Facebook Status Sentiment Analysis
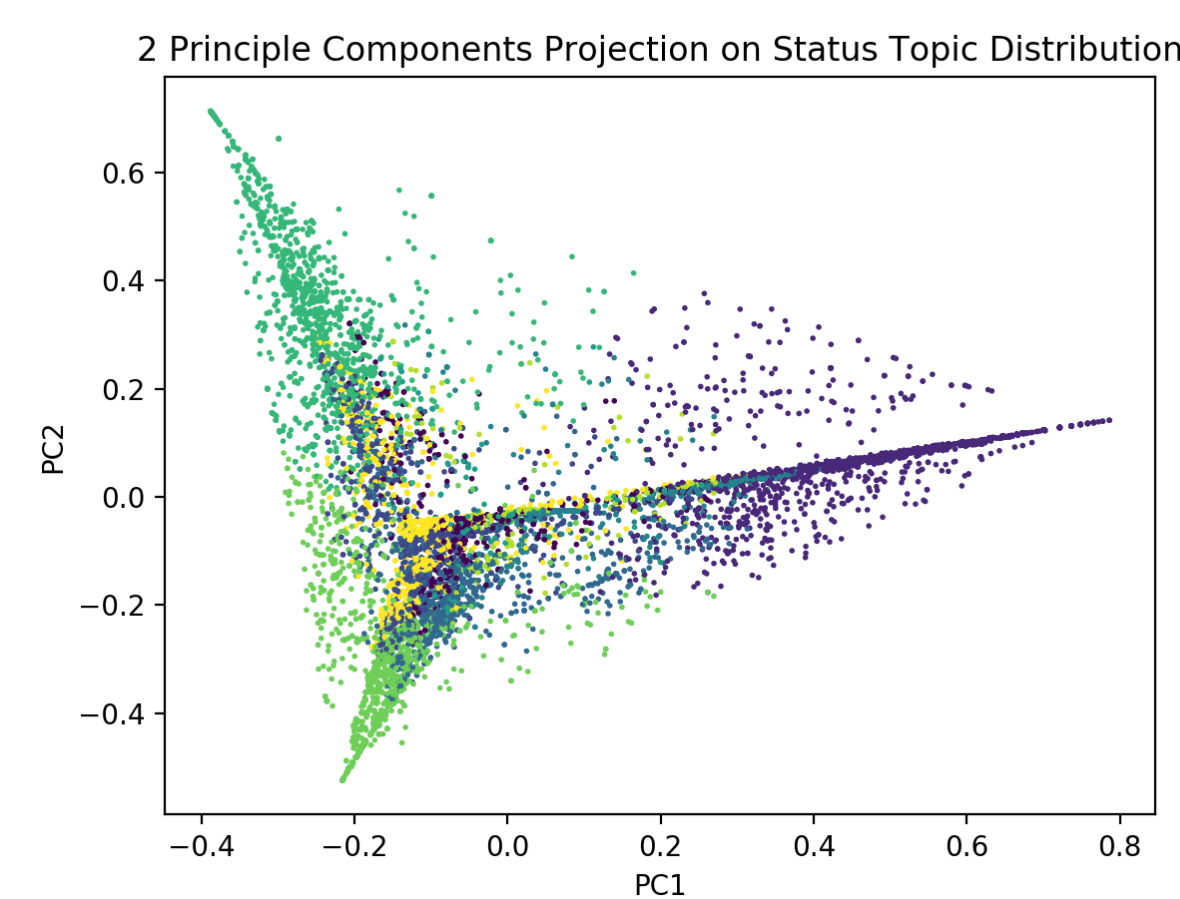