Spring 5-22-2019

Influence Analysis based on Political Twitter Data

Jace Rose
San Jose State University

Influence Analysis based on Political Twitter Data

A Project

Presented To

The Faculty of Department of Computer Science

San José State University

In Partial Fulfilment

Of the Requirements for the Degree

Master of Computer Science

By

Jace Rose

May 2019

SAN JOSÉ STATE UNIVERSITY

The Undersigned Thesis Committee Approves the Thesis Titled

Influence Analysis based on Political Twitter Data

By

Jace Rose

APPROVED FOR THE DEPARTMENT OF COMPUTER SCIENCE

Dr. Teng Moh, Department of Computer Science 05/14/2019

Dr. Melody Moh, Department of Computer Science 05/14/2019

Dr. Thomas Austin, Department of Computer Science 05/14/2019

ABSTRACT

Studies of online behavior often consider how users interact online, their posting behaviors, what they are tweeting about, and how likely they are to follow other people. The problem is there is that no deeper study on the people that a user has interacted with and how these other users affect them. This study examines if it is possible to draw similar sentiment from users with whom the target user has interacted with. The data collection process gathers data from Twitter users posting to popular political hashtags, which the highest at the time published were #MAGA and #TRUMP, as well as the tweets of people to whom they have tweeted. By applying weights based on the type of interactions as well as the amount, study how close the sentiments that the original user expressed are compared to the users they tweeted to. The weighting formula described above will be known as the Inferred Sentiment Score, or ISS for short. This study presents this scheme of gathering data to build user profiles and ISS to determine how similar a user's sentimental expression is to the people they communicate with on Twitter. The main results of this study show that by using the ISS formula that there is a strong correlation of the sentiments expressed on Twitter by a user and the users that they communicate with.

Keywords—Twitter, Sentiment, User relationship, Stanford NLP

ACKNOWLEDMENT

I want to thank my project advisor, Dr. Melody Moh for guiding and working with me not only in her classes but through the entirety of this project which would not have been possible if not for her. I also extend my appreciation to the committee members, Dr. Teng Moh and Dr. Thomas Austin for their teaching and willingness to work with me and give their time to assess this project and study.

I also thank the department for its resources because of which I acquired the knowledge and skillset necessary to see this project to fruition.

I express a great heartfelt gratitude towards my family, friends, and peers for their unconditional support, help, and faith.

Table of Contents

Table of Figures

Table of Formula

Table of Tables

1. Introduction

A lot can be said about users on social media based on their data. Users are posting pictures, what other users are sharing, what their stances are on certain topics, and what they like. Facebook, Instagram, Snapchat, and Twitter are all popular ways for people to channel themselves into cyberspace. Social media presents a plethora of information researchers can leverage to gain insights into people's lives and opinions.

Twitter is a widely used platform for users to post their opinions and is first of its kind. Users write their opinions and create hashtags as a way to form groups of related information. For example, if I post my opinion and put #icecream in it, it will add my tweet to a list of tweets where people have also put #icecream in their tweets. This makes it possible for a user who wants to collectively look at a subject to find other user's opinions on it; all one has to do is search for the hashtag. A hashtag can be thought of as a keyword representing the whole tweet. Moreover, Twitter allows for multiple hashtags per tweet, thereby creating a graph of related information that allows users to branch out and explore relatable content.

This makes Twitter an increasingly usable platform for research studies. The data relationships already exist and an abundance of what users share allow for incredible applications. Researchers use this data for clustering, location statistics, sentiment analysis, and improving algorithms for how and with whom users interact. Researchers strive to keep making the analysis more accurate by creating word recognition and slang translations. More recently, researchers are using this data to help classify users into different categories based on their interests.

The motivation to extend the current study is to analyze if users truly surround themselves with users with similar sentiments towards certain subjects. For this, the basis is the way in which users interact with each other and the ability to determine the sentiment of their tweets. This purpose of this study is to explore and find if a weighted strategy can be employed in ISS to examine if a user's sentiments are converging to or diverging from the sentiments expressed by the users interacting with target user. To achieve this, weights are placed on the amount of interactions, the types of interactions, and the sentiments expressed that are shared among users on Twitter. This study's approach is to collect users from political hashtags and build profiles of their tweets and the users they communicated with in those tweets. Once all the user's profiles are built, all the interactions and the types of interactions these users have with other users are used to calculate individual scores. Then it determines the sentiments of all other users. ISS as proposed by this study will provide a percentage of positive and negative sentiments expressed by users who communicated with the original user. By considering the percentage of positive and negative tweets , they will be compared against the originally collected user's positive and negative sentiment percentages.

2. Background and Related Studies

This next section will go over the related works that were read with relevant information. Related work on Trends, User Behavior, Algorithms, Stanford NLP, and data.

2.1 Trends

Determining trends and how trends started are important for understanding context. Without knowing how a negative or a positive trend is started and continued, it cannot be determined how users' sentiments propagate through other users. According to Shahab Saquib and Rashid Ali's study "Understanding dynamics of trending topics in Twitter", tweets with negative sentiments are promoted by influential users who have a particular interest or who want to instigate a sentiment toward a subject[1]. However, this may not be the case with positive trending topics where users tend to share their feelings. Here, users tend to relate more to a subject[1]. Expanding on the current results, another study has shown that someone with many followers can sway users to their side, given post-factual information about the trend that was running [2]. Over time, given that factual information is provided, users who felt a certain way about a subject are more likely to change their views[2].

Given how information spreads, it is possible that the collection process derives a lot of spam or fake accounts. In the USA, especially around the 2016 election, there was a significant amount of spam and misinformation being placed on Twitter[5]. The amount of this spam varies per topic, but overall, larger the sample sizes of the tweets, the more accurate is spam detection[5]. The more spam a user sees, the more it influences them to feel the intended sentiments towards particular subjects. This is due to the fact that the more of that opinion users see, the more they will accept it as factual, even if not. Spam can have a large effect on users who are bombarded with information. The amount of content a person sees on a subject will engage them more in that subject and help shape that person's opinion.

