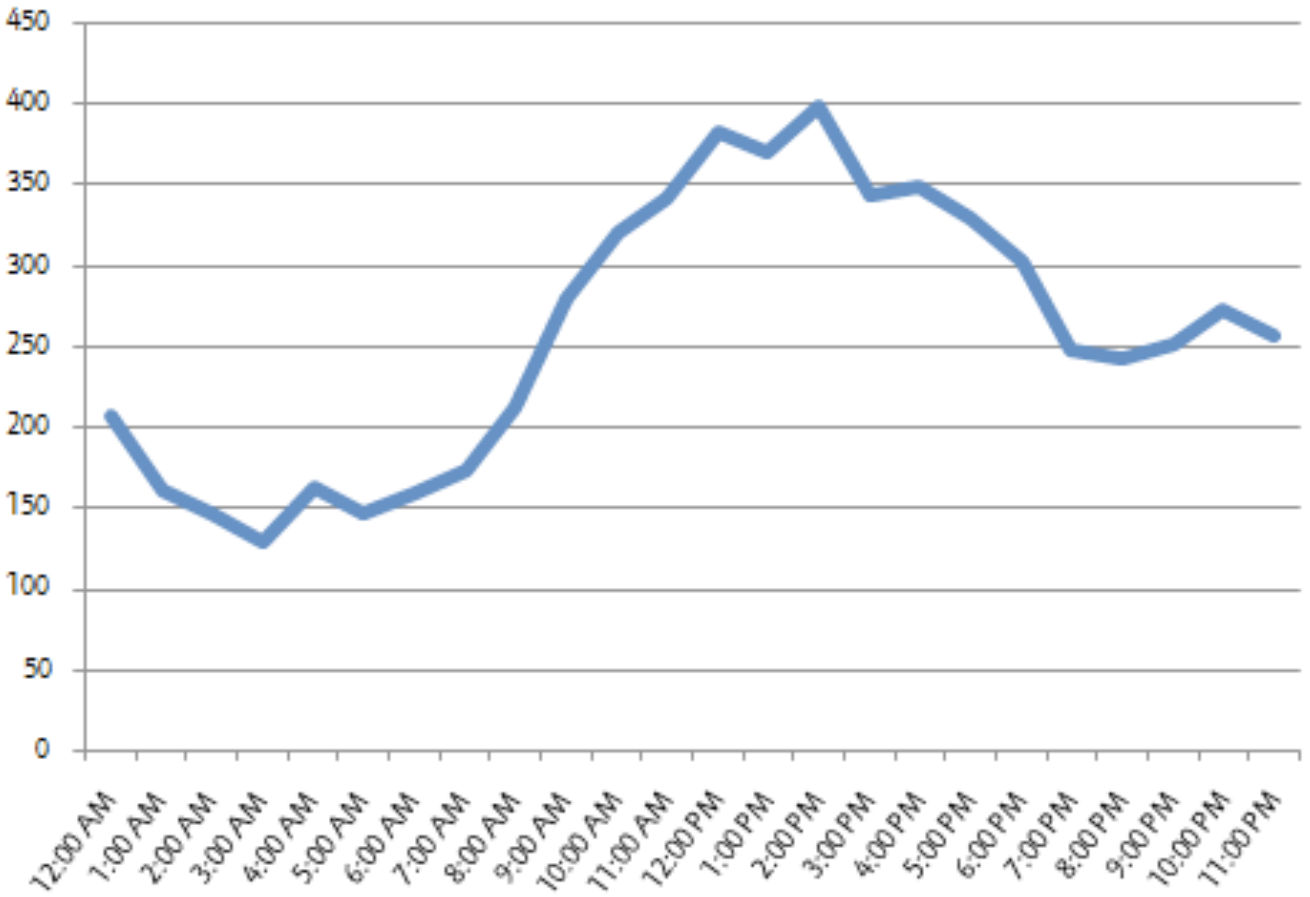
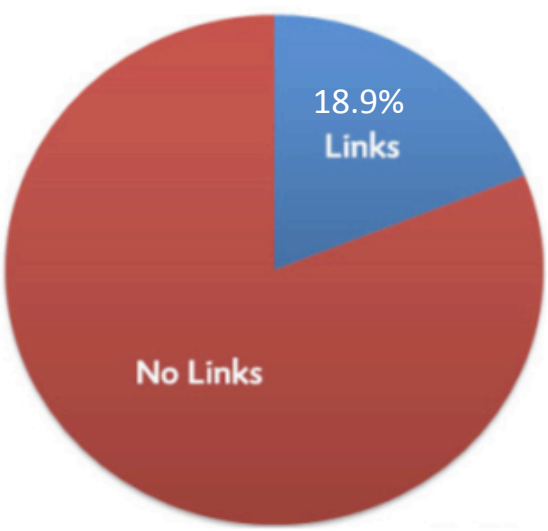
Simple Twitter Stastics