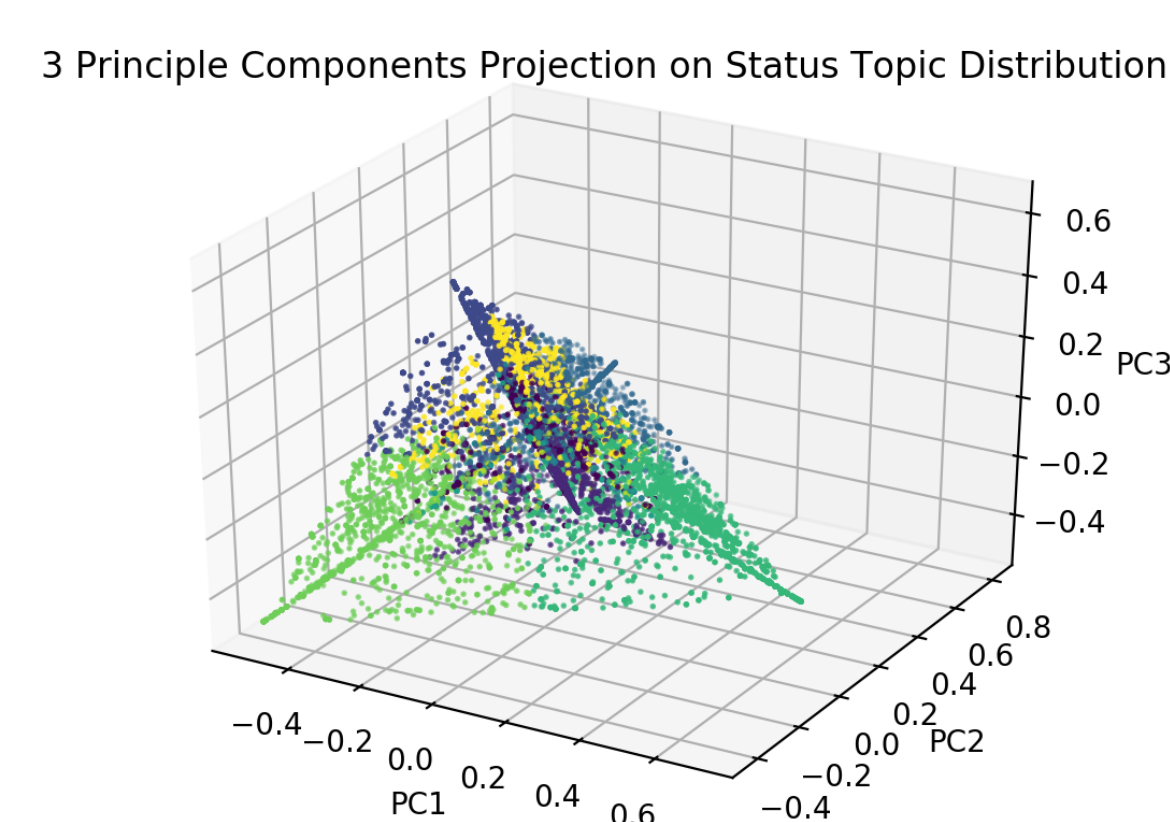
Hailey Lee, Alex Ahn, Jae Park

