

A comparison of depression-related text across social media platforms

Agenda

- 1 Background & Project Objective
- 2 Social Media Platform Deep Dives
- ③ Social Media Platform Comparisons
- 4 Conclusions & Next Steps

Background & Project Objectives

Part 1

Background

- People suffering from various mental health illnesses often turn to social media for support and resources.
- It is often difficult to identify depression, so organizations are looking for new ways to find and reach out to those suffering with depression.
- It can be insightful to understand how depressed users interact with different social media websites, since behavior varies greatly among these outlets.

Text Associated with Depression

- First-person singular pronouns increased use of I (Rude et al., 2004)
- Absolutist words ex. always, totally, entire, must, have to, never, etc. (Al-Mosaiwi & Johnstone 2018)
- Negative sentiment and general sadness (Tausczik & Pennebaker, 2009 & De Choudhury et al., 2013)
- Anxiety (Coppersmith et al., 2014)
- Anger (Coppersmith et al., 2014)
- Swearing (Resnik et al., 2015)
- Death interest (Coppersmith et al., 2015)
- Key life domains such as work or school (De Choudhury et al., 2013)
- Expression of medical concern and religious involvement (De Choudhury et al., 2013)

| Project Objectives

- Gain insight into how depression is expressed and discussed across social media and forums
- Explore similarities and differences of the expression of depression across each platform and relate these findings to previous research
- With this information, we hope that public health organizations and mental health advocacy groups can better target their outreach to support depressed individuals

Methodology

twitter

rtweet package

Depressed Sample: 169,472 Tweets containing the word 'depressed'

Random Sample: 71,896 Most recent tweets

reddit

RedditExtractoR

Depressed Sample: 385 Posts and 2197 Comments from depression-related subreddits

Random Sample: Unable to pull, so used eRisk (CLEF) data (below)

tumbir.

Screen Scraping & tumblR

Depressed Sample: 1607: #depressed, 2497: #depression, 4218: #depressing, 4290:

#depressive, 2192: #sad

Random Sample: 1512 trending posts

Blog & Forum

Screen Scraping

Depressed Sample: 34,307 posts from thread in depressed forum in Beyond Blue and

Psychcentral

Random Sample: 113,469 Posts from general discussion threads in Straight Dope

Early Risk Detection (CLEF):
Data from eRisk Lab

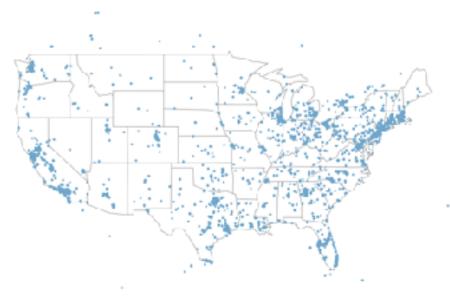
Labeled reddit data (45,241 depressed user posts and 334,764 non-depressed user posts) from 2017

Part 2

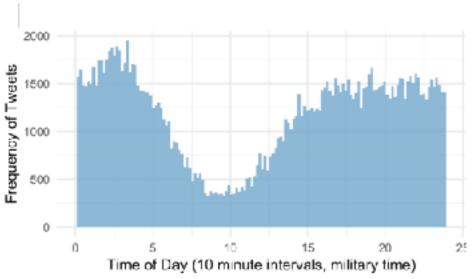
Social Media Platform Deep Dives



'DEPRESSED' GEOLOCATIONS



FREQUENCY OF 'DEPRESSED' TWEETS BY THE HOUR

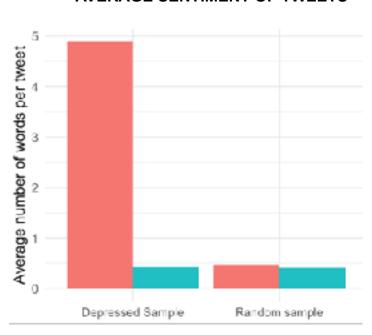


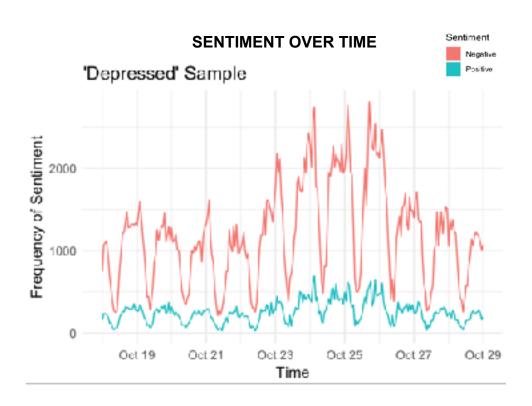
Source: Data collected from Twitter's REST API via rtweet

Twitter - SENTIMENT ANALYSIS



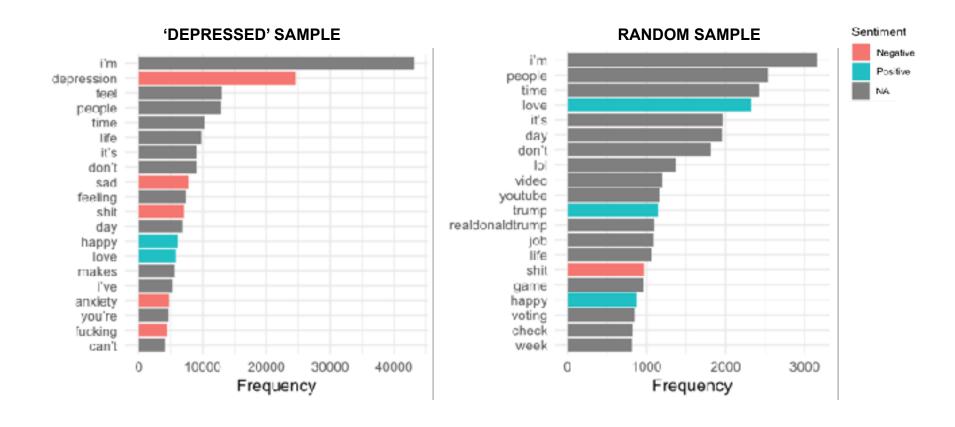
AVERAGE SENTIMENT OF TWEETS





Twitter - SENTIMENT ANALYSIS



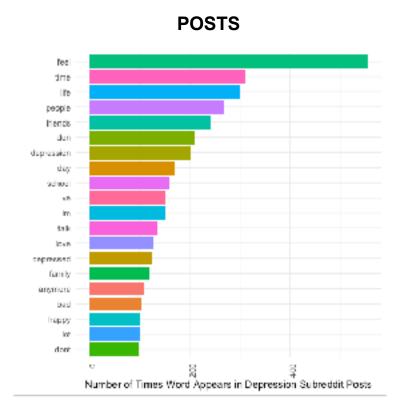


Twitter Summary

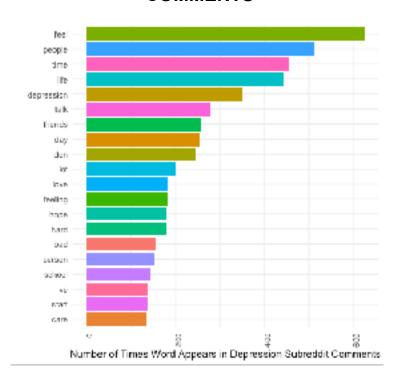
Mirrors previous research on text and depressions: Increased use of first person singular pronouns, negative sentiment, absolutist words, swearing

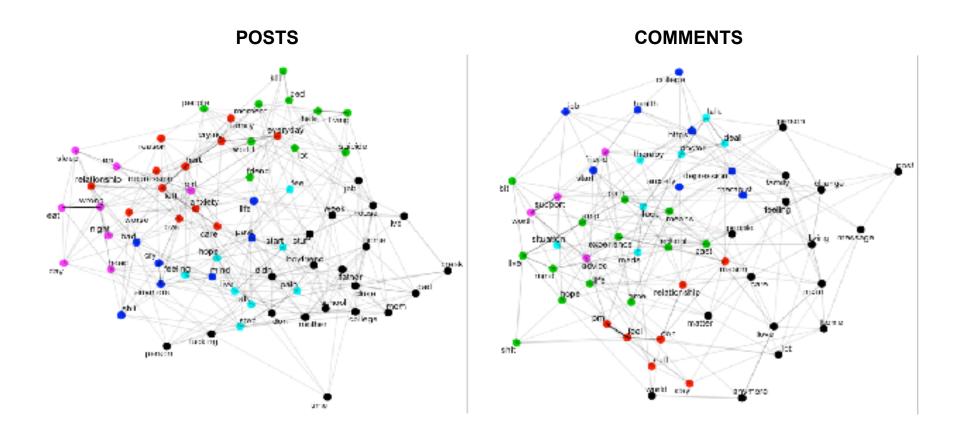


Reddit - FREQUENT WORDS



COMMENTS





Reddit Summary

Broadly looking at unigrams, users talk about key life domains and receive support and empathy from commenters. Some users treat Reddit as a social media site (talking about daily frustrations, feelings, etc.), while others treat it as a support forum (looking for advice and guidance).



