Characterizing and Predicting Long-term Emotion in Reddit Comments, CS224W

Jaron Shalom Rottman-Yang / jaronry / 05817302

December 10, 2017

1 Introduction

As people spend more and more time online [1], it becomes increasingly important to know the effects such behavior has. It is already known that emotions such as happiness can be transferred through real-world social networks, and that this spread can happen broadly across the network [2]; a major question then arises, which is whether online communities can have effects on individual mental state and health. Numerous studies have demonstrated that spending large amounts of time on the Internet can have negative effects. One found that online interactions have both positive and negative effects on user well-being, with negative effects on depression, academic performance, general emotional health, stress, and loneliness.[3] Another found that exposure to uncivil comments led to increased hostile cognition; however, the effect was not found to scale with the proportion of incivility.[4] This may have been because of the maximum level of incivility or the small number of incivil comments presented.

It is clear that online activity cannot be considered to have no influence on real-world mental state. Knowing that emotions can be transferred through social networks, and that online interactions can have effects on well-being, this project will consider whether emotions can be transferred to an individual through online networks, how emotions may spread to an individual in online networks, and what network characteristics might influence an individual's susceptibility.

This paper will present one attempt at measuring emotional contagion in Reddit, an online network which has not yet been subject to an analysis such as those done for Facebook[5] or Twitter.[6]

2 Related Work

2.1 The Spread of Behavior in an Online Social Network Experiment [7]

This paper details an experiment in which the adoption of a health behavior in a social network was characterized. The authors found that the health behavior spread more widely in a clustered network, with an average of 53.77% of users adopting the behavior versus 38.26% of users adopting the behavior in a random small-world network. In addition, the authors found that redundant ties were very important; a user receiving many email notifications was significantly more likely to adopt the behavior than a user receiving fewer notifications. Such social reinforcement also had a significant effect on the users' level of engagement with the behavior. If emotional contagion can spread similarly to behavior, then this paper suggests a more clustered network will allow for stronger emotional contagion.

2.2 Sentiment Flow Through Hyperlink Networks [8]

This paper investigates the flow of sentiment through cascades in blog networks. It uses the MemeTracker dataset with millions of blog posts and hyperlinks between blog posts. The network used for the analysis of sentiment flow assigned blog posts to nodes and inserted an edge from one node to another if the first node cites the second node in the post. Sentiment was measured on two axes (positivity/negativity and objectivity/subjectivity) relative to baseline sentiment for the author, and was based on bag-of-words averages as well as emoticon tagging. It provides a different look at sentiment flow from the approach in this paper, following individual flows rather than the cumulative effect of sentiment flow.

2.3 Experimental evidence of massive-scale emotional contagion through social networks [5]

This paper details a study in which Facebook users' News Feeds were altered and the users' resulting activity was evaluated. The study was performed in order to determine whether emotions spread like a contagion through online social networks. News Feeds were altered by reducing the number of positive or negative posts that appeared, where positivity and negativity were determined by the presence of 'positive' and 'negative' words. The authors found that reducing negative emotional content on a user's News Feed led to that user posting fewer positive words and more negative words as a percentage of their total output.

From these results, the authors concluded that emotional contagion does occur via text-based computer-mediated communication. It is worth noting that the amount by which a user's emotional expression is affected is very low, with the percentage of positive or negative words in a user's output changing on the order of 0.1%. This forms a basis on which to base the ideas in this paper; emotional contagion in social media may be generalizable to a different type of network.

3 Methods

3.1 Data Set

For our dataset we used Reddit data provided by the SNAP group, consisting of all comments posted on the politics subreddit in 2014. Reddit data is appropriate to the task of evaluating the spread of emotion over time through online material due to its nature as a large forum that links all comments written by one person to one account. The politics subreddit 2014 dataset is presented as a list of comments, with each entry noting important metadata such as the poster's account, the parent comment, the timestamp, and the score, as well as the comment text. The comment text is pre-processed to replace sentence-level punctuation with 'EOS' words, and contains other noise (such as contractions being separated into two words - don't into do n't).