INTRODUCTION

Sentiment Analysis is more than figuring out how people feel about in the social media. With sophisticated analysis on how people react to certain topics, sentiment analysis can predict the following: campaign success, marketing strategy, product messaging, customer service, and stock market price.

Recently, **Facebook** extended "like" by including "**angry**", "**wow**", "**haha**", "**love**", and "**sad**" reactions on posts, comments, and even messages.
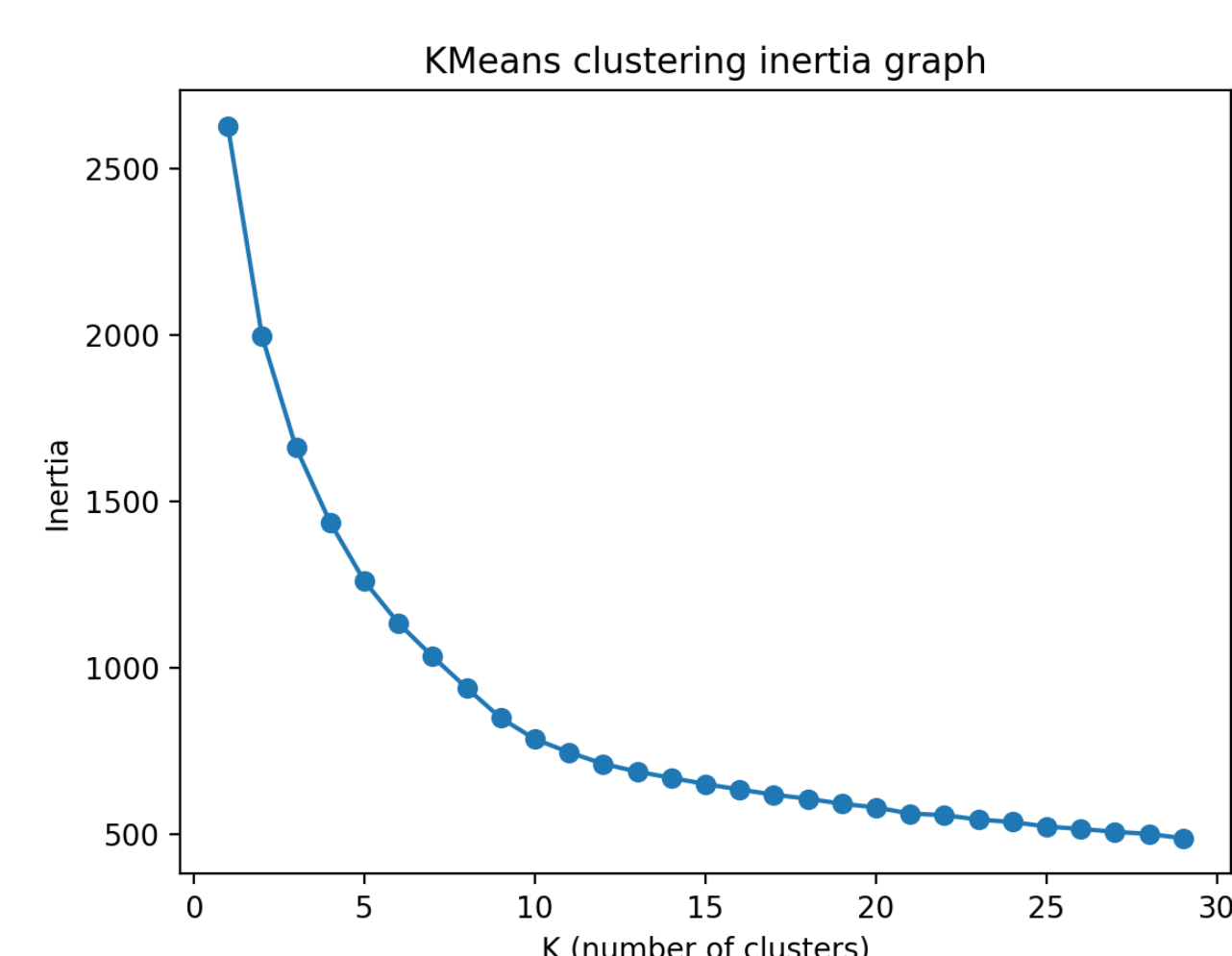
FIGURES