Tumbir - DICTIONARY ANALYSIS

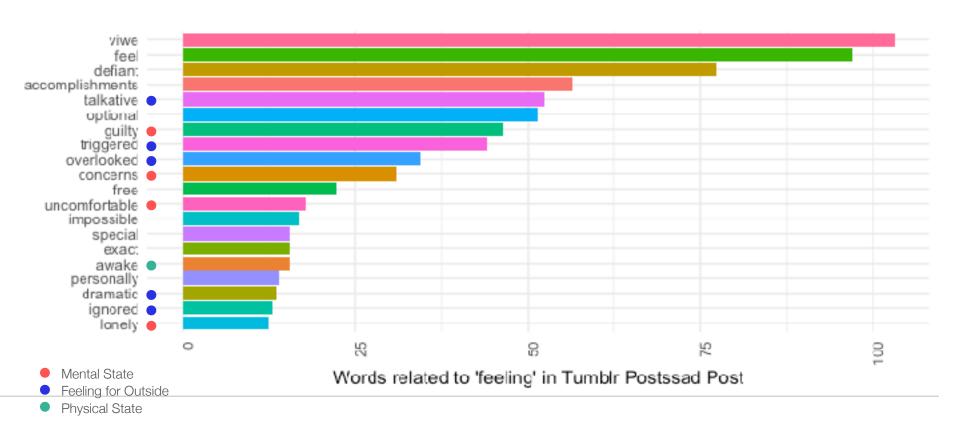




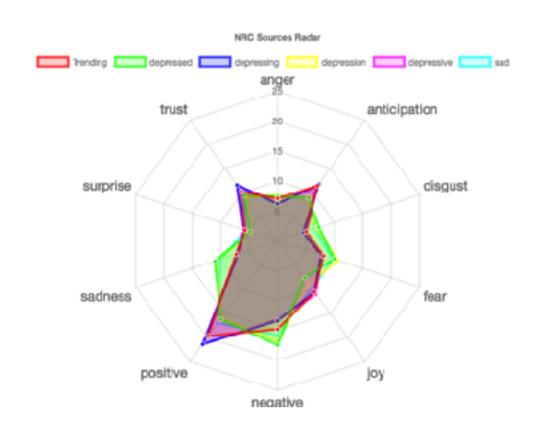


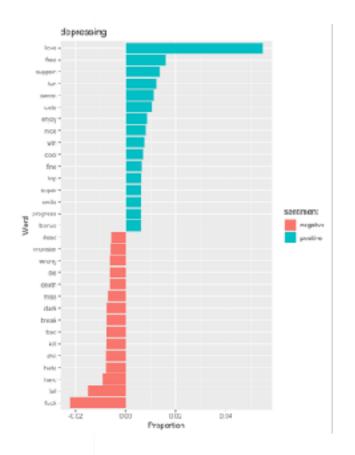
Tumblr - TEXT ANALYSIS - Dive into 'Feel'

tumbir.







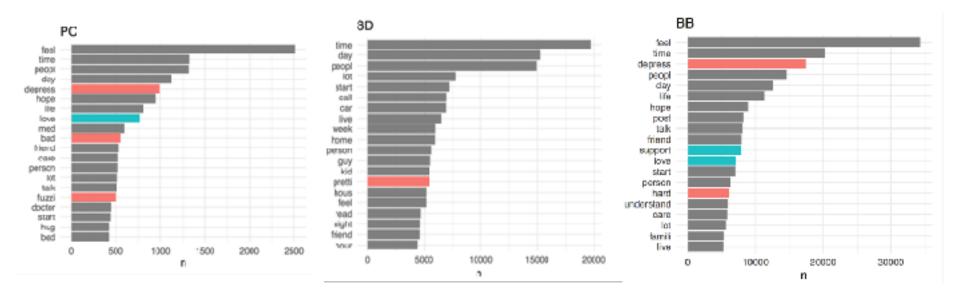


Tumblr Summary

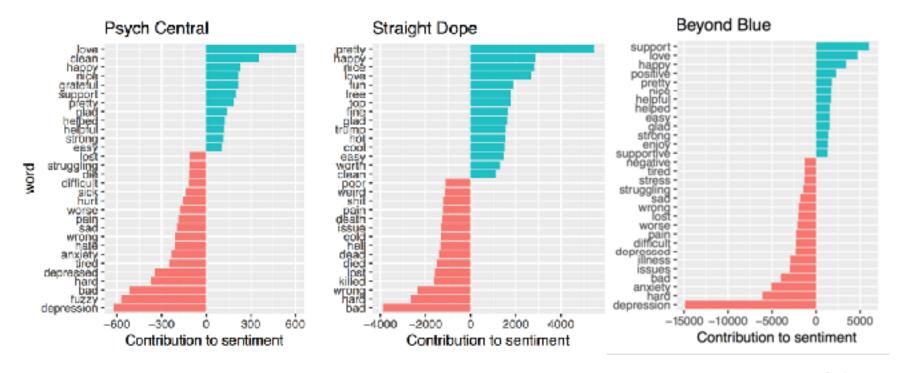
Express personal feelings like 'guilty', 'concerned'

'#Depressing' Posts tend to be more supportive



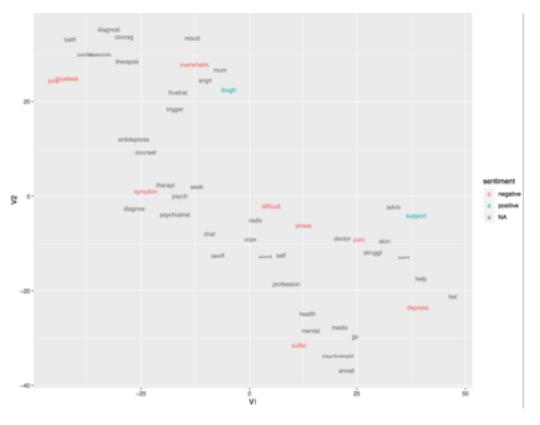








TSNE PLOT WITH 50 WORDS SIMILAR TO DEPRESS



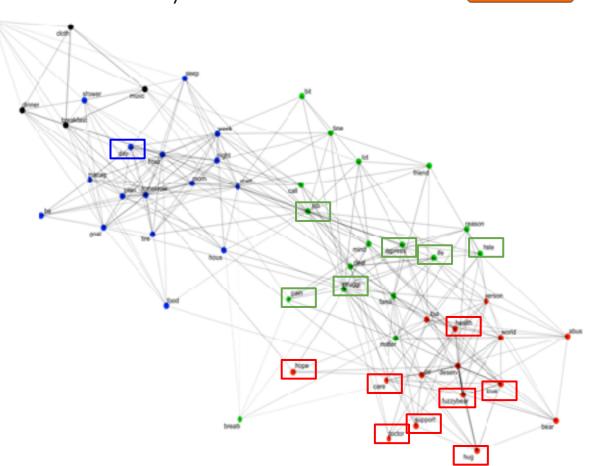
| Forums & Blogs - Text Network Analysis

Psychcentral (PC)

Nodes represent words

Edges represent the use of two nodes(words) by the same person

Drop nodes and edges with low frequencies



Forums & Blogs - Text Network Analysis

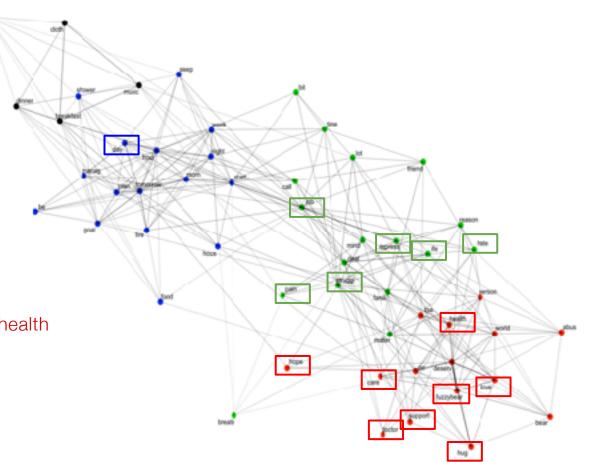
Psychcentral (PC)

Blue & Black: daily words

Green: depressive words

Red: supportive words

Central: stuggl, depress, fuzzybear, health



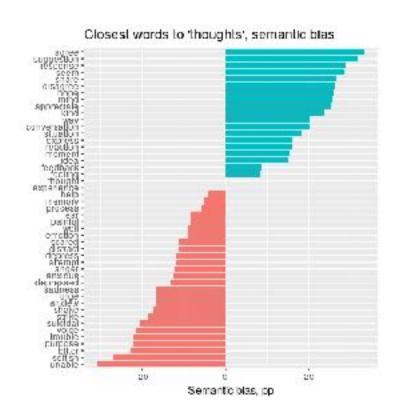
| Forums Summary

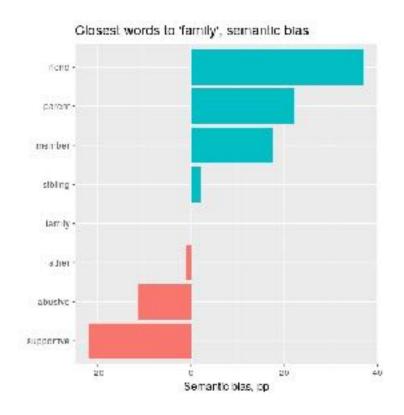
People seek for support and help in forums, they are open to share more personal things like the process of therapy and daily lives.

We found that absolutist words were more frequent in depressed forums compared to the random sample.