It is difficult to tell solely from the comment data what comments a user may have seen in the course of browsing the website. Therefore, for this project, the focus was on direct user-to-user interactions within threads. The year of comments was divided into four quarters (Jan. to March, April to June, July to Sept., Oct. to Dec.), and a separate directed direct-reply multigraph was created for each quarter. Within the networks, nodes represent users, and an edge between any two nodes A and B is added if A responds directly to or is directly responded to by B. A score is assigned to each edge according to the content of the comment; if A comments on a comment by

B with comment of score 10, then the $A \to B$ edge has a score of 10 (and we say A has produced a comment with emotional score 10, while B has received a comment with emotional score 10). Each of the four quarters is evaluated separately for network characterization.

For the predictive models, a single dataset for the full year is used, with samples consisting of one user over two consecutive quarters. Only users who commented at least 5 times in each of the two consecutive quarters are considered. User features from the first quarter are used to predict user features from the second quarter. So, a user who comments 10 times in each quarter will appear as three samples: one for the first and second quarters, one for the second and third quarters, and one for the third and fourth quarters. This allows for the inclusion of users who joined or left the Reddit userbase during 2014.

3.2 Sentiment and Emotional Tone Scores

Two methods were chosen to represent comment scores in the networks. Both methods attempted to capture the emotional value of the comment.

In the first method, the TextBlob sentiment analyzer is used.[10] On each comment, the TextBlob analyzer computes a polarity score between -1 and 1 corresponding to the comment sentiment and a subjectivity score between 0 and 1 corresponding to the comment subjectivity. The score of edge e is then computed by

$$S_{TB}(e) = 100 p_e s_e$$

where p_e is the polarity of edge e and s_e is the subjectivity of edge e. This score maintains the sign of the comment polarity but scales the score by the comment subjectivity (under the assumption that more subjective comments are more emotional). Scores range from -100 to 100; a score of -100 indicates a highly emotionally negative comment, while a score of 100 indicates a highly emotionally positive comment.

The second method relies on LIWC dictionaries. The LIWC text analysis tool provides an emotional tone metric for a given block of text.[11] The metric falls between 0 and 100, with a value of 0 indicating a highly negative emotional tone (greater anxiety, sadness, or hostility), and a value of 100 indicating a highly positive emotional tone (more positive, upbeat). For this project, the score of edge e with emotional tone t_e is given by

$$S_{ET}(e) = t_e - 50$$

in order to center emotional tone at 0. Scores range from -50 to 50, with a score of 0 suggesting ambivalence or lack of emotionality.

3.3 Network Characterization

To characterize the Reddit direct-reply network, we computed network statistics using SNAP for each quarter network, using both TextBlob sentiment and LIWC emotional tone scores. In addition, we produced a number of plots to better understand the nature of emotional expression within the networks.

3.4 Prediction Models

We developed several different predictive models to test the feasibility of determining the future emotional expression of a user from their current interactions in the network. Target classes were

defined in terms of the change in average produced emotional score from one quarter to the next; to simplify the problem we attempted binary classification (increased average emotional score or decreased average emotional score). Feature vectors for each sample included the degree, total received emotional score, and average received emotional score. The small feature vector allows us to treat any success of the predictive task as indicative of emotional contagion.

We used the following models as classifiers:

- K-Nearest Neighbors: Examines the 3 nearest neighbors by Euclidean distance to the query sample, and returns the majority label.
- Logistic Regression: Attempts to fit a function $P(x) = \frac{1}{1+e^{-\theta x}}$.
- Naive Bayes: Constructs a probability model and chooses the most likely label based on that model.
- Multilayer Perceptron: Uses a one-layer 100-neuron artificial neural network with ReLU activation.
- Random Forest: A collection of 10 decision trees trained on subsets of the data.
- Support Vector Classifier: Uses the Gaussian kernel to map samples to higher-dimensional space.

All models were developed in Python with scikit-learn. Classes were already well-balanced so there was no need to balance. Models were evaluated with 5-fold cross validation.

4 Results and Discussion

4.1 Network Statistics

We calculated the clustering coefficient and average scores for each of the four quarter networks, seen in Table 1.