In a study by Abbasi, Rehman, Lee, Riaz, and Luo, they explored temporal user topics[13]. What they found was that while they cannot determine how a user's interest will change over time, they were able to determine what the user's current interests were based on their tweets[13]. This information is pertinent to back the notion that if a user interacts with another person's tweet in any way, that user has an interest in that subject. It is also important to note that users do not change their information frequently, which means that influential trends get longer with time and information [9]. As a result, the sentiments expressed at the time will truly reflect in the collected data. This is how users behave online and how someone of a larger spherical influence can change users' perception in general which may lead to different sentiments being expressed.

2.2 User Behavior

User behavior is how users act on a social media platform. Platforms use it to provide content relevant to the users. In a paper by Dewi, Yudhoatmojo, and Budi, they explored how rumors spread on social media[3]. They created a directed weighted graph based on retweets, likes, replies, and mentions by followers and users of a confirmed rumor[3]. With this approach, they determined that quote, retweet, and reply meant more than what just a mention did[3]. They do this by using a weighting scheme based on the amount and where the user was in the rumor detection. This means that the more a user interacts with another user, the more influence a user will have over the other user[3]. For instance, the more a user interacts with another user, the more they are likely to influence each other[3]. It is important to understand how rumors spread, and even more importantly for this paper, how much one person's sentiment in a post may affect the sentiment of another person who sees that post.

In addition to how users behave through Twitter, a user's location is tweetable and shareable. With the location of a user in a tweet, it is possible to examine the average opinion of the area and possibly categorize the users of that area into bias groups. For this information, there have been studies that consider location data during disasters[4][10][14]. These studies on disasters provide vital information about location tracking of tweets, given a distressful situation [4][10][14]. In the study by Luca Venturini and Evelina Di Corso, they concluded that about 1 in 100 tweets were directly georeferenced to a particular area [4]. In Another study by Eiji Aramaki, Asako Miura, and Mai Miyabe, they revealed that while users may not post in the distressed area, users retweeted about the distressed area in the larger surrounding area [14]. This may or may not be sufficient for emergency services, but it gives a general location that helps define a demographic, such as the demographic of the users who are influencing others.

Instead of taking into account sentiments, another paper which follows a similar path determined how the emotional state relates and influences other users [16]. In this paper, Kiichi Tago and Qun Jin monitored the emotional state and applied their own algorithm to determine how a user affects other users. The factor to determine was emotional states and not sentiments, it employed a similar approach.

2.3 Algorithms

The important factor of determining how users who post feel about a certain subject is recognizing that there is more than one sentiment that may be expressed in a tweet. This is because there is no one driving force behind a tweet. Ahuja, Gupta, Sharm, Govil, and Venkataraman have completed multi-algorithmic analysis using, Support Vector Machines (SVM), K-Nearest Neighbor, Naïve Bayes, and Decision trees[6]. They have discussed in their approach how they used these algorithms to remove stop words, converted skewed words back to their original forms, and used those keywords pulled to create vectors of the tweets and trained a model to determine the accuracy of this approach. Their modelling determined that SVM was the best means to classify emotions expressed in tweets with a 5-fold validation accuracy of ~90%.

In this study, SVM is implemented and tested as a classifier for sentiment classification of data. This study is considering sentiment as a factor – mainly, positive and negative. A base training set is used for classifying sentiments and tweets are used to see where they fall in the SVM interpretation. As shown in Fig. 1, SVM can be tuned for both positive and negative sides of a plane, which is then used to classify the sentiment of a user's tweet.

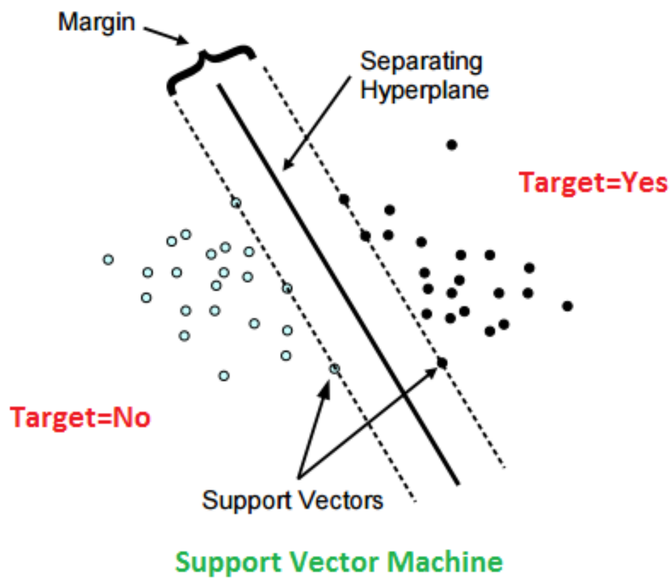

Figure 1. An SVM Implementation of Yes/No

To break down a tweet into meaningful words that can be used for a score in SVM, TF-IDF and bag-of-words are used to supplement SVM. TF-IDF has been used to enhance SVM and improve classification [7]. Bag-of-words provides a set of positive and negative words that are used to filter the tweets. This classifies a users' sentiment so it can be classified for their particular sentiment in a tweet. Bag-of-words is also used to only consider the words that express sentiment or that have a sentiment attached to them.

The results of the SVM, TF-IDF, and bag-of-words implementation, however, did not show promising results. To remediate this and to get better results for sentiment analysis, a different approach is chosen to determine sentiment. Stanford University has developed a sentiment analysis library that provided satisfactory sentiment results on the test data sets.

2.4 Stanford NLP

Since SVM, TF-IDF, and bag-of-words did not perform as well in terms of accuracy as Stanford NLP, a myriad of different options were considered. Google and Amazon both have solutions for this but unfortunately, they are expensive due to their data-quantity pricing model that which is not suitable cost-wise for the large amount of data collected in this study. To solve this problem, Stanford's Natural Language Processor (NLP) library which is written in Java is used. Stanford's NLP uses a combination of binarized tree with a Recurrent Neural Network(RNN). The binarized tree allows for sentences to be broken down and analysis of different parts of speech to be identified. Once a sentence or phrase has been identified, an RNN is used to help select a sentiment for that part of the sentence as seen in the binarized tree in Fig. 2.