2 Principle Components Projection on Status Topic Distribution
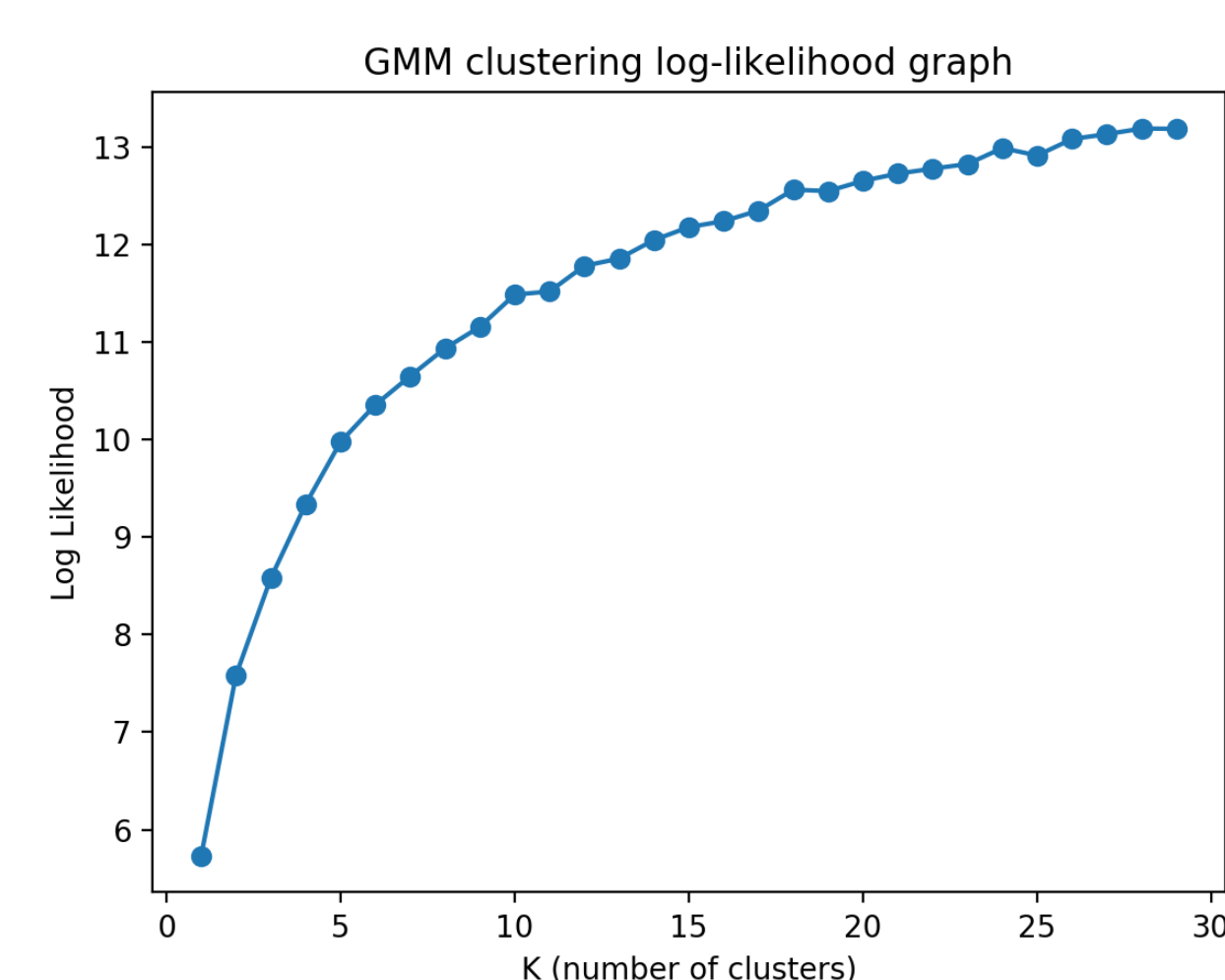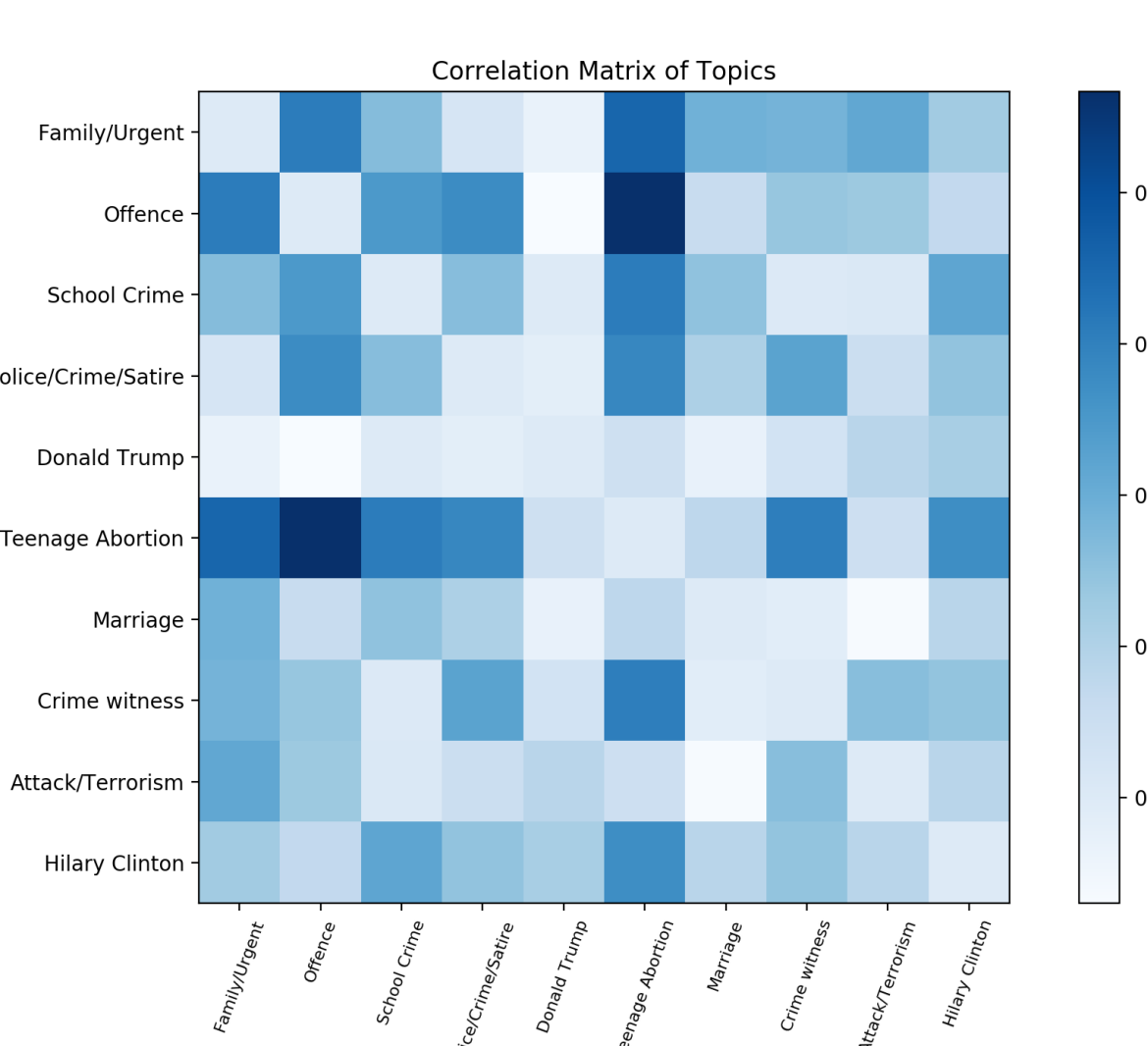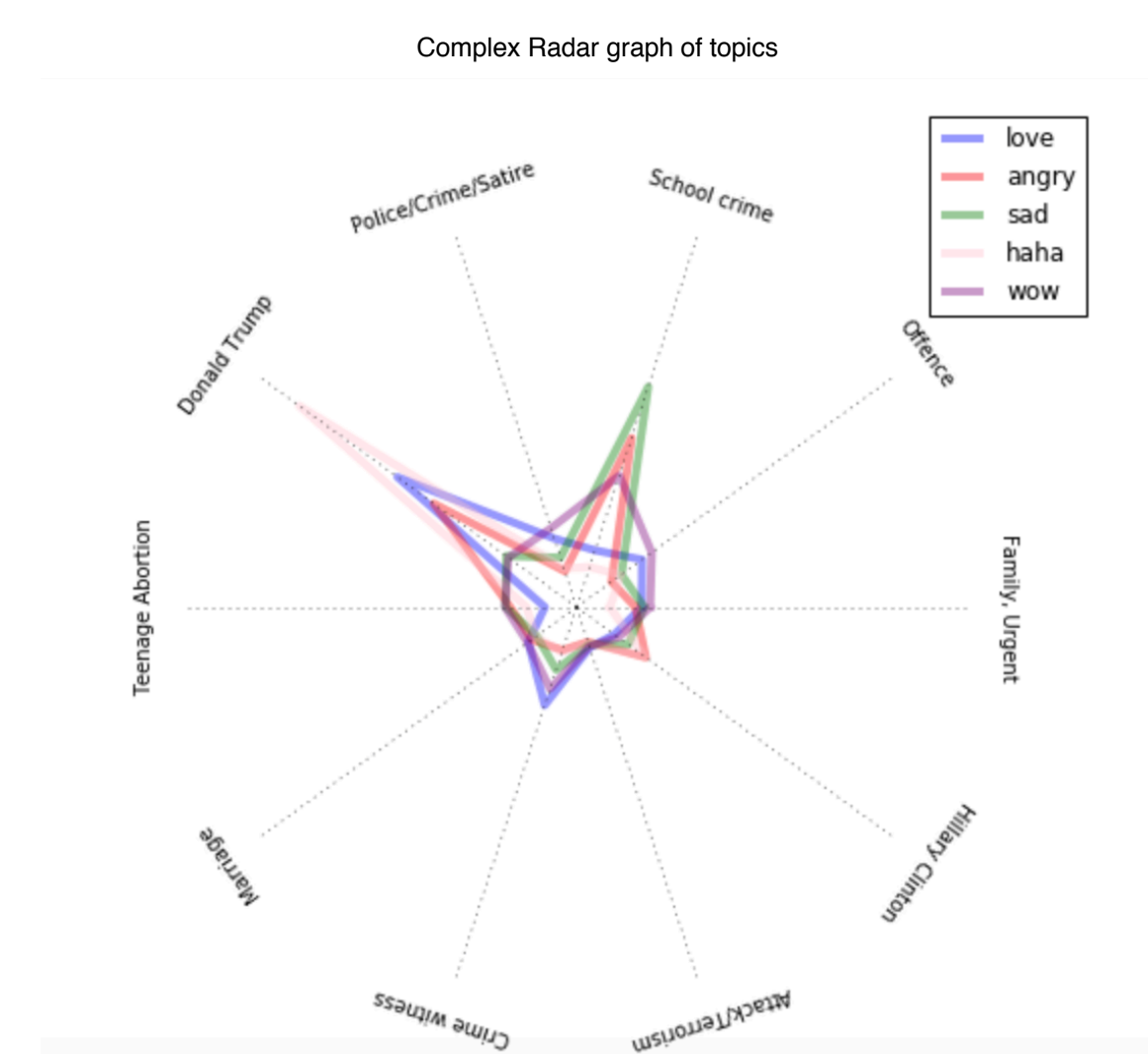
3 Principle Components Projection on Status Topic Distribution