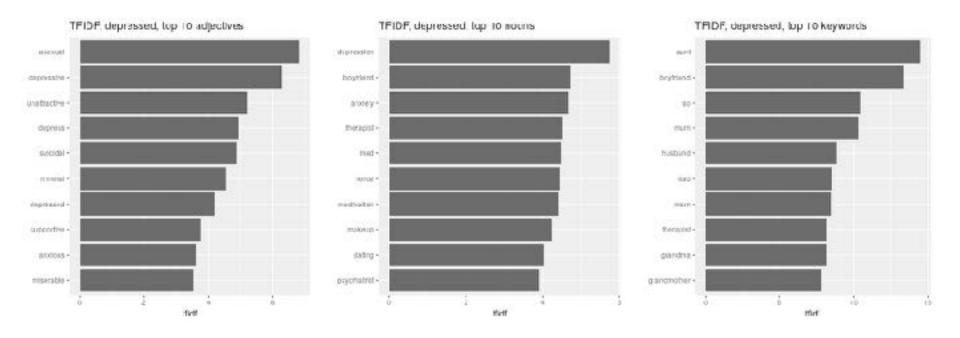


Early Risk Detection - WORD2VEC SEMANTIC BIAS: THOUGHTS, FAMILY





Early Risk Detection - TFIDF: TOP WORDS IN DEPRESSED PEOPLE

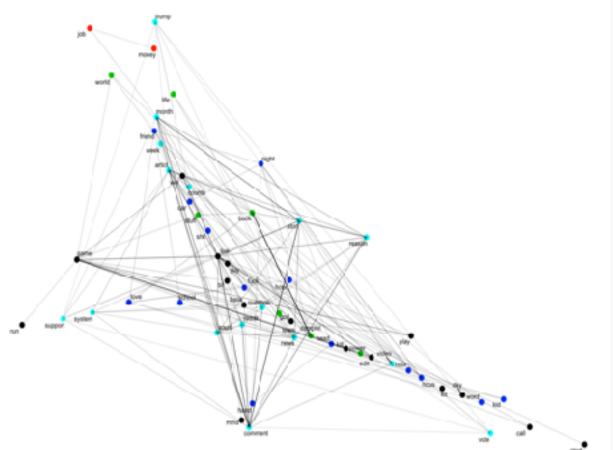


| Early Risk Detection - Text Network Analysis



ER Depressed

| Early Risk Detection - Text Network Analysis



ER General

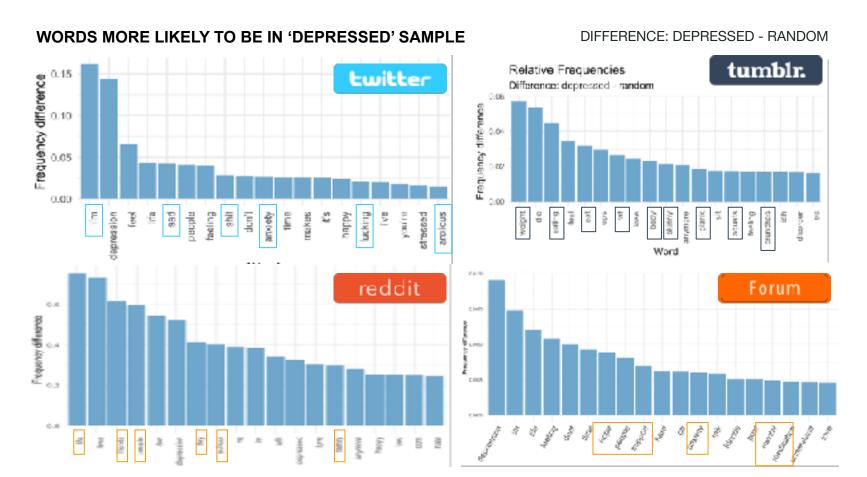
| Early Risk Detection Summary

Using word2vec we were able to detect a semantic bias between depressed people and not depressed people.

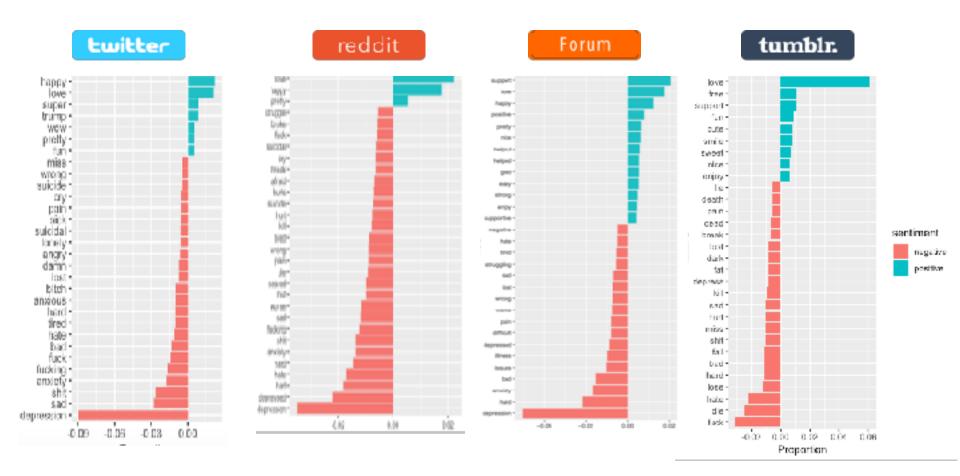
Part 3

Social Media Platform Comparisons

Dictionary-Based Analysis: Relative Frequencies

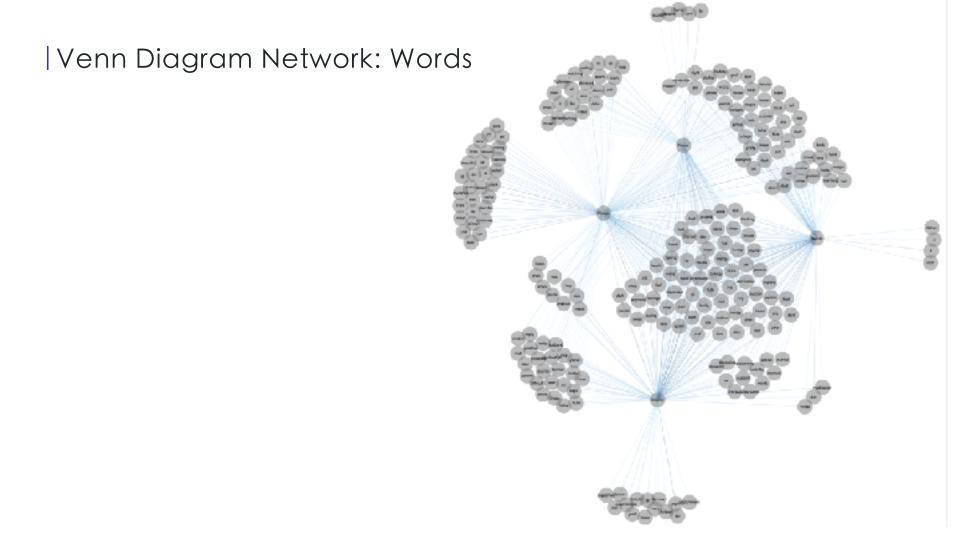


Sentiment Analysis: Bing Dictionary - 30 TOP WORDS

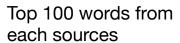


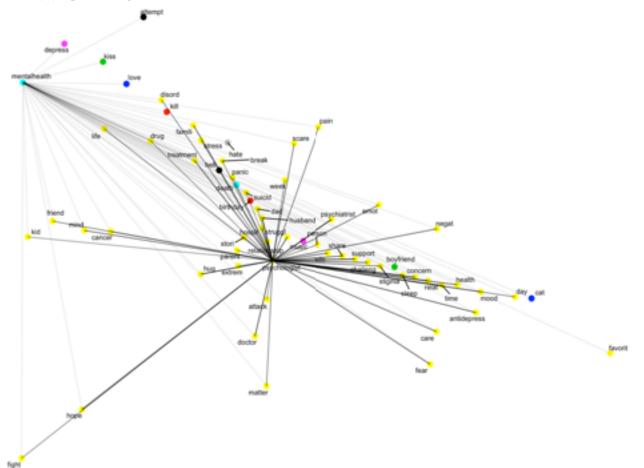
DEMO

NCR Combined Source



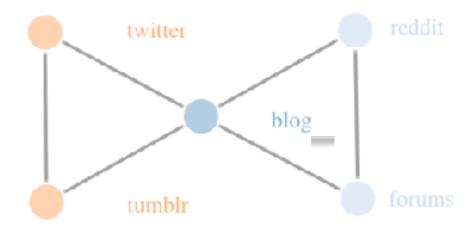
Combined Network - WORDS





Combined Network - PLATFORMS

Separation between social media platforms and blogs/forums



Conclusions & Next Steps

Part 4

Conclusions

- Depression is not expressed uniformly across social media platforms and blogs
- Text across a variety of social media platform and forums reflect the findings of previous research examining language of those suffering from depression
 - High use of first person singular pronouns, increased negative sentiment, swearing, anxiety, death, everyday life
- Language around depression on Reddit more closely resembles forums, while text on social media sites Tumblr and Twitter are similar to each other.
- Public health organizations can use this delineation to better target care campaigns (ex. Focus eating disorder outreach on Tumblr)

Challenges

- Issues related to scraping depression-related data
 - Lack of ground truth
 - Differences in blog and forum formatting
 - Anonymity vs. Identification
 - Subjectivity to time period of data collection
 - Special Characters
- Difficulty interpreting topic models and text networks
 - Depressive sentiment is often found in hidden subtext
 - More time and data might yield different results
- Dealing with broadness of 'depression'
 - Mental health disorders are overlapping in symptom nature (ex. Anxiety, suicidal ideation, etc.)
 - Users themselves often do not know how to express what they are going through

Next Steps

- Additional analysis
 - LIWC
 - Part-of-speech tagging
- Further comparison between our findings and previous research
- Follow-up study to find a solid ground truth
 - Psychological assessment
 - Social media handles
- Improve upon networking and topic modeling approaches to gather new insights on depressed population.
- Build a classification algorithm to identify text from people potentially suffering from depression.

Thank You

Anna Berman, Viggy Kumaresan, Azucena Morales, Allen Wang, Sicong Zhao



References

- Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Quantifying mental health signals in Twitter. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pages 51–60.
- Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pages 1–10.
- Philip Resnik, William Armstrong, Leonardo Claudino, and Thang Nguyen. 2015a. The University of Maryland CLPsych 2015 shared task system. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pages 54–60.
- Mohanned Al-Mosaiwi and Tom Johnstone. 'In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation'. *Clinical Psychological Science*, Vol. 6(4), 2018 pp. 529–542.
- Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. ICWSM, 13:1–10.

References cont.

Stephanie Rude et al. "Language Use of Depressed and Depression-Vulnerable College Students." *Cognition & Emotion*, vol. 18, no. 8, 2004, pp. 1121–1133.

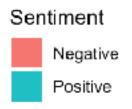
Yla R. Tausczik and James W. Pennebaker. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods." *Journal of Language and Social Psychology*, vol. 29, no. 1, 2009, pp. 24–54.

