who retweets who, and perhaps who “likes” who (in terms of liking a Tweet).

It’s an interesting problem, but one difference between this and website linking is that Twitter is mostly huge celebrities that may not link to other big celebrities. My hypothesis is that people who retweet have very few people who retweet them, but this project is to explore that.

Getting the data is pretty easy through Twitter’s API. I just need to scrape a large enough dataset, and then just apply Pagerank to the constructed graph. I’m planning on implementing the “probability” of taking the edge to another account as simply counting the proportion of retweets/likes to that account. Another possible problem is that this graph may be really, really sparse. But that also seems to be a problem with web pages too, so maybe it won’t be an issue. It would be much more neat if I could get an irreducible MC from the data.