In Fig. 2, the tree form of the Stanford NLP library breaks down one sentence into subjects and then multiple adjectives to describe each subject. When there is more than one subject or break in a sentence, it will analyze them and pass back a map of sentiments. The values passed back from Stanford's NLP library are "very positive", "positive", "neutral", "negative", and "very negative".

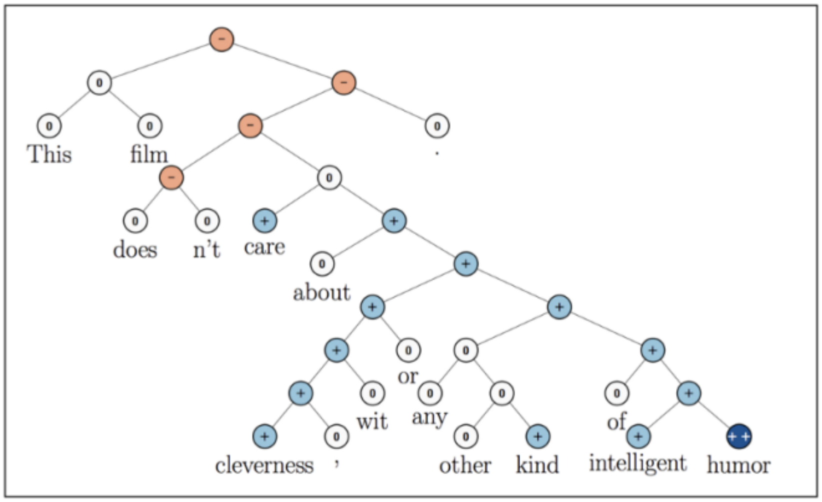

Figure 2. Binarized Tree of Sentence

There is often more than one sentiment passed back for one tweet, as there can be multiple thoughts expressed in a tweet. Only "positive" and "negative" will be used for the tweets in this study. Hence, the sentiments need to be recategorized. Once the sentiments have been recategorized for a tweet, each sentiment Stanford NLP outputs for that tweet is given a numeric value: -1 for negative, 1 for positive, and 0 for neutral. For the data that was collected, the number of negative tweets is far more than positive tweets. Because most of the tweets are negative and being recategorized into positive and negative, neutral tweets are given a bias towards negative and value changed to -1. To give the positive side a bias of its own, a tie breaker between negative and positive is changed to positive. The reasoning behind this is that if the output is neutral and changed to negative, then a tweet that contains two positives, a neutral, and a negative, will be in limbo between positive and negative and not know which side to choose from. With this approach if the situation above occurs, it is clear that it is likely a positive tweet

2.5 Data

There is a multitude of issues with users' speech that tend to limit how researchers classify an emotion towards a particular issue. A major one is acronyms and emoticons. To rectify this issue, users have compiled large dictionaries to translate these acronyms and emoticons into their English equivalent [10][12]. This means when a user types "lmfao" or when a user puts an emoticon such as ":-)" or ":(", they can be factored into the overall sentiment of the sentence. This is an issue with Twitter as users originally have a character limit of 140 characters, now 280 characters. Often, abbreviations are used in their posts to share their opinions. There is no definitive way of incorporating this into Stanford's NLP library that will make a difference in the overall sentence sentiment determination, these pieces of text are disregarded as a result.

Another issue with textual analysis is that users repeat letters in a word to emphasize a feeling. When a user tweets something like "hellllllllllllo", a translator would need to shorten it to the actual word "hello", so a dictionary can properly categorize it in the proposed algorithm[12]. This factor is not covered in this paper but is considered as something worth noting is when a user types a word like this, it can only be said that the sentiment is strong whether its happiness, sadness, or sarcasm.

3. Proposed Solution

Based on the discussion of the related work on user behaviors, sentiments, and interactions, this study proposes an expanded methodology for understanding users and how they influence others online. To do so, a series of steps are outlined in Fig. 3 to analyze and group users. In the first step, data is collected on users based on popular political hashtags on Twitter, #Trump and #MAGA. This is done to collect users who post more frequently as it is easier to build their profiles. How the data is collected is an important part of this study as well.

This paper defines the term ‘base user’ for the user who originally tweeted on hashtags. Based on this, a ‘leaf user’ is defined as the user who was tagged or mentioned by the base user. This study requires monitoring of not just base users, but also leaf users.