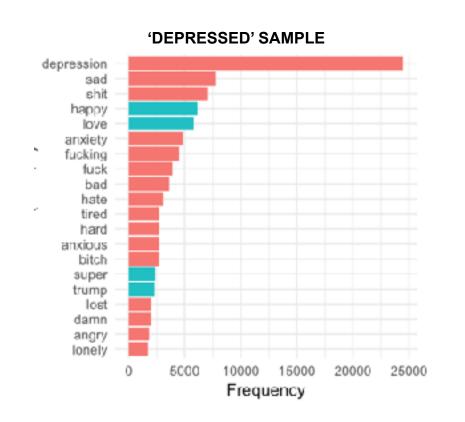
Appendix

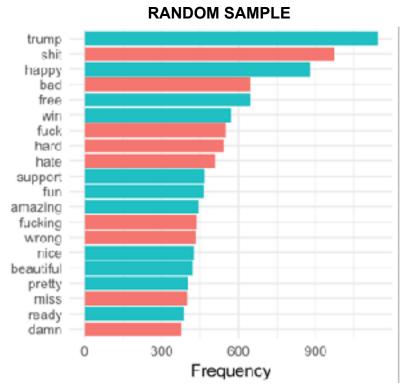
Individual Sources

Twitter

SENTIMENT ANALYSIS

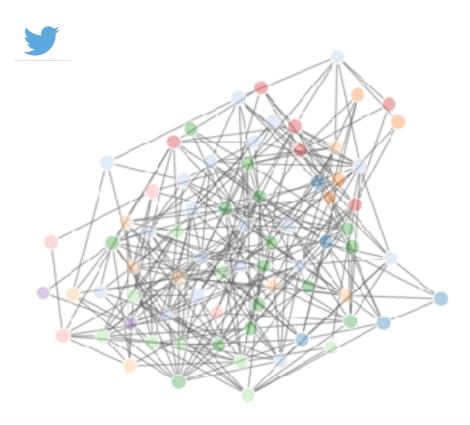


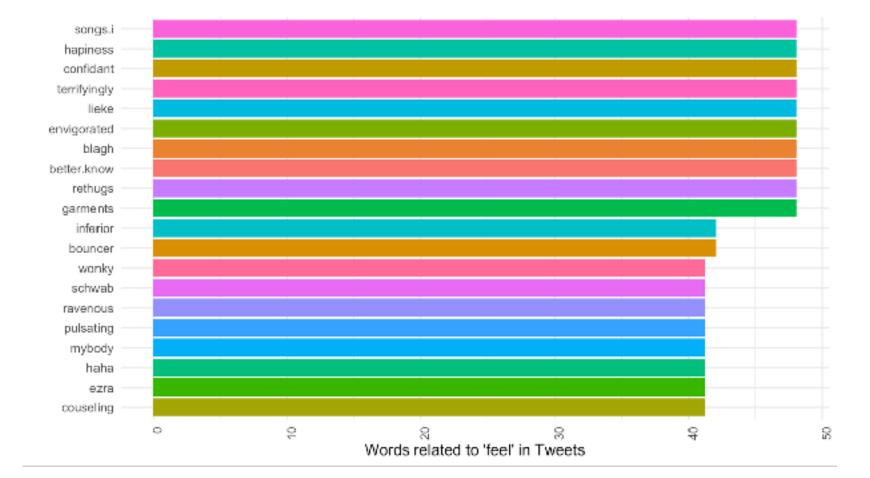




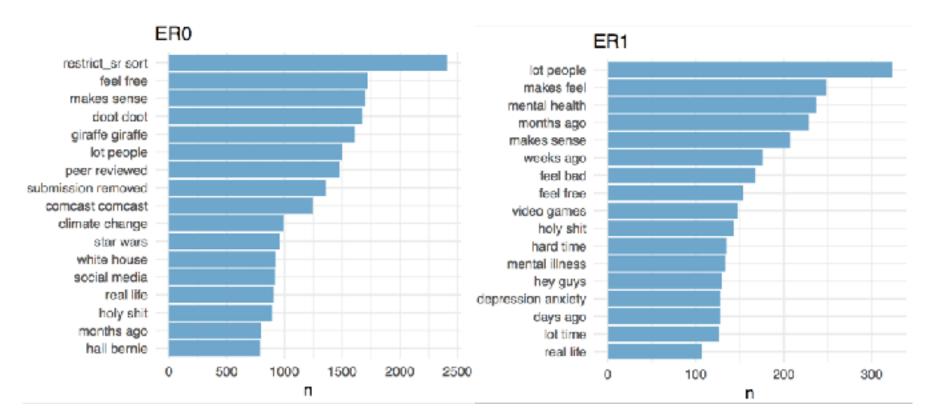
Networks

WITHIN PLATFORM COMPARISON

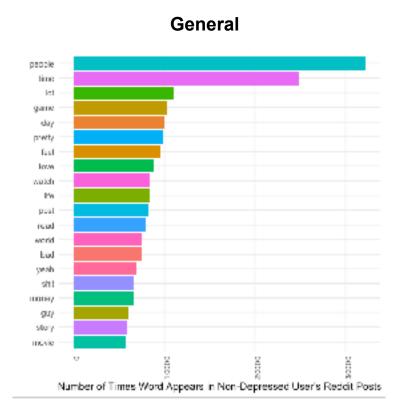


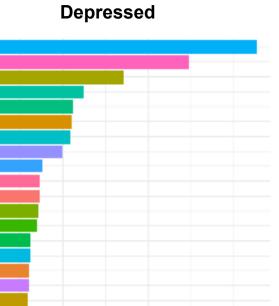


FREQUENT BIGRAMS



FREQUENT WORDS





Number of Times Word Appears in Depressed User's Reddit Posts

people

time

feel

life day

laws:

pretty.

yeah

litiends.

game

hard

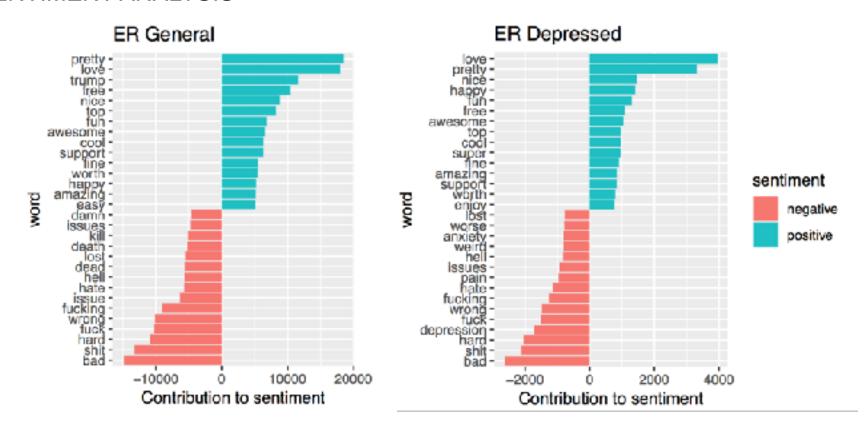
makes

echnol

atacted.

depression

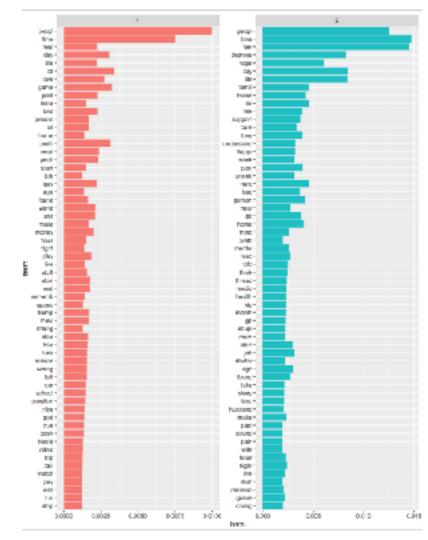
SENTIMENT ANALYSIS



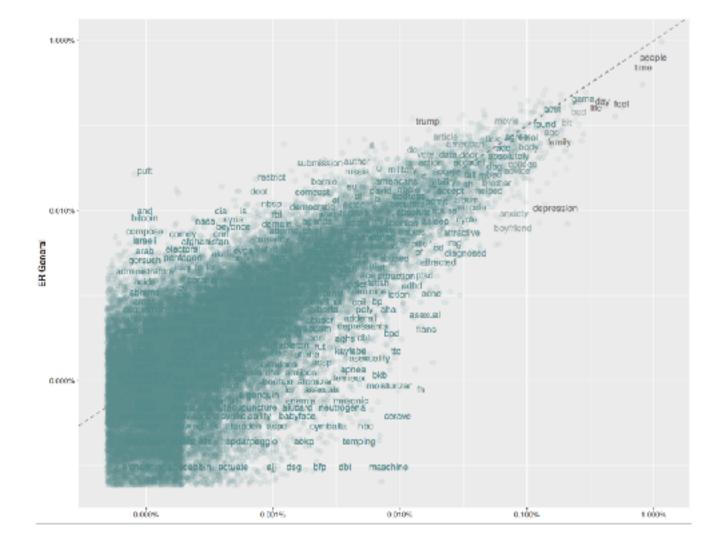
ection

Is this a combination of all forum and ER data?

What's the takeaway?

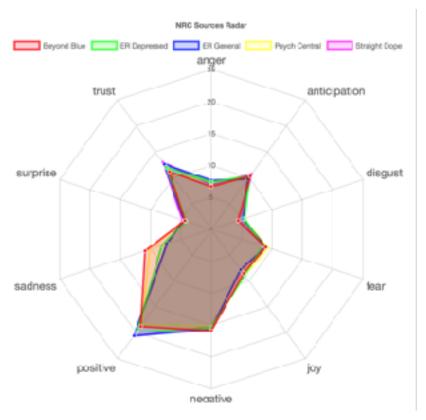


SHARED WORDS



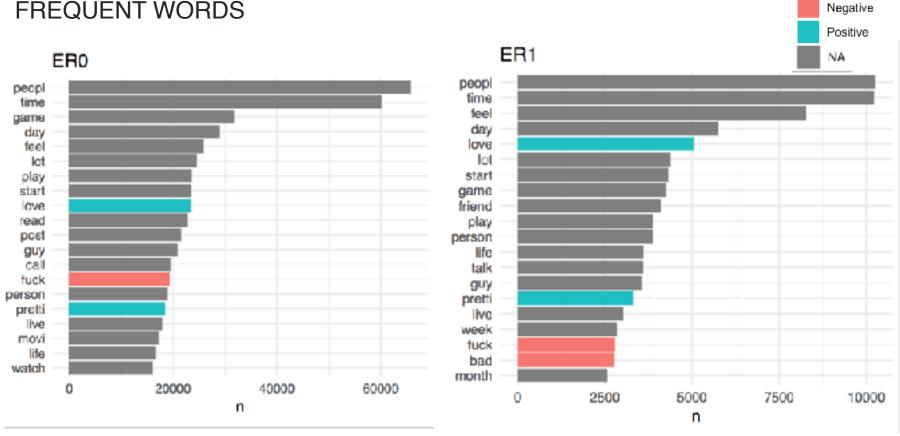
Forums & Early Risk Detection

NCR RADAR



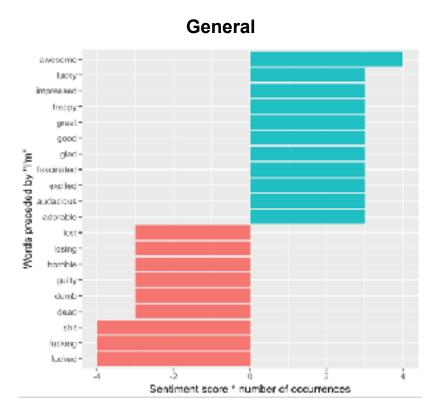


FREQUENT WORDS

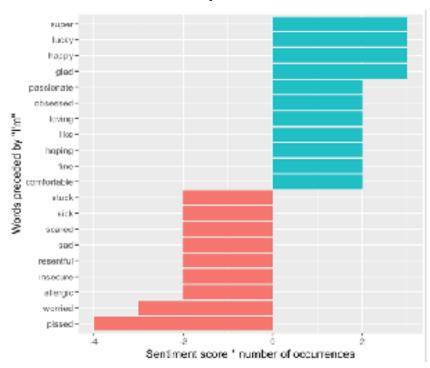


Sentiment

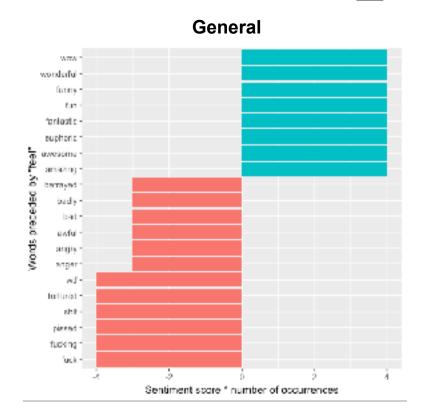
BIGRAMS SENTIMENT: 'I'M ___'