Qtr	No. of Nodes	No. of Edges	Avg Clust. Coef.	Avg Score (TB)	Avg Score (ET)
Qtr 0	68672	813180	0.0021	2.9053	-9.6355
Qtr 1	66562	833192	0.0023	2.9082	-10.1221
Qtr 2	66380	799438	0.0016	2.7850	-10.9747
Qtr 3	71180	880192	0.0018	2.6501	-11.05808

Table 1. Statistics for each of the four quarter networks, for both TextBlob sentiment scores (TB) and LIWC emotional tone scores (ET).

These statistics show that there is not a significant difference between the four quarter networks. For all networks, the clustering coefficient is very small, a sign that the network does not have strong local community structure. The difference in sign between the average score using the TextBlob score versus the LIWC score is notable as well.

4.2 Network Plots

We produced various plots to clarify the network structure. Figure 1 shows the degree distributions for the TextBlob-scored networks as well as the LIWC-scored networks.

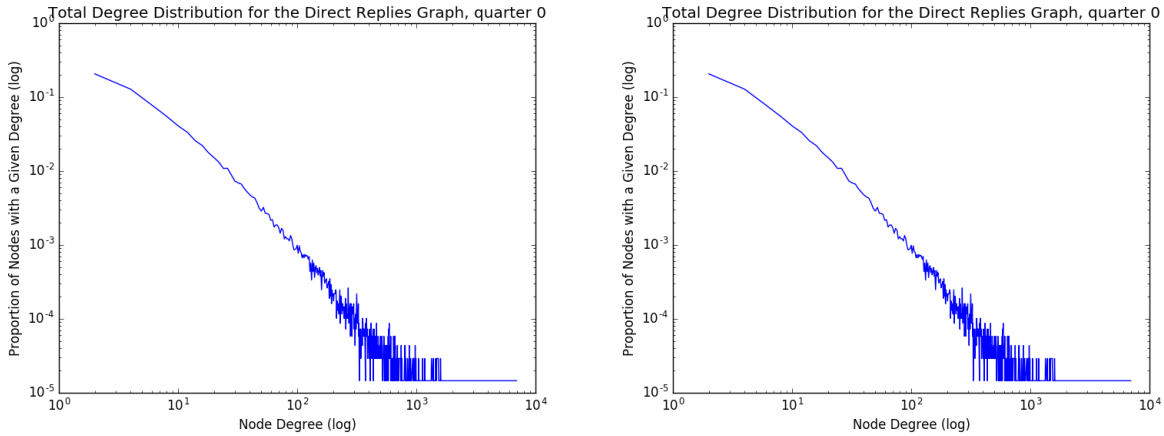


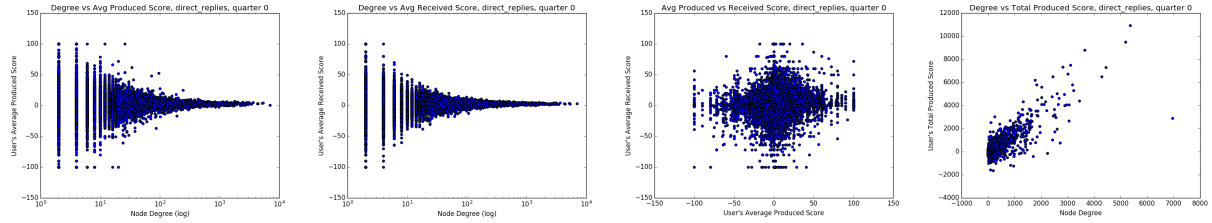
Figure 1. Degree distribution for the TB scores network on the left and the ET scores network on the right, for the first quarter, with log-log axes. Other quarters have similar appearances.

Upon viewing the degree distribution, it is apparent that the Reddit direct-reply networks follow a power-law distribution. There are a large number of users who comment very infrequently, and a select group of 'power users' who comment far more frequently. This also suggests that if emotion can flow through the comment network, those 'power users' will have enormous influence over the flow of emotion to the rest of the network.