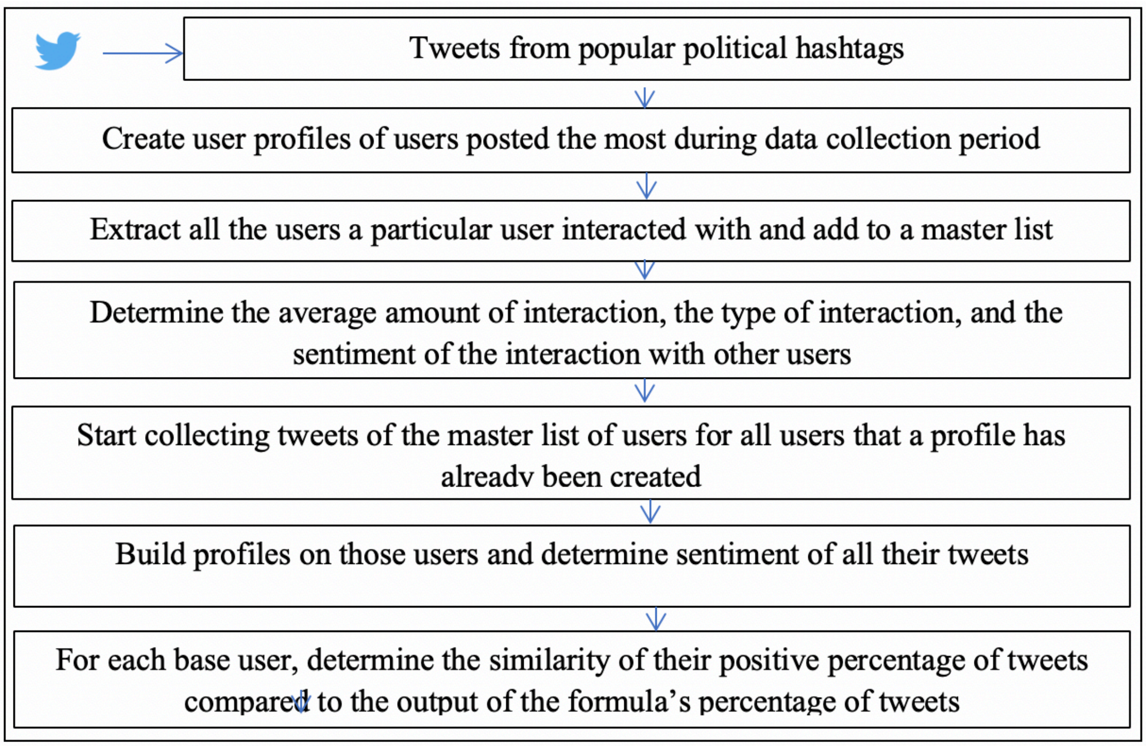

Figure 3. Process to Determine Sentiment Relationship

First, base users and data related to base users is collected and established. Then, data is collected on leaf users. This provides a different amount of leaf users for each base user and a different number of tweets per leaf user. For example, if data collection starts identifying a new leaf user at the end of the collection process, fewer tweets are collected from that new leaf user. However, even if this collection process runs for a long time, there is no guarantee that a leaf user would still post. To help counteract this, leaf user identification is stopped a month before data collection is.

Once data has been collected on the base users and their respective leaf users, the next step is to determine the amount of the original base users have the most collected leaf users. This filters users who have sufficient amount of data to look at their overall sentiment behaviors and their interactions with enough number of leaf users whose data is collectible. If there are not a lot of these leaf users whose tweets are collectible, the base user is removed from this study as there is not enough data to analyze. This step is important in determining who the base users should be because there needs to be satisfactory amount of data to determine similarities between users.

3.1 Identifying Users

The first part of this study is determining base users who comment and interact with other users on Twitter. For this study, base users were required to have commented on other users' posts, liked users' posts, and written their own opinions. An easy method to find a group of users like this is to pull posts on currently trending topics. To get these trending topics, Twitter has a trending list of all the hashtags that are easily accessible, and part of their API allows interested individuals to collect data based on these hashtags. The first way is looking for users who post more on the trending political topics, #MAGA and #TRUMP. The more a user tweets in general and on these topics, the higher chance that another user will interact with that user. Sometimes these users are called “super users”, or rather users who are always communicating their thoughts on what is currently happening in the world. This process will exclude users such as celebrities, public officials, and users of high public stature. Because even if they are a “super user”, public pressure could offset our results as they are not always allowed to speak their true feelings towards certain topics.

3.2 Building User Profiles

Now that a set of base users has been determined, a process to build the profiles needs to be defined. The profile that is built for the user is slightly different from that of the users that user interacts with. The user profile consists of types of communications the user used with other users, the amount that the user tweeted other users, and the sentiment for each tweet that user tweeted.

Gathering interactions and determining the average interactions per user are important first steps for building user profiles. With this, the amount of impact different individuals has on the original user is determined. For example, if a person interacts with another person, Person A, more than some other person, Person B, Person A has more influence and more impact on that person. The number is used in our calculation as a factor to determine how much of the sentiment is imposed on from the original user. Stanford NLP is applied to run on each tweet to determine the overall sentiment of the user. The overall sentiments of the base users’ tweets are the amount of percentage of positive and negative tweets. The percentages are compared against the output of ISS, that is the sum of the leaf users’ weighted sentiments percentages. This shows how similar the base users sentiments expressed are compared to those of the leaf users around them. The type of interactions is considered in the ISS as well. The more a user comments, or retweets, or quotes, the more those actions mean to that user[3]. This is because each user has different preferences on communicating with other users. In other words, if a user interacts with users more in one way than another, then the more popular the interactions would have more weight for the at user. In this case the interactions are quote, tweet, and retweet.

For leaf users, there is no need to collect the same amount of data as base users, only more than one or two tweets. Only a few tweets from these leaf users that the base user interacted with are required. Using Stanford NLP again to classify what the sentiment expressed is in a certain tweet. The base users are mapped to the leaf users that have been collected, which is made possible by keeping a list of users that the user has tweeted in their profile. Now that the base user and the leaf user's profiles are built, the next step is to determine how all this data of their interactions and interaction types, and sentiments can be combined.

3.3 Development of ISS

There are three different parts of the ISS; the amount of weight assigned to an interaction type, the amount of interaction that the leaf user has with the base user, and the sentiments expressed.

In the first part of ISS, the interaction weight factor is the number of interactions the base user has with the other user. This is because without diminishment of the weight for the interaction between the base user and the leaf users, every interaction would have the same impact on the base user. To determine this factor, this paper proposes the diminishment affect in Fig. 4.

$$\frac{AMT_{U-K}}{\sum_{K=1}^{N} AMT_{U-K}}$$

AMT_{U-K} = amount of user K was interacted with in all tweets for a particular user

Figure 4. Diminishment Factor Based on Interactions

This diminishes the influence of user that interacts with another user; the more two users interact, the more influence they have on each other. Only the base users' interactions with the leaf users that are collectable are considered. This makes the overall study and data collection relative to the data that is collected instead of absolute, which is impossible to collect.

For the next part of ISS, the type of interaction is considered. As stated before, there are two different approaches to factor this into the base user's sentiment estimation. One approach is using the values of the type of interactions determined in a study by Fatia Kusuma Dewi, Satrio Baskoro Yudhoatmojo, and Indra BudiI [3]. For this study, the weighting scheme is based on each individual user rather than the overall amount of type of interaction percentages as seen in Fig. 5.