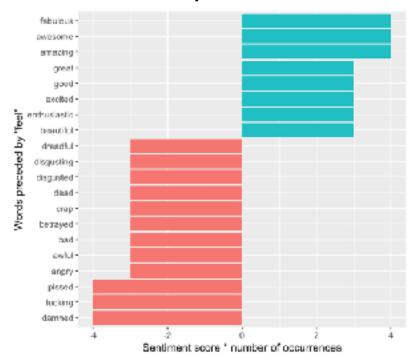
Depressed



BIGRAMS SENTIMENT: 'FEEL ___'

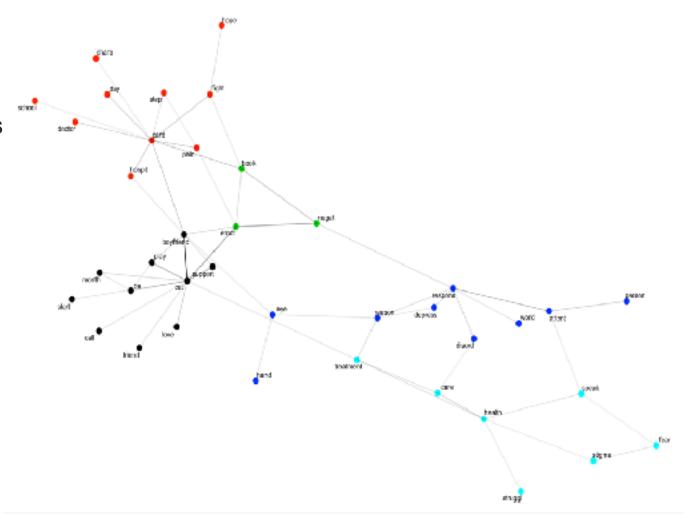


Depressed



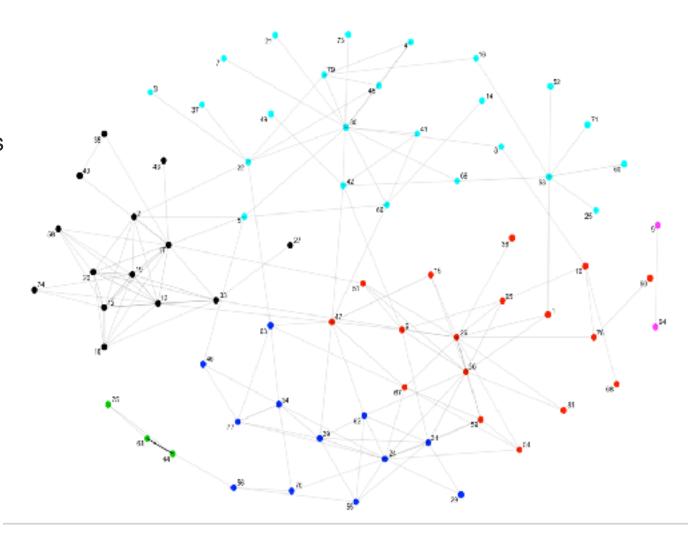
Text Network Analysis

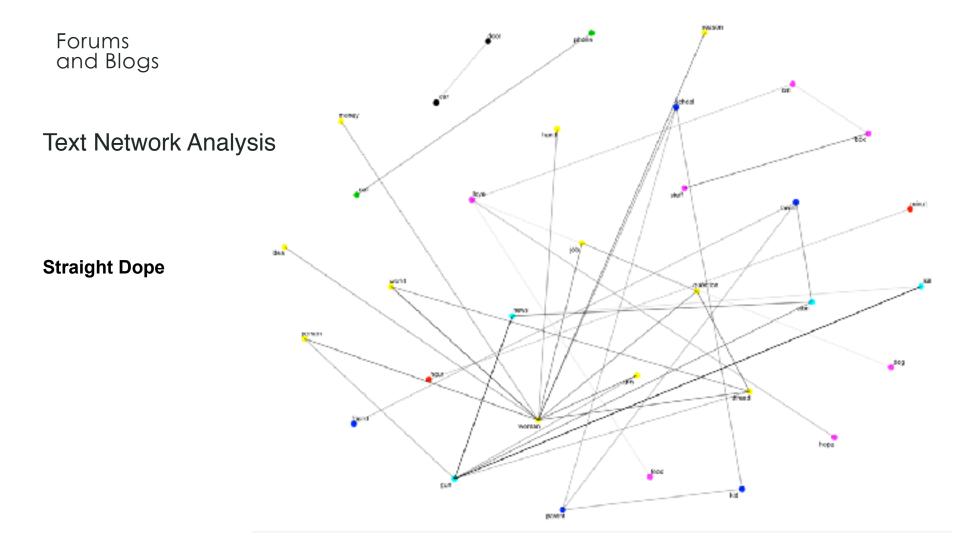
Depression Army



Text Network Analysis

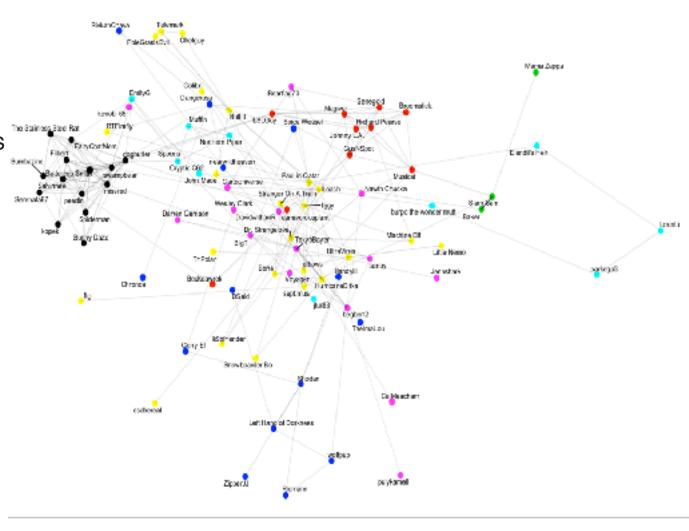
Depression Army





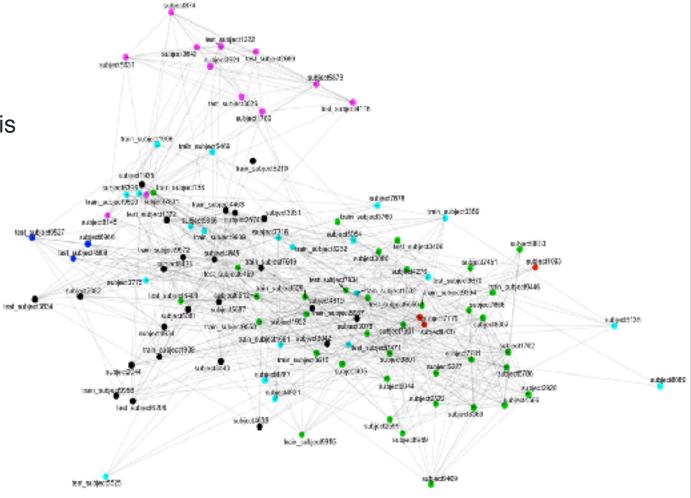
Text Network Analysis

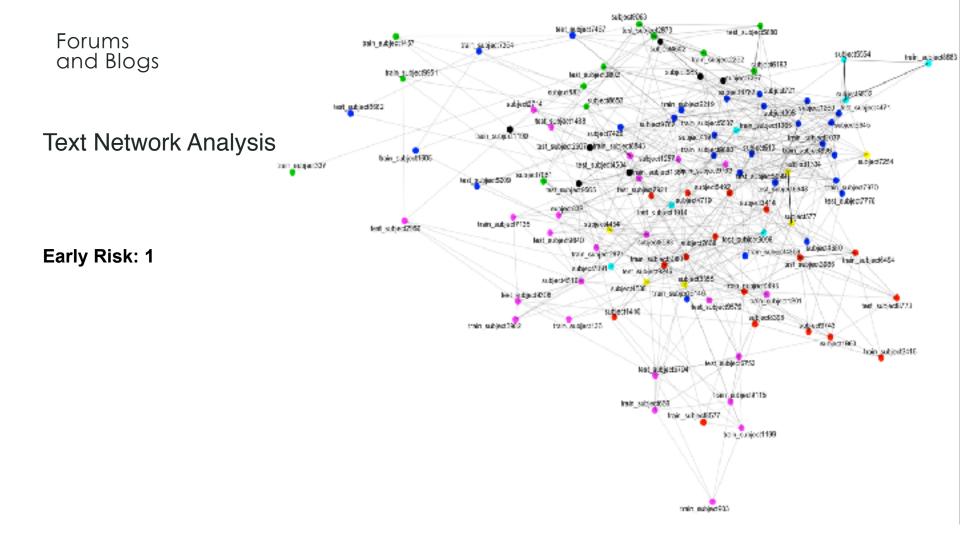
Straight Dope



Text Network Analysis

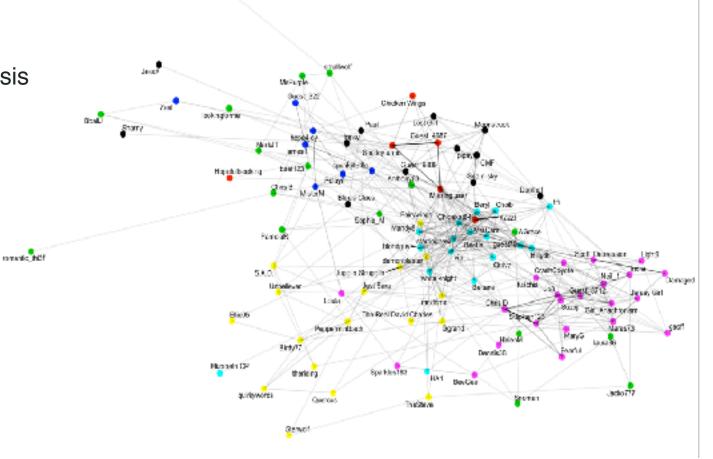
Early Risk: 0





Text Network Analysis

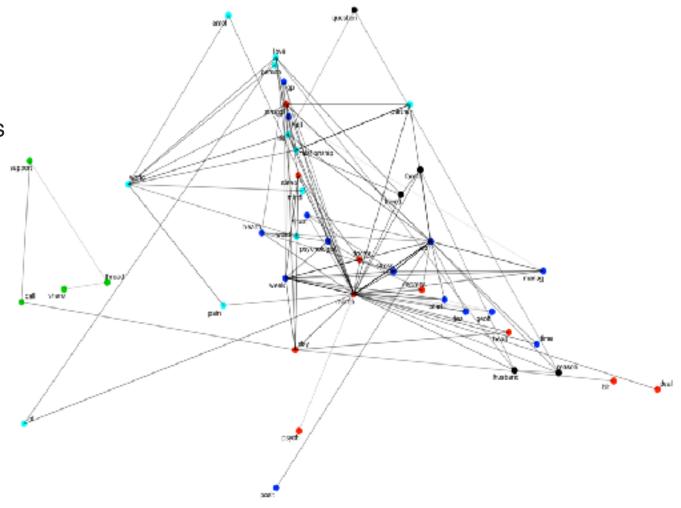
Beyond Blue



Feeling Lone V.

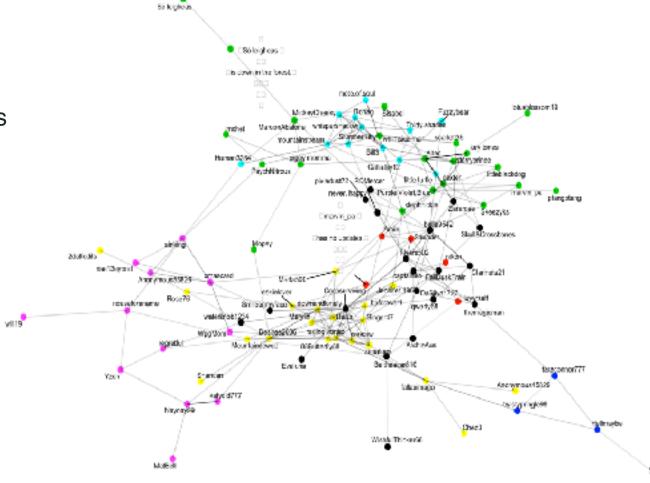
Text Network Analysis