Figure 2 shows various node properties plotted against each other. From the left-most plots (node degree vs average produced score), it is clear that high-frequency posters tend toward overall emotional neutrality. The shape of the plots also suggest that the more extreme average produced scores of lower-degree nodes may just be due to variance; with fewer comments to contribute to the average, one highly emotional comment can heavily skew the average produced score. The second-to-left-most plots show that a similar pattern emerges with average received score.

The second-to-right-most plots investigate the relationship between average produced score and average received score; do more emotionally positive users invite more emotionally positive replies (or are emotionally negative users spurred to reply to other emotionally negative users)? The plots suggest that there is no concrete relationship between the two properties.

The right-most plots show the effect of the differing sign of the average TextBlob and average LIWC scores. With a positive average score, the TextBlob-scored network has total produced score increase with degree, while the LIWC-scored network has total produced score decrease with degree.



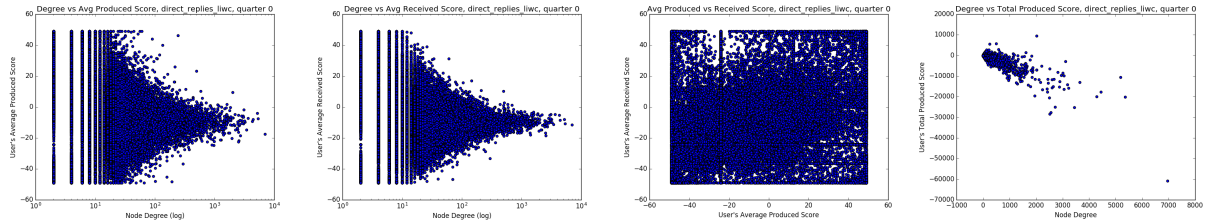


Figure 2. All for quarter 0 networks. From left to right: node degree (log scaled) on x-axis, average produced score on y-axis; node degree(log scaled) on x-axis, average received score on y-axis; average produced score on x-axis, average received score on y-axis; node degree on x-axis, total produced score on y-axis. Top is for TextBlob scores networks; bottom is for LIWC scores networks.

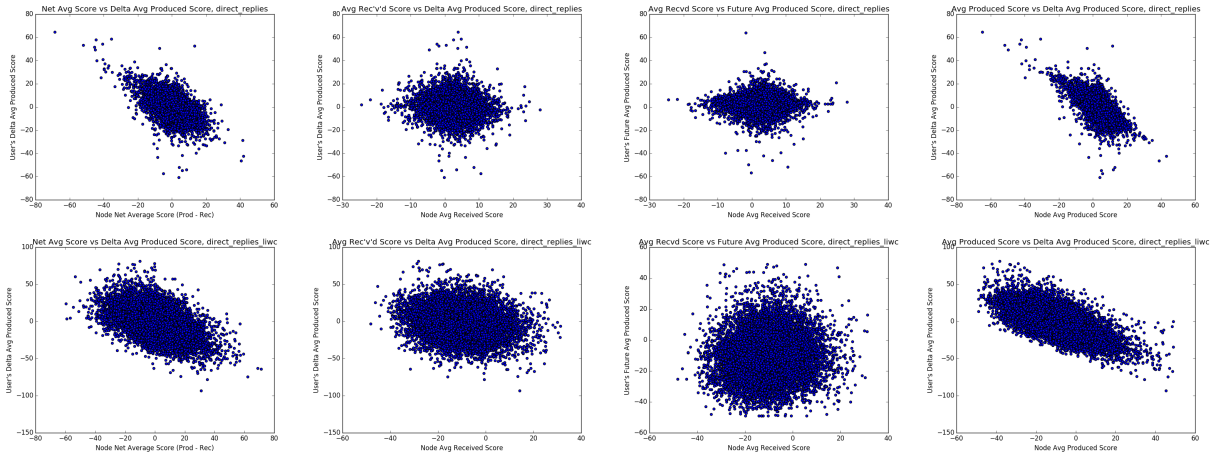


Figure 3. All for the network of samples over two consecutive quarters. From left to right: node net average score (average produced - average received) on x-axis, user's change in produced score from one quarter to the next on y-axis; average received score on x-axis, user's change in produced score from one quarter to the next on y-axis; average received score on x-axis, user's average produced score in the next quarter on y-axis; average produced score on x-axis, user's change in produced score from one quarter to the next on y-axis. Top is for TextBlob scores networks; bottom is for LIWC scores networks.