In this scenario, the more the base user commits a certain action, such as a like, or a retweet, or a mention, the more the interaction is worth. This is based on the assumption that the more frequently an interaction is committed compared to others, the more important that action is for the base user. In other words, if a base user only tweets a few times but retweets a lot of other users' tweets, the retweets mean more than tweets when determining their sentiments.

$$\frac{\sum_{i=1}^{tweets_{U-K}} tweet\text{–}\,type_i}{\sum_{i=1}^{tweets_{U-K}} tweet_i}$$

$tweets_{U\text{-}K}$ = all the tweets for a user K

$tweet\text{-}type_i$ = retweet, tweet, or quote

Figure 5. Weight Factor Based on Type of Interaction from Study

The final part to consider is the sentiment. Sentiment has no diminishing factor as it is the percentage of a series of tweets, not the individual tweets. Combining all of these factors is the next and final part of the sentiment analysis of the leaf users. ISS is split into two parts, positive and negative as seen in Equation 1. To arrive at the diminished result, each tweet's sentiment is multiplied by both interaction types and interaction amount with the base user to factor in the fact that different users have different degrees of influence based on those weights.

Considering all the sentiments expressed by the users around the base user, it is considered that a mix of those emotions do exist in the base user. The result of this study has two different outcomes: (1) overall sentiment representation of positive and negative tweets. If a user's positive tweets are over a certain percentage of that user's tweets, then that user is positive, or else they are negative, and (2) the distance between users and their overall sentiments to compare how similar they are in their posting behaviors. This way, there is a clear method to determine if a user is positive or negative as well as a similarity measure to show the level of correlation between base users and their leaf users.

$$\mathrm{ISS}_{P,U} = \sum_{K=1}^{N}\left(\frac{AMT_{U-K}}{\sum_{K=1}^{N} AMT_{U-K}}\right) * \left(\frac{\sum_{i=1}^{tweets_{U-K}} tweet\text{-}type_i}{\sum_{i=1}^{tweets_{U-K}} tweet_i}\right) * Pos\text{-}tweet_{U-K}$$

$$\mathrm{ISS}_{N,U} = \sum_{K=1}^{N}\left(\frac{AMT_{U-K}}{\sum_{K=1}^{N} AMT_{U-K}}\right) * \left(\frac{\sum_{i=1}^{tweets_{U-K}} tweet\text{-}type_i}{\sum_{i=1}^{tweets_{U-K}} tweet_i}\right) * Neg\text{-}tweet_{U-K}$$

$$\mathrm{ISSP}_{P,U} = \frac{ISS_{P,U}}{ISS_{P,U}+ISS_{N,U}}$$

$$\mathrm{ISSP}_{N,U} = \frac{ISS_{N,U}}{ISS_{P,U}+ISS_{N,U}}$$

$\mathrm{ISS}_{P,U}$ =	Positive Inferred Sentiment Score of base user U's leaf users
$\mathrm{ISS}_{N,U}$ =	Negative Inferred Sentiment Score of base user U's leaf users
$\mathrm{ISSP}_{P,U}$ =	Positive Inferred Sentiment Score Percentage of base user U's leaf users
$\mathrm{ISSP}_{N,U}$ =	Negative Inferred Sentiment Score Percentage of base user U's leaf users
Neg-Tweet$_{U-K}$ =	Base user U's leaf user K's Negative tweet determined by Stanford NLP
Pos-Tweet$_{U-K}$ =	Base user U's leaf user K's Positive tweet determined by Stanford NLP
Tweets$_{U-K}$ =	Base user U's leaf user K's total collected tweets

Formula 1. Inferred Relative Sentiment Score, ISS

To provide details of Equation 1, Fig. 6, ISS in Equation 1 is combining all the weights with the sentiments for a particular base user for positive and negative. The following diagram in Fig. 6 gives an overview of what this equation hopes to achieve and shows its intended user.

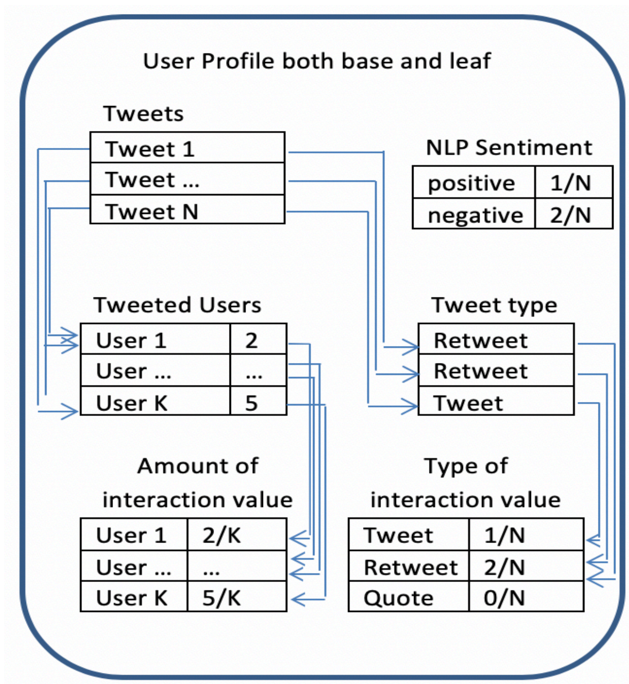

Figure 6. User Profiles are Stored and Calculated

4. Performance Evaluation

This section is going to go over the results from different parts of the study. These different parts include data collection, SVM, Stanford NLP, user interactions, sentiments, user profiles, ISS results.

4.1 Data Collection Results

Table 1. Results of Data Collection

Total Base Users	4,866
Total Leaf Users	8,676
Total users a Base communicated with	101,332
Total users a Leaf communicated with	237,221
Total Base user Tweets	130,518
Total Leaf user Tweets	232,452

Table 1 represents data collected over three months. As shown, the match between the leaf users and the total users that the base user communicated with is not perfect. This is largely in part to profiles being set to private and limitations on collecting the tweets. There is also large discrepancy between the number of base user tweets and the number of leaf users. This can be attributed to duplicated tweets, users whose data was not collected, and multiple tags in the same tweet.