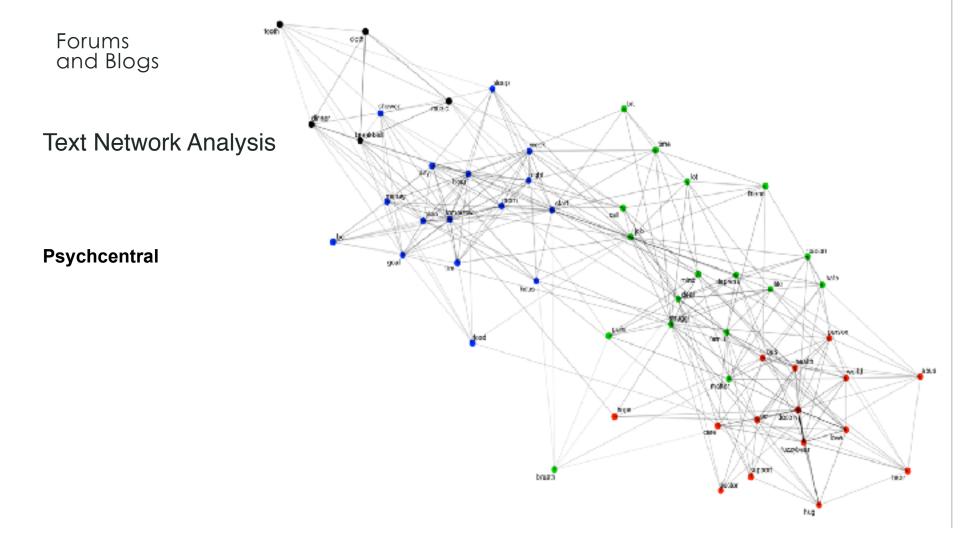
Beyond Blue



Text Network Analysis

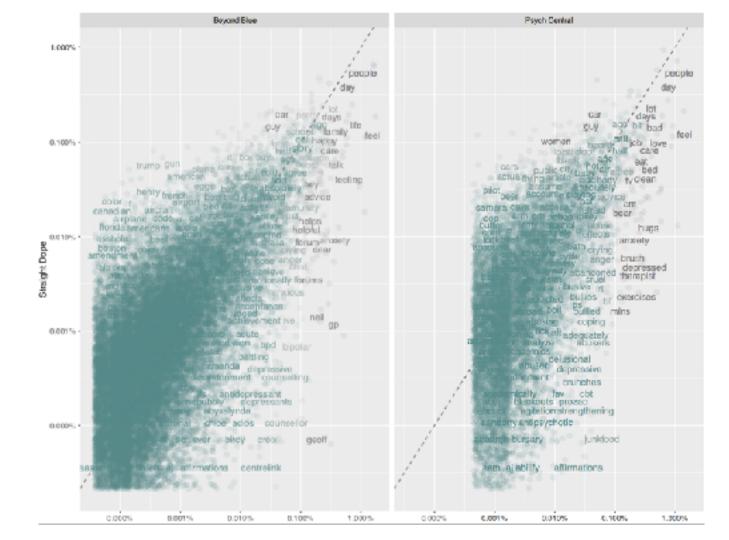
Psychcentral





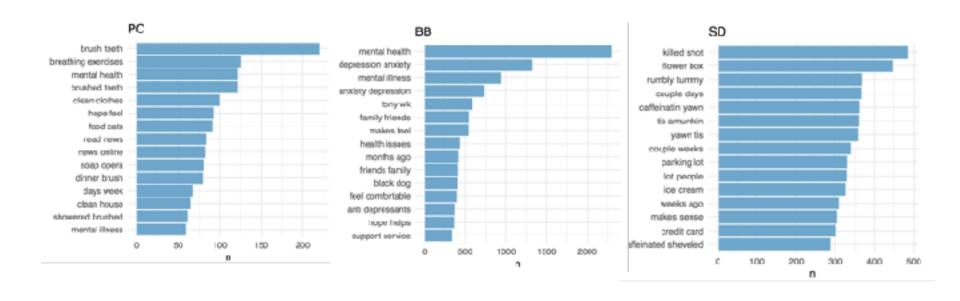
Forums

SHARED WORDS



Forums

FREQUENT BIGRAMS

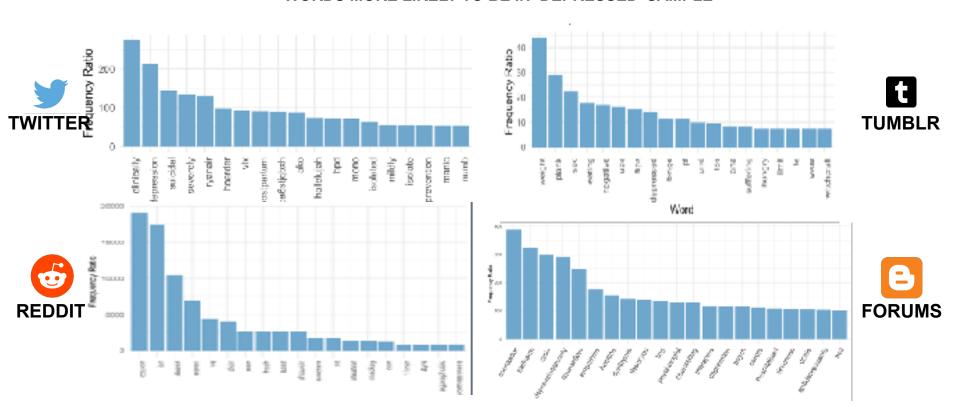


Platform Comparisons

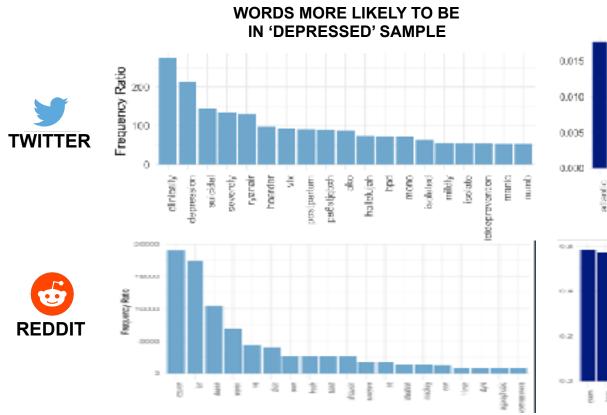
Dictionary-Based Analysis: Relative Frequencies

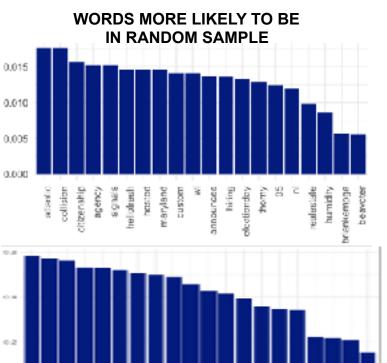
RATIO: DEPRESSED / RANDOM

WORDS MORE LIKELY TO BE IN 'DEPRESSED' SAMPLE



RATIO: DEPRESSED / RANDOM





RATIO: DEPRESSED / RANDOM

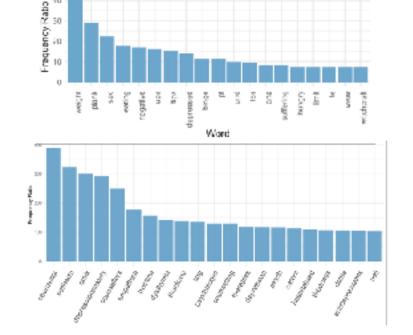
Relative Frequencies

Ratio: depressed / random

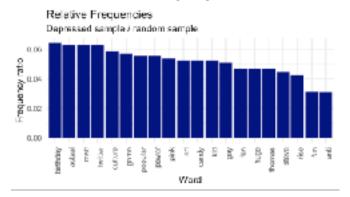
WORDS MORE LIKELY TO BE IN 'DEPRESSED' SAMPLE

