For the exploratory part we wish to explore some possible features. Features can be of 2 types: per tweet features and per user features.

1) We will develop a unified function that takes in bundles of features from the two groups and then return metrics (eg: chi square etc) to determine how important the feature is.

2) We also wish to study the changes in a single user in the ‘per tweet’ features. For example perhaps we can find some patterns on how the features evolve over time. We may find a significant difference in some features before and after the onset of schizophrenia.

3) We can apply PCA to some of the low dimensional features and plot them in 2D or 3D and explore if we can observe clusters (manually or through K-NN).

Below is a list of features that we wish to explore through the above mentioned approaches.

**Depression through twitter paper**

Possible features based on twitter specific interactions:

1) Engagement: Volume of posts, Proportion of reply posts, fraction of retweets, proportion of links shared, insomnia index (time of the day tweets were made)

2) Social graph: Construct a graph of user and the people he/she interacts with and then consider the graph features like node, dyadic and network properties

*The rest of the stuff in this paper is analysis of tweet sentiment etc which you guys are writing (LIWC, tre based complexity etc).*