Figure 3 shows the relationship of various node properties to their future behavior. These plots suggest that there is no strong linear relationship between a user's received score and their change in behavior in the next quarter. In fact, the right-most plot shows that as a user's average produced score increases, they tend to regress back to the mean average produced score in the next month.

In total, these results indicate that there is little emotional contagion information to be gleaned from the emotional scores of comments that a user directly interacts with. It may be that emotion cannot be transferred through large-scale online forums in the same way that it can be transferred through Facebook or Twitter. It may be that emotional contagion in online networks is short-term only. It is also possible that emotional contagion is spread not through active interaction but through passive consumption; in the Facebook and Twitter studies, users' emotions were influenced just by viewing their feeds, not by commenting on the emotional material.

Of course, these results could also be due to the relatively inaccurate tools for emotional analysis used here. Notably, both the TextBlob and LIWC analysis tools are unable to deal with

sarcasm, irony, etc., which are common in online discourse. The noise in the comments could be removing vital signals of emotion, and other contextual information, such as a comment's vote score, that was discarded could also be useful.

4.3 Classifier Results

TextBlob	KNN	Log. Reg.	Naive Bayes	MLP	Random Forest	SVM
Accuracy	0.5037	0.5324	0.5171	0.5223	0.5020	N/A
Precision	0.50	0.53	0.52	0.52	0.50	N/A
Recall	0.50	0.53	0.52	0.52	0.50	N/A
F1	0.50	0.53	0.51	0.52	0.50	N/A
AUROC	0.5012	0.5446	0.5375	0.5250	0.5005	N/A

LIWC	KNN	Log. Reg.	Naive Bayes	MLP	Random Forest	SVM
Accuracy	0.5182	0.5607	0.5604	0.5351	0.5171	0.5066
Precision	0.52	0.56	0.56	0.53	0.52	0.50
Recall	0.52	0.56	0.56	0.54	0.52	0.51
F1	0.52	0.56	0.56	0.52	0.51	0.49
AUROC	0.5220	0.5871	0.5806	0.5450	0.5206	0.4950

Table 2. Results of the various classifiers on both TextBlob (above) and LIWC (below) -scored networks.

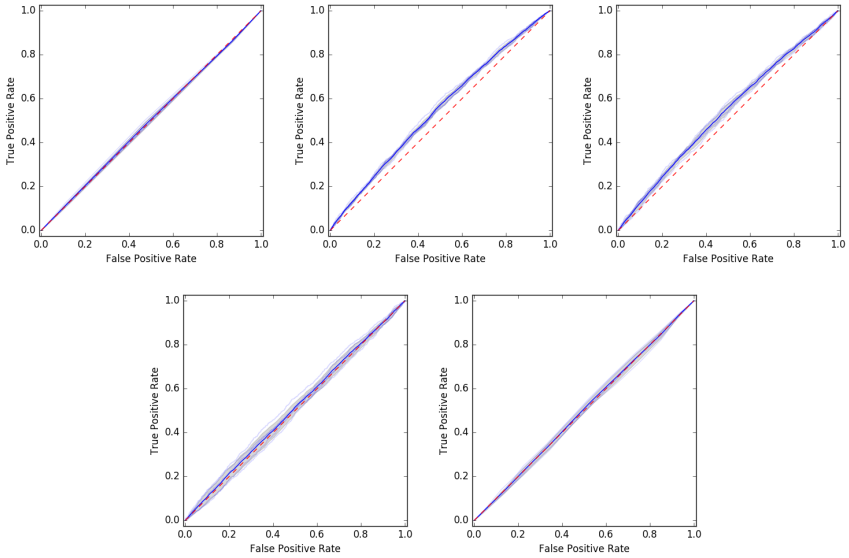


Figure 4. ROC curves for TextBlob-scored networks. Top row, left to right: K-nearest neighbors, logistic regression, naive Bayes. Bottom row, left to right: multilayer perceptron, random forest.