4.2 SVM Performance

To be able to use SVM, the tweets need is turned into a value. To solve this problem, a combination of TF-IDF and bag-of-words is used to break down, weigh the terms frequency, and determine the sentiment of qualifying terms. First, the tweets are stripped off of any words that were not in the bag-of-words that had positive or negative connotations. Then TF-IDF determines the weight of each word from bag-of-words extracted from the tweet and determines its importance. From here, the sum of the TF-IDF of the bag-of-words in a tweet create the label. This same technique for creating a label was tested on the given set of tweets that already had labels. Using 5-fold validation, as seen in Fig. 7., a high accuracy of 70% and a low accuracy of 51% is not satisfactory for this study.

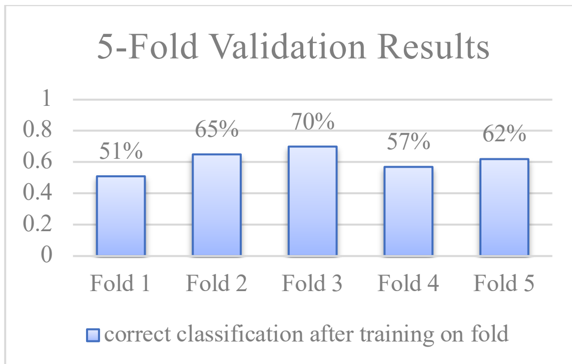

Figure 7. Results of SVM 5-fold Validation

Instead of looking at optimizing this part of the study, a decision was made to look for groups that have already created sufficient ways of labelling sentiment, such as the Stanford Natural Language Processing (NLP) research group.

4.3 Stanford NLP Performance

Since SVM, TF-IDF, and bag-of-words did not produce any usable results for sentiment classification. A new sentiment classifier, Stanford's CoreNLP library is used to classify the tweets with a sentiment. Stanford's CoreNLP tool uses Socher et al's sentiment model with a binarized tree of the sentence to create a Core Mapping to predict the sentiment of that tweet[17]. The annotators from their core annotators ("very negative",

"negative", "neutral", "positive" and "very positive") are determined for the tweet. To test the accuracy of Stanford's CoreNLP library, a test set of 2,980 hand labelled tweets with labels for positive and negative were compared against Stanford's NLP library's outputs.

Table 2. Results of Original Stanford NLP

Tweet Sentiment	Count
Positive	313
Neutral	1,053
Negative	1,614

Table 3. Results of Manually Labelled Tweets

Tweet Sentiment	Count
Positive	1,021
Negative	1,959

In Table 2 and Table 3, there is a mismatch not only in the sentiment labels but the proportions. In the manually classified tweets, there are only positive and negative. The negative tweets consume about 66% of the total tweets and the positive accounts for only 34%. In the Stanford CoreNLP, there is a neutral sentiment classification. To determine the best way to handle the neutral tweets, the overall data sentiment is considered. This study requires positive and negative labels. A bias was introduced from neutral to negative after determining the tweet's context and type. This is done because most the tweets labels as neutral were too short and did not provide full context appropriately or was quoting a game or The Bible. Another issue is if all neutral tweets are treated as negative, there is still a possibility of misclassification of positive and negative tweets by considering the new numbers in Table 4 and Table 5.

Table 4. Stanford NLP Results Given the Negative

Tweet Sentiment	Count
Negative Bias	2,646
Positive	334

Table 5. Reviewed Hand Labelled Tweets Results

Tweet Sentiment	Count
Negative	111
Positive	2,869

After reviewing all the tweets that are hand labelled, some are changed; some from negative to positive and from positive to negative after considering possible biases based on the tweet's contents. Introducing the bias and reviewing the hand labelled tweets for any personal biases, the accuracy of the Stanford NLP output of the test set is 90%. SVM used 5-fold validation to confirm the accuracy of its classification. Stanford's NLP won't be using 5-fold validation as it does not depend on previous data to create a label. Taking that into account, the tweet analyzed by the Stanford NLP will not generate a different label.

4.4 User Interactions Results

In this study, considering the interactions between users is an important factor in how much a user communicates with each other in their tweets.

Table 6. Base User Tweets and User References

Total Unique User references in tweets	101,332
Total Base User tweets	130,518
Average unique Base user interactions	1.29

In Table 6, a large portion of users is being referenced with fewer overall tweets. This is approximately 1.29 user mentions per tweet. When examining overall how many users used these different types in our data, the distribution of tweet types for all collected tweets is 64% retweets, 1% quotes, and 35% tweets. This is informative because it gives a rough idea of how users are tweeting and what the individual communications collected by this study may look like.

4.5 Sentiments Expressed Results

The results of the sentiments expressed in the whole dataset differ from that of the subset Stanford's NLP used. It is important to note again that Stanford's NLP doesn't require training, it obeys a set of rules that cannot be changed on how to break down text. The amount of overall positive and negative tweets.

Table 7. Sentiment amount of All Tweets Collected in This Study

Total Tweets	362,970
Total negative tweets	334,891
Total positive tweets	28,079

In Table 7 that the percentage of positive to negative tweets is askew. This is expected and is acceptable because our training set was a random sample of the overall data collected. When choosing the random dataset, it was possible to get only positive or only negative tweets. Examining the amount of positive and negative tweets in Table 7, it is reasonable that a larger portion was negative.

4.6 Building User Profiles

Each user is given a user profile in this build. The user profile tracks the tweets of that user, a list of the sentiment expressed for each tweet, the amount of interaction per user, the interaction type per user, all the interactions per user, and the final score for the user given. To examine this study's collection of data drilled down to a user perspective, there is the following table, Table 8.