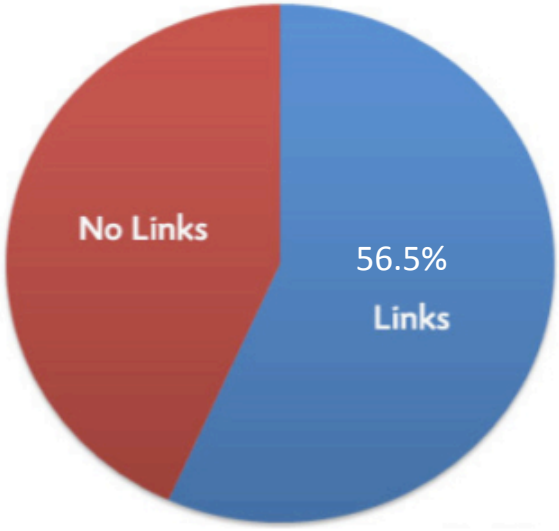
When Retweets occur the most



How many time a link occurs in a retweet



All Tweets



ReTweets

Common Words

Most Retweeted

1. you
2. twitter
3. please
4. retweet
5. post
6. blog
7. social
8. free
9. media
10. help

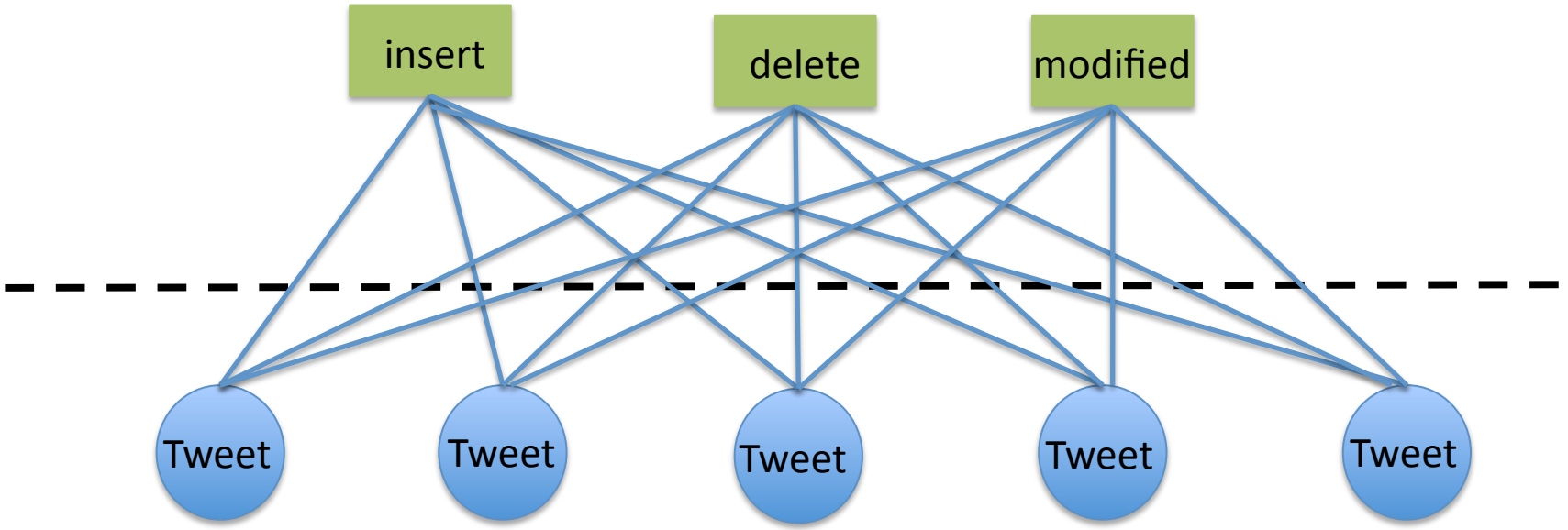
Least Retweeted

1. game
2. going
3. haha
4. lol