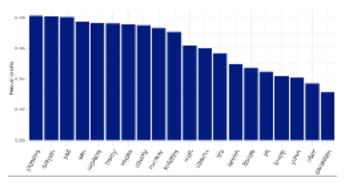


FORUMS

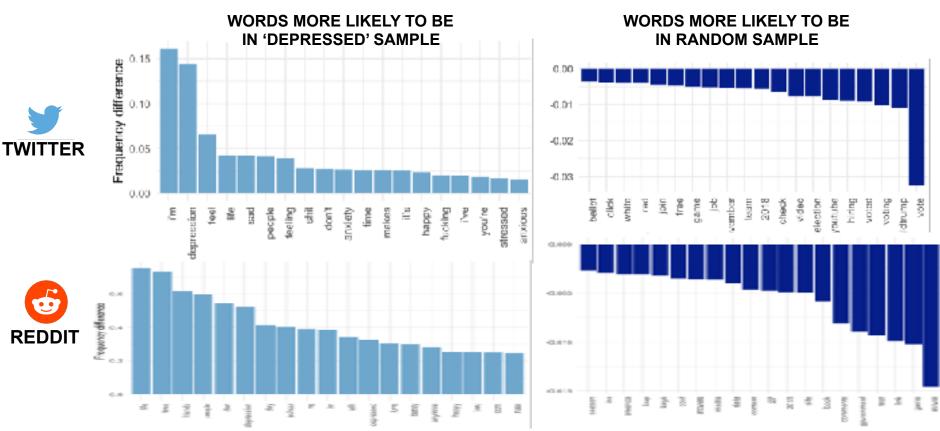


WORDS MORE LIKELY TO BE IN RANDOM SAMPLE





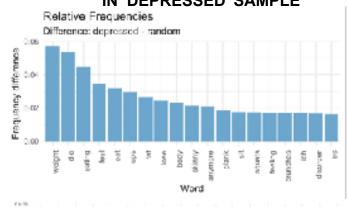
DIFFERENCE: DEPRESSED - RANDOM



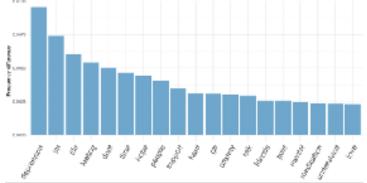
DIFFERENCE: DEPRESSED - RANDOM

WORDS MORE LIKELY TO BE IN 'DEPRESSED' SAMPLE

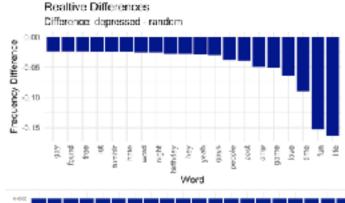


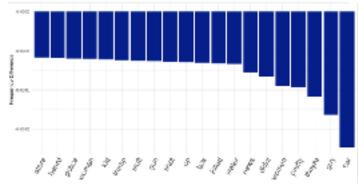






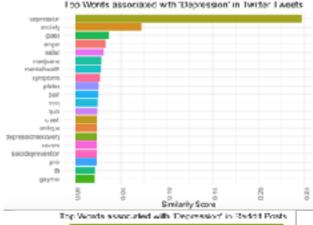
WORDS MORE LIKELY TO BE IN RANDOM SAMPLE



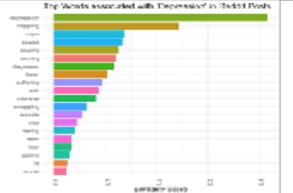


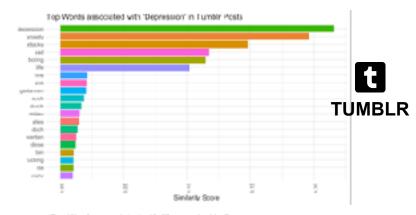
SIMILARITY: DEPRESSION

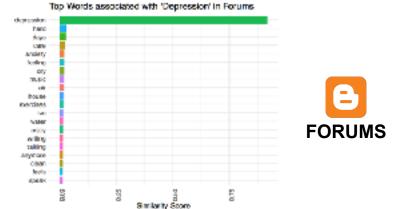






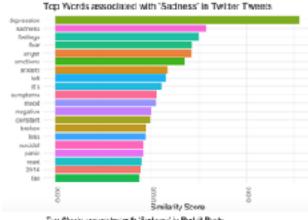


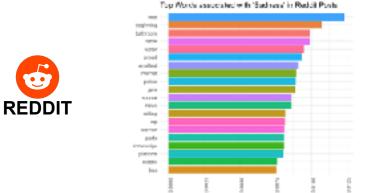




SIMILARITY: SADNESS



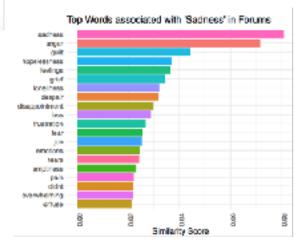




In tumblr post, there is no sadness after stemming...

Does it necessary to include sad & sadness at the same time?