Explained variance with 2 eigenvectors: 0.390087%

Explained variance with 3 eigenvectors: 0.531836%

Complex Radar graph of topics

Correlation Matrix of Topics

GMM clustering log-likelihood graph

KMeans clustering inertia graph

OBJECTIVE

1. **Extract** data from a public Facebook page
2. **Feature Learning:** learning topic distributions of status using LDA (Latent Dirichlet Allocation)
3. **Visualize** Facebook statuses using PCA
4. **Clustering:** perform clustering to analyze distributions of learned features

METHODS

1. Select a public Facebook page for user sentiment analysis
2. Collect data including status and user responses
3. Perform LDA on status corpus and generate topic vectors
4. Assign most relevant topic label to each status
5. Perform PCA to visualize topic distribution
6. Visualize correlation matrix of the topic distribution
7. Visualize user responses to each topic
8. Cluster status corpus using topic-features (KMeans, GMM)
9. Plot likelihood/inertia for each K-number of clusters for each method
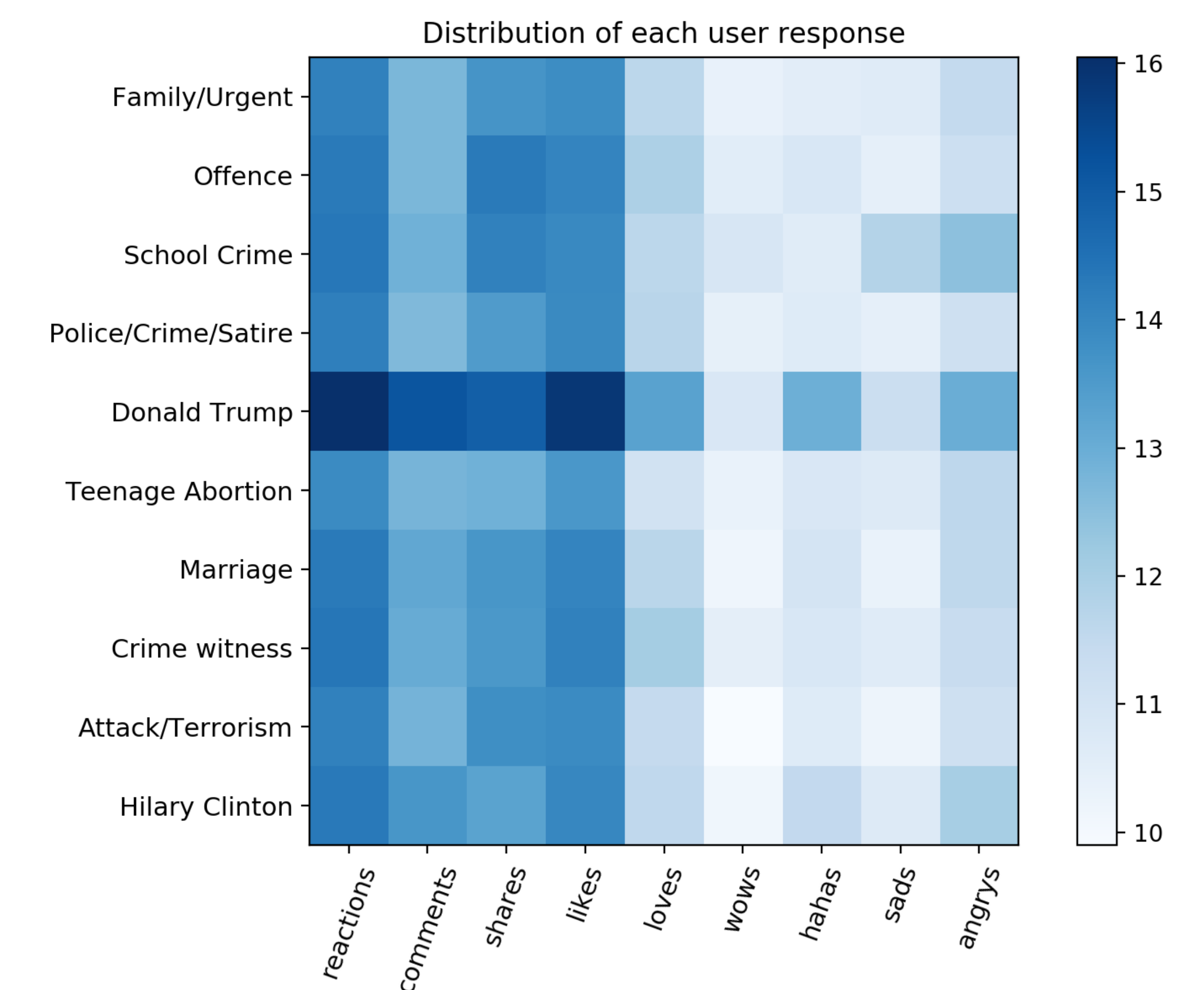
Generated Topics

Topic #1 (Offence):
woman man young words knew began didn judge stop suddenly got went took felt inside later received way home offended

Topic #3 (Police/Crime/Satire):
police car officer woman saw wasn life pulled just officers left door arrived won couple looked hard cop driver soon
.
.
.

RESULTS

Distribution of each user response

CONCLUSION

Able to observe/tracked distributions of different topics under "Opposing View" in the Facebook group page.
We chose 10 topics to be able to interpret the results more clearly.
From August 2016 to April 2017, as we can see from the graph, Donald Trump took the majority of attention, especially because it is politically highly debated topic.
From clustering, (both K-means and GMM), we observed that likelihood approximately converges after 10 cluster mean. This means that topics are usually combined with one another for each status.
According to the correlation matrix, we noticed that Teenage abortions, Family issues, and Offense are highly correlated with one another. Police Crime Satire is likely correlated with crime witness, while Donald Trump is not correlated with anything in topic distribution of each status.
Further implications: apply to the other page, and we can possibly predict how people react sensitively to certain topics.

REFERENCES

[1] https://github.com/minimaxir/facebook-page-post-scraper
[2] https://medium.com/@baditaflorin/understanding-facebook-reactions-using-sentiment-analysis-f17b6e561ff3