5. but
6. watching
7. work
8. home
9. night
10. bed

My Hidden Markov Model

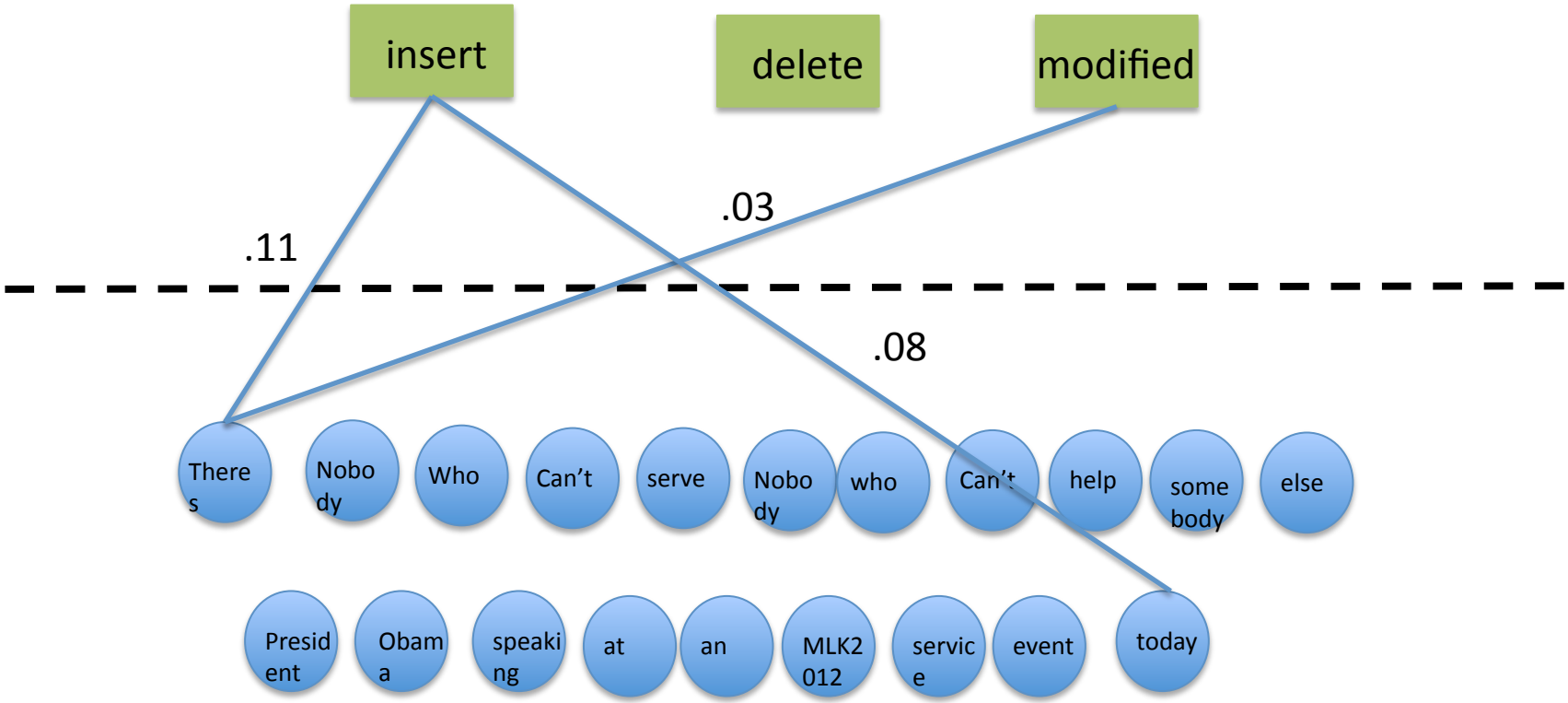
- Super Simple HMM
  - In theory every word in the tweet is an observed state and would need to predict what the probability of having an insertion or deletion
  - Insertion or deletion occurs more frequently at the beginning or the end