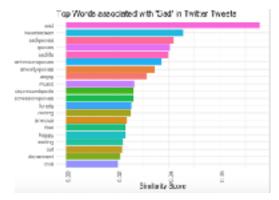


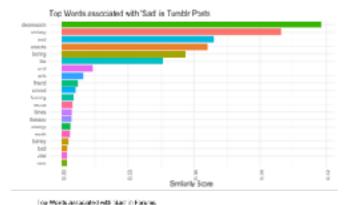




SIMILARITY: SAD

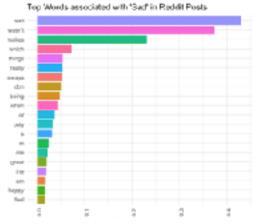


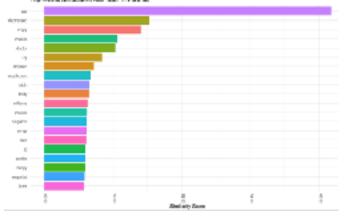






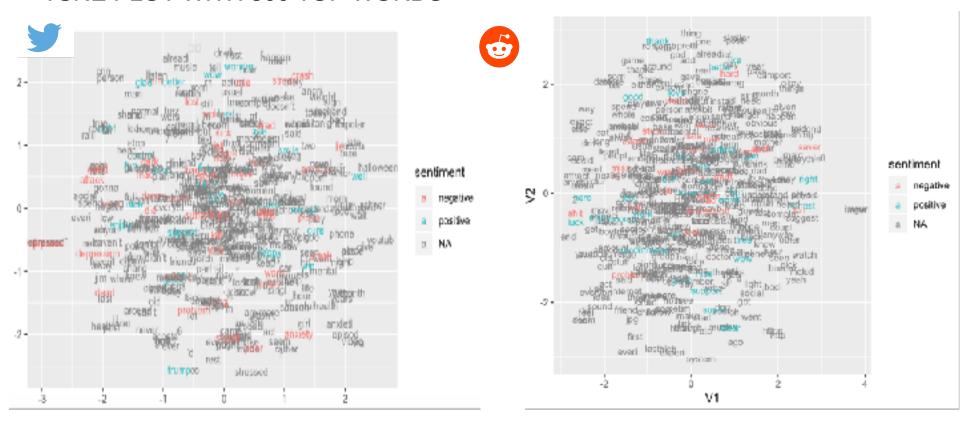




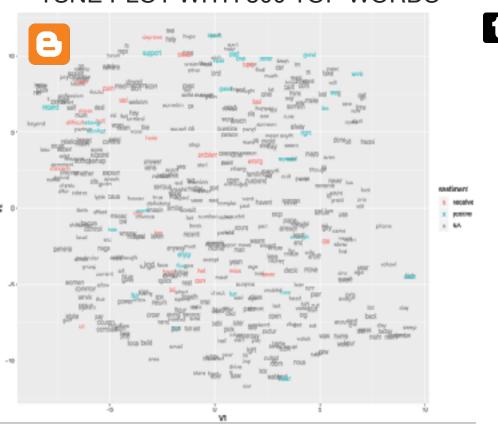


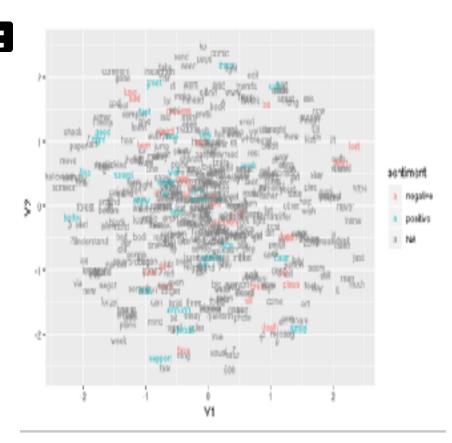


TSNE PLOT WITH 500 TOP WORDS



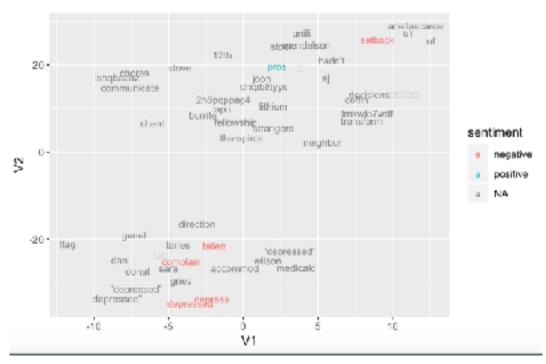
TSNE PLOT WITH 500 TOP WORDS



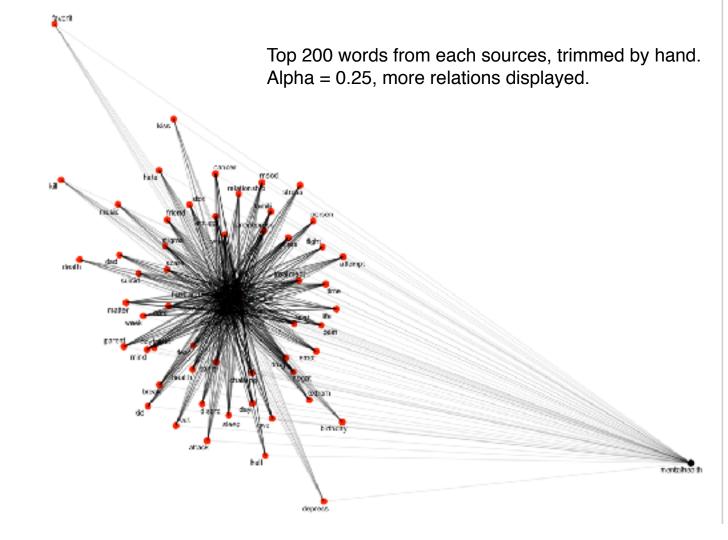


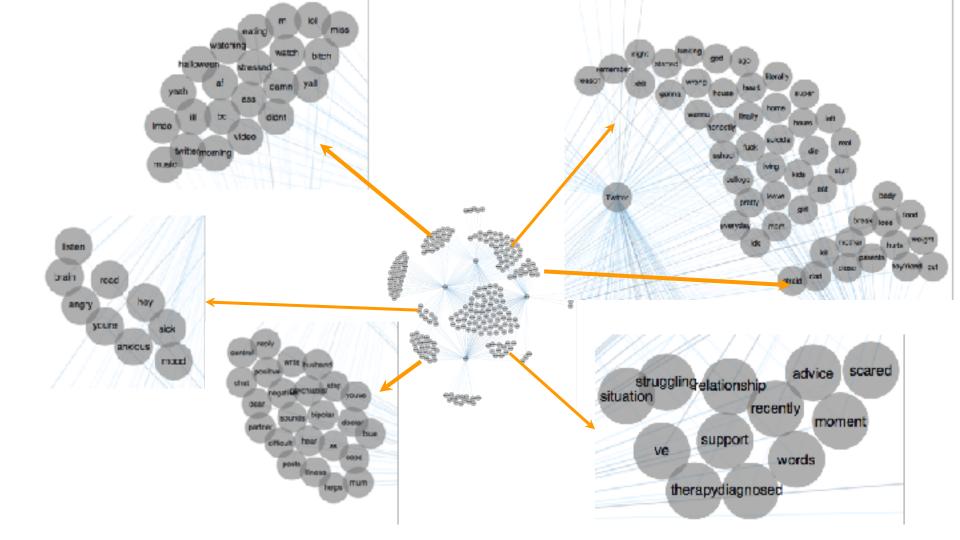
TSNE PLOT WITH 50 WORDS SIMILAR TO DEPRESS





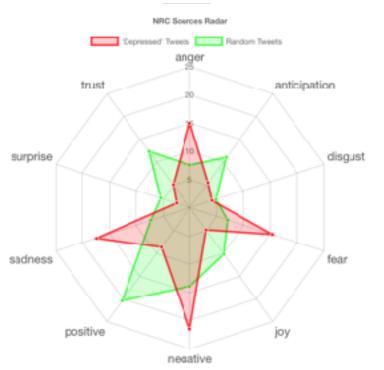
Combined Network: Words



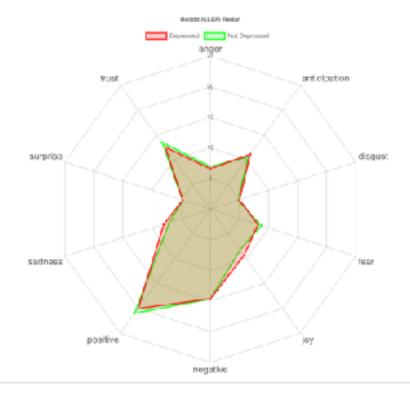


NCR Radar





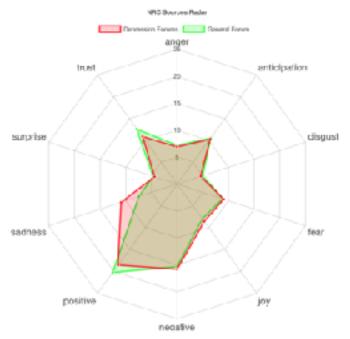




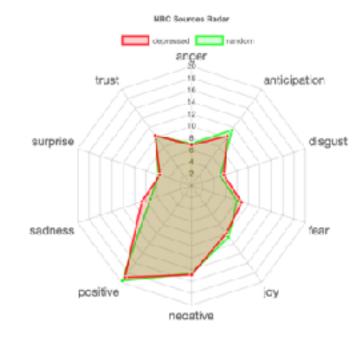
NCR Radar



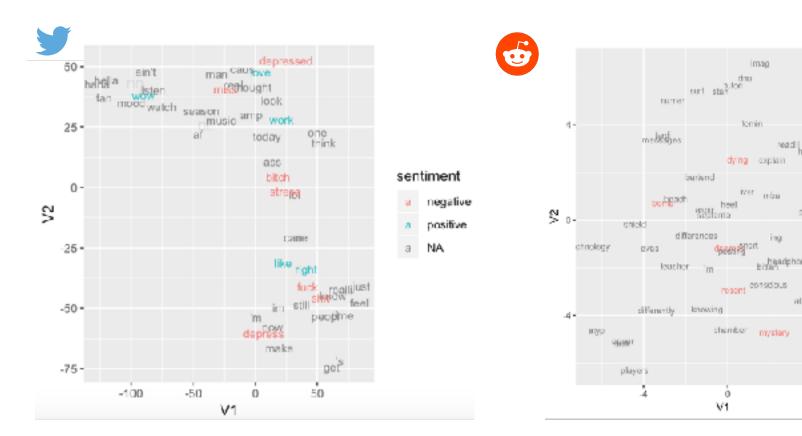
FORUMS



TUMBLR



TSNE PLOT WITH 50 WORDS SIMILAR TO DEPRESS



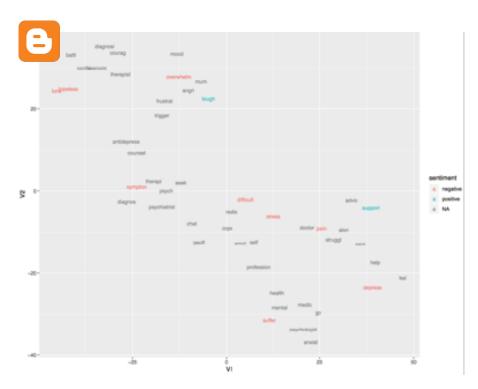
featur

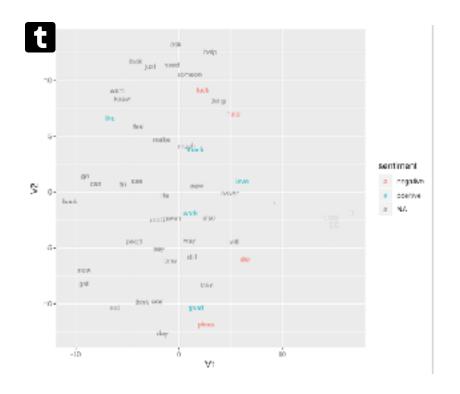
attienton

sentiment

a negative

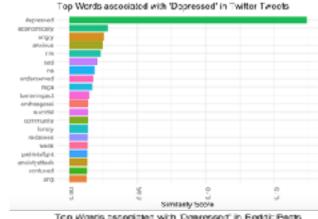
TSNE PLOT WITH 50 WORDS SIMILAR TO DEPRESS



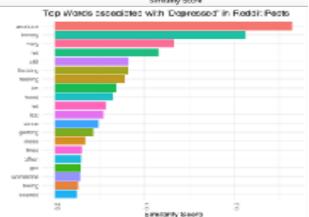


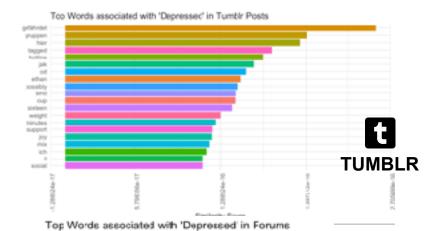
SIMILARITY: DEPRESSED

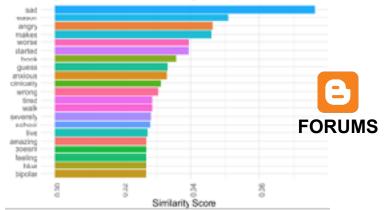






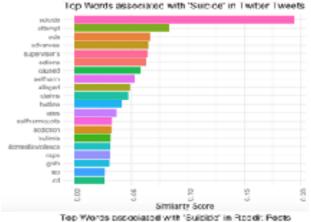




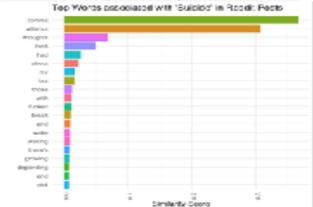


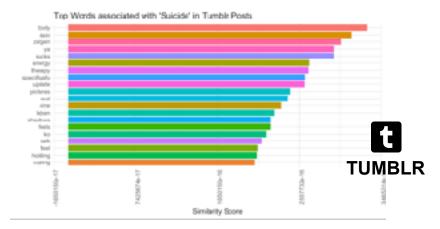
SIMILARITY: SUICIDE

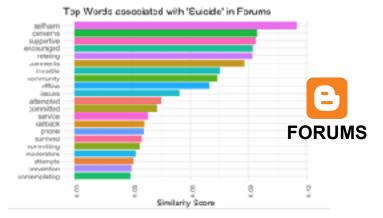








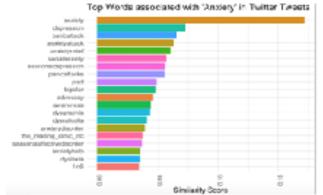


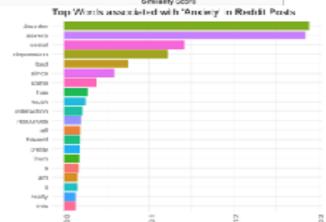


SIMILARITY: ANXIETY

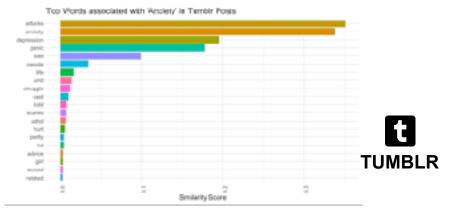


REDDIT





Similarity Score



Similarity Score

FORUMS