Table 8. User Tweet Statistics

Average tweets per user	26.80
Average negative sentiment tweets expressed per user	24.74
Average positive sentiment tweets expressed per user	2.06
Average retweet per user	14.70
Average quote per user	.07
Average original tweet per user	12.03

In Table 8, it shows the average tweets of each user that has been collected, both base user and leaf user, only tweeted around twenty-six to twenty-seven times on average over the data collection process. This does not mean that they only tweeted that many times in the two to three months data collection period, but that the data sampling for this was only able to capture around twenty-six to twenty-seven tweets on average for this dataset. The percentages of positive to negative are slightly skewed from what was observed in Stanford NLP section; this is because this is not only the test set, but the full dataset with a lot more data being considered.

4.7 ISS Results

There are two different results that will be presented in this study. The first result classifies users as positive if the percentage of positive tweets that were collected reach a certain threshold. The second considers how close a base users' negative and positive tweet sentiment percentages correlate to the output ISS given their leaf users.

The first result shows how effective ISS is in determining the base user's positive sentiment predicted class compared to the outcome of the leaf users predicted class. To do this, different percentages of what defines a user as positive or negative are used. This is done for two different reasons: there is no strict definition that states the percentage of positive or negative tweets that defines a person positive or negative, and the fact that most of the tweets collected in this dataset were negative.

Tables 9 and 10 are defined such that 'Actual' represents the base users and 'Predicted' is the output of the formula this paper presents. There are also percentiles listed under the predicted header that show the threshold at which a user is determined to be positive. For example, the section that read "Positive > 50%" is the threshold for defining users as positive, i.e. if a user's tweets were over 50% positive, then they were defined as a positive user.

Table 9. Results of Base User Positive Percentages Compared to the percent Output from ISS, with Different Percentiles Defining a "Positive" User

	Predicted								
		Positive > 50%		*Positive > 45%*		*Positive >40%*		*Positive >35%*	
Actual		*Class: positive*	*Class: negative*	*Class: positive*	*Class: negative*	*Class: positive*	*Class: negative*	*Class: positive*	*Class: negative*
	Class: positive	0	33	0	36	0	50	2	57
	Class: negative	1	4832	3	4827	7	4809	11	4796
		Positive > 50%		Positive > 45%		Positive >40%		Positive >35%	
Results	Precision:	0		0		0		.15	
	Accuracy:	.99		.99		.99		.98	
	Recall:	0		0		0		.03	
	F-measure:	0		0		0		.05	
	G-measure:	0		0		0		.087	

Table 10. Results of Base User Positive Percentages Compared to the Percent Output from ISS, with Different Percentiles Defining a "Positive" User

		Predicted							
		Positive > 30%		*Positive > 25%*		*Positive > 20%*		*Positive >15%*	
Actual		*Class: positive*	*Class: negative*	*Class: positive*	*Class: negative*	*Class: positive*	*Class: negative*	*Class: positive*	*Class: negative*
	Class: positive	5	99	12	199	43	415	92	543
	Class: negative	15	4747	54	4601	103	4305	242	3989
		Positive > 30%		Positive > 25%		Positive >20%		Positive >15%	
Results	Precision:	.25		.18		.29		.28	
	Accuracy:	.98		.95		.89		.83	
	Recall:	.05		.06		.09		.14	
	F-measure:	.08		.09		.14		.19	
	G-measure:	.11		.10		.16		.20	

Predicting the positive class was difficult as seen in the above tables 9 and 10. Most of the tweets were negative; defining a user as a positive or negative person was challenging. Traditionally, a positive person would be defined as someone who expressed positive sentiments more than 50% of the time in their tweets. It is not until the threshold is lowered to 35% positive tweets that define a user as a positive user that there is some intersect. The more the threshold is lowered for what defines a user as positive, the more f-measure rises.

Table 10 also shows that when the actual class is positive, the predicted class is getting increasingly negative as the threshold is lowered. This shows that the leaf user's sentiments are getting increasingly negative as well. There is similar growth for base users who were determined to be negative, but the surrounding users are generally more positive. Considering the fact that the tweets collected from the start were political, it was expected to have a low positive percentile rate of their tweets for these users. Though as the definition of a positive user changes to lower, the f-measure does increase.

For the second part of the results, this study examines how close users' sentiments are to the users they interact with. This is an important distinction from the first part of the results, which give a rather true or false result, whereas this second part will examine more at how similar or different base users and leaf users are after our formula is applied.

The graph in Fig. 8 describes the users who had a similar positive or negative amount within a given percentage. For example, if a base user tweeted 30% and the threshold was 10%, and if the outcome for the leaf users are below 40% and above 20%, then they are similar. The graph is also given in a percentiles decimal form, to match the y axis. Examining the graph above there is strong evidence to support that base users are pretty similar to their leaf users. There are three fields in this graph: "Within Targeted Range", "More Negative", and "More Positive". "Within Targeted Range" means that the output of ISS outcome percentage fell within the given range of the base user's sentiment percentage. If the output of ISS is more negative than within the target range of the base user, then they are added to the "More Negative" field. The "More Positive" is therefore the leaf user sentiment outcome percentage that were more positive than the base user they have communicated with.

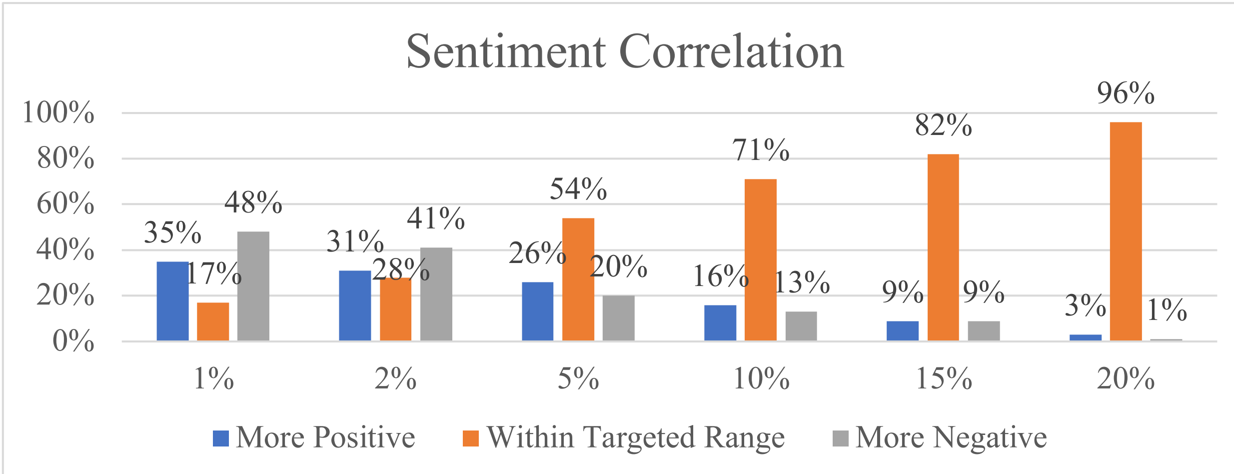

Figure 8. Sentiment Percentage Correlation between Base User and Their Leaf Users using ISS given a Percentage Threshold