Sample Example



BarackObama Barack Obama
"There's nobody who can't serve. Nobody who can't help somebody else."—President Obama speaking at an #MLK2012 service event today

This tweet has 4,236 retweets

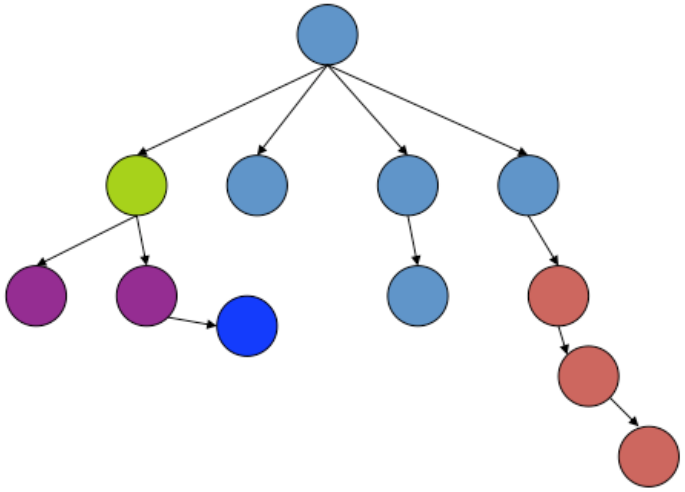


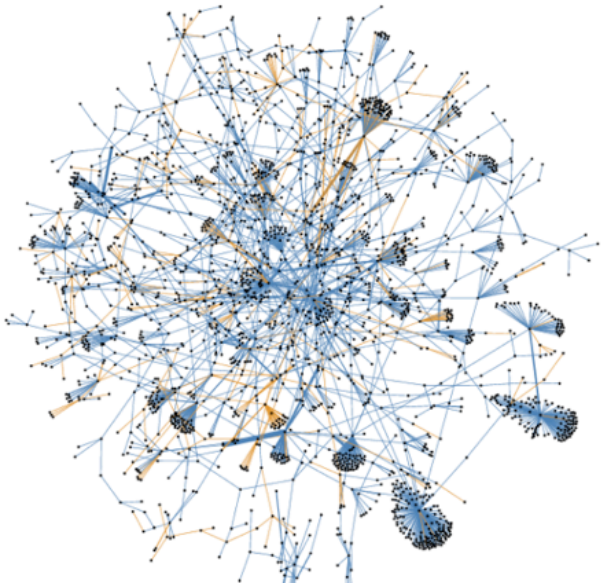
Most significant probabilities. There were more insertions at beginning and End and some modifications at the beginning. Not so much other places

- Using this HMM model, do a multiple sequence alignment
  - Still not getting it to work, my model still may be wrong, but getting closer
- Interesting observation is that most of the retweets is the majority of retweets happen between 2 people with a depth of 2
  - We need to pick prominent people on twitter to have more depths to study, like the Obama tweet shown previously

Another Idea

- So I want to create two different trees using the same data
  - One tree would be created by the multiple sequence alignment
  - The other tree would be created just by the twitter data, like time and who retweeted who
    - Already have examples of these done by other people
    - Want to compare with my multiple sequence tree





Multiple Sequence Alignment VS Followers directed network