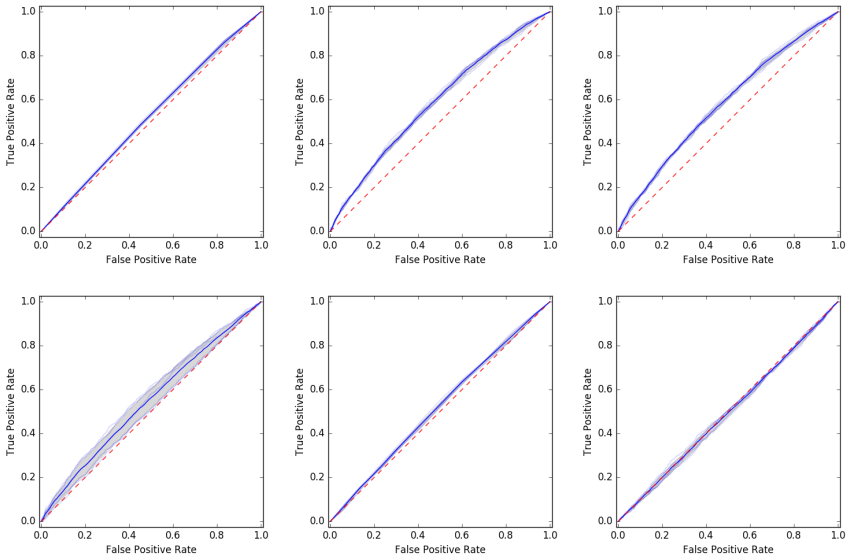


Figure 5. ROC curves for LIWC-scored networks. Top row, left to right: K-nearest neighbors, logistic regression, naive Bayes. Bottom row, left to right: multilayer perceptron, random forest, support vector classifier.

The result for all of these models is poor. None of the models can do significantly better than random in classifying the change in average produced emotional tone given information about past received score. From the earlier plots, it appears that there is no relationship between past received score and change in average produced emotional tone.

5 Conclusions

In this paper, we attempted to generalize previous results about emotional contagion in online social networks to predict long-term emotional contagion from users' direct interactions with Reddit comments. The results show that this is not possible with the conception of the problem presented here; either long-term emotional contagion cannot be observed in Reddit direct-reply networks, or the tools used in this paper to accomplish the task were incorrectly chosen. Future work on this topic should explore other methods of emotional analysis, or should account for other mechanisms for emotional contagion through Reddit.

References

[1] Elizabeth Anderson. *Teenagers spend 27 hours a week online: how internet use has ballooned in the last decade.* The Telegraph. 11 May 2015. Online.

[2] JH Fowler, NA Christakis. *Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study.* BMJ 2008;337:a2338.

[3] LF Pendry, J Salvatore. *Individual and social benefits of online discussion forums.* Computers in Human Behavior, Volume 50, 2015, 211-220.

[4] L Rosner, S Winter, NC Kramer. *Dangerous minds? Effects of uncivil online comments on aggressive cognitions, emotions, and behavior.* Computers in Human Behavior, Volume 58, 2016, 461-470.

[5] Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. *Experimental evidence of massive-scale emotional contagion through social networks.* PNAS 2014 111 (24) 8788-8790

[6] E Ferrara, Z Yang. *Measuring Emotional Contagion in Social Media.* PLoS ONE 2015 10(11): e0142390

[7] D. Centola. *The Spread of Behavior in an Online Social Network Experiment.* Science, 2010.

[8] M. Miller, C. Sathi, D. Wiesenthal, J. Leskovec, C. Potts. *Sentiment Flow Through Hyperlink Networks.* In Proc. ICWSM, 2011.

[9] C Arthur. *Facebook emotion study breached ethical guidelines, researchers say.* The Guardian. 30 June 2014. Online.

[10] Loria, Steven. "Textblob: Simplified text processing." Python release v0.15.0 (2017).

[11] *LIWC2015.* Pennebaker Conglomerates. 2017.