At 1%, the lowest correlation between base years and leaf users is observed, at only 17% of leaf users falling within a 1% sentiment outcome range of the base users score. This tells us that while in the first part of the results, our classification of positive and negative may have been ideal with the sentiment distribution of the tweets collected, the base users and the leaf users still tweeted along very similar tweet percentages. Notice as the range expands from 2% to 20%, the number of leaf users who fall within the "Within Targeted Range", or rather the users range increase steadily. A broad statement can then be made about the tweet users collected in this study: while the base users were more negative, the users that they communicated with around them were generally as negative if not more negative. Another important note about the numbers in this graph is that more leaf users are more negative than the base user than more positive.

5. Future Work

There are many paths that this paper lays out for the future with regards to online interactions. It completes a large piece of the larger puzzle of users interacting online. One of the future works that can leverage these results could be determining what sentiment is mapped to certain topics in tweets. Using ISS, if there is a correlation between the sentiment the base users and leaf users output on the topics being tweeted about. Using this, another study can see the sentiment correlation between topics.

Another future work that this study could be used for is expanding out to determine how users feel about the particular subject in the tweet if they can identify the sentiment expressed toward the subject of the tweet. This is a long-term study trending for these users that can track timelines of when negative sentiments are expressed and correlate that to the users around them.

With more data collection and a timeline, another study could see if the base users surrounded by more positive users reach some form of equilibrium between where the base user was to where they are at the new point in time. This paper lays the ground work for the general sentiment expressed by others are related to each other. these are some general areas one might use this study to move forward in those areas.

6. Conclusion

There is definitely a strong link between users and the users that they communicate with online. ISS as presented in this paper shows how close users are similar given different factors. By using ISS, that base users are surrounded by both more negative and more positive people than them. The data also shows that even though these users tend to be more negative, some are still communicating with more overall positive users. Given ISS, considering the factors of how users are communicating, how much they are communicating, and the sentiments they are expressing, this study shows that users share the similar sentimental posting behavior to those they communicate with.

References

[1] Shahab Saquib, Rashid Ali, "Understanding dynamics of trending topics in Twitter", (ICCCA), 5-6 May 2017

[2] Anna Guimarães, Liqiang Wang, Gerhard Weikum, "Us and Them: Adversarial Politics on Twitter", (ICDMW), 2017

[3] Fatia Kusuma Dewi, Satrio Baskoro Yudhoatmojo, Indra BudiI "Identification of opinion leader on rumor spreading in online social network Twitter using edge weighting and centrality measure weighting", (ICDIM), 12-14 Sept. 2017

[4] Luca Venturini, Evelina Di Corso, "Analyzing spatial data from twitter during a disaster", IEEE international Conference on Big Data, 11-14 Dec. 2017

[5] Yifang Wei, Lisa Singh, "Understanding the impact of sampling and noise on detecting events using twitter", IEEE international Conference on Big Data, 11-14 Dec. 2017

[6] Ravinder Ahuja, Rohan Gupta, Saurabh Sharma, Ayush Govil, Karthik Venkataraman, "Twitter based model for emotional state classification" (ISPCC), 21-23 Sept. 2017

[7] Chungho Lee, Incheon Paik, "Stock market analysis from Twitter and news based on streaming big data infrastructure", (iCAST), 8-10 Nov. 2017

[8] Kwan Hui Lim, Shanika Karunasekera, Aaron Harwood, "ClusTop: A clustering-based topic modelling algorithm for twitter using word networks", IEEE international Conference on Big Data, 11-14 Dec. 2017

[9] Jinsei Shima, Mitsuo Yoshida, Kyoji Umemura, “When do users change their profile information on twitter”, IEEE international Conference on Big Data, 11-14 Dec. 2017

[10] Saki Kitaoka, Takashi Hasuike, “Where is safe: Analyzing the relationship between the area and emotion using Twitter data”, (SSCI), 27 Nov.-1 Dec. 2017

[11] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau, “Sentiment Analysis of Twitter Data”, (LSM '11), Pages 30-38, June 23 - 23, 2011

[12] Siddu P. Algur, Rashmi H. Patil “Sentiment analysis by identifying the speaker's polarity in Twitter data”, (ICEECCOT), 15-16 Dec. 2017

[13] Rashid Abbasi, Gohar Rehman, Jia Lee, Farhad Muhammad Riaz, Bin Luo, “Discovering temporal user interest on twitter by using semantic based dynamic interest finding model (TUT)” (ITNEC), 15-17 Dec. 2017

[14] Eiji Aramaki, Asako Miura, Mai Miyabe, “Use trend analysis of twitter after the great east japan earthquake”, (CSCW '12), Pages 175-178, February 11 - 15, 2012

[15] Hajime Watanabe, Mondher Bouazizi, Tomoaki Ohtsuki, “Hate Speech on Twitter A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection”, IEEE Access (Volume: 6), Page(s): 13825 – 13835, Date of Publication: 15 February 2018

[16] Kiichi Tago, Qun Jin, “Influence analysis of emotional behaviors and user relationships based on Twitter data”, Tsinghua Science and Technology, 104 – 113, Feb. 2018

[17] Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60. [pdf] [bib]