

Teenage Development observed via Twitter

Parth Patel

Montclair State University

Upper Montclair, 07043

Patelp94@montclair.edu

Abstract

This paper examines social media's twitter members as they grow and develop from teenagers to early adults. The purpose of this study is to learn and understand how the development of human maturity can also be observed through their social media accounts. The study will tackle to prove this theory by performing three examinations. First, by analyzing linguistic features of the user's twitter tweets. Second, by investigating emotions found among those tweets, and third, by calculating the statistic about the user's tweets. With the results, one can perceive how trails left behind by twitter users can potentially show their progression from teenage to adulthood.

1 Introduction

As we enter an era of new technological advancements from virtual reality to developing artificial intelligence, the mindset of human beings is changing to adapt to this new environment. In the early 2000's when MySpace, and Facebook rose to popularity, many older generations were furious and disappointed to those who were dependent on these time-consuming applications. Because of the addiction that social media and smartphones provided, multiple middle schools and high school strictly banned these devices as they served as a distraction from

obtaining education. While it was clearly a nightmare for those teenagers, the older generations finally came around. In our current time period, the thought of new technology is looked upon rather than quickly judged, and eliminated. Even millions of old-fashioned individuals are browsing through their desired social medias right now. There are billions of people waking up to their smart phones and scrolling down to their personalized pages to find helpful information, communicate across the globe, judge others, and state their opinion. Especially teenagers, who are adjusting to their body's new transformations and changes which brings out many different types of moods. Ample times these moods and attitudes are then shared via social media applications. By observing and studying these changes, one can predict a conclusion about how changes and development in the human body and brain can also be seen on social media. Some teenagers might mature quickly than others which can be evaluated from their social media's statuses and tweets.

In this paper, I will share my results of studying twitter users to observe how the linguistic aspects of tweets develop as the user advances in their age. Some of the topics in this portion of the paper will include spelling development, bad word and acronym usage. The goal is to conclude whether the following topics show an increase or a decrease on the twitter dataset. The second portion of this paper will focus on emotions that are observed from tweets upon the chosen users. This portion of the paper will address emotions in

tweets near events such as Christmas, Thanksgiving, and New Years by performing sentimental analysis. Lastly, a statistical analysis will be completed in which the number of user tweets, re-tweets, image posting, self-reference, and 2nd person references will be calculated. Overall, many would agree that social media provides multiple benefits and its becoming an essential to the modern world. As we humans continue to progress to adulthood, do we leave a trail behind of our development on the social media's we use?

2 Motivation

If you are reading this paper then somewhere in your early life you were taught how to read, and you practiced it multiple times until your brain finally build a structure that lead you to understand the words on this paper today. My point is that as humans, we develop maturity as we age, and it's important to study these changes to fully understand the human-nature. Psychologist and neuroscientist are the once responsible for studying brain development and this paper can provide a motivation or an idea for their upcoming studies. While the ideal way to study human development is by observing the test subject and monitoring their changes in moods, behaviors, and body language, social media can also be a great tool to acquire data that can be beneficial. In this modern generation, a teenager in a first world country is almost guaranteed to have some form of social media. In a study conducted by Amanda Lenhart (2015) about the statistics of modern teenagers, she discovered that 91% of teens reported going online daily to their desired social media website, and 71% had more than one social media application downloaded on their phone. While teenagers and adults are pouring their hearts and souls on their social medias, with their consent, psychologist can use this social media addiction to their advantage by studying the tweets and statuses uploaded by users. With this study, scientist in the field can determine

how behavior, attitude, and maturity changed in their test subject throughout their social media lifespan. In this paper, I'll provide a small glimpse of a larger study that can be conducted.

3 Related Work

In order to understand human maturation, several studies have been conducted to expand why and how we develop from an infant that requires continuous attention, to a grown individual with the capability of understanding the theory of relativity. While psychologist can certainly perform experiments and evaluations on adolescence and following them up to adulthood, a simpler approach is to study their social media and see how changes occur within the tweets of the test subject.

In a study conducted by Pantic (2014) focuses on the addiction people have when it comes to social media and how certain comments and tweets attacked toward some users can escalate their depression and decrease self-esteem. Throughout the paper, a great explanation of how social media plays a huge role in mood swings and manages user's self-esteem especially for adolescents is argued. One of the studies explains that users with lower self-esteem tend to spend more time on social media compared to users with higher self-esteem. Pantic quotes, "The results indicated that individuals with lower self-esteem are more active online, especially in terms of having more self-promotional content on their SNS profiles" (Pantic, 2014). The outcome of this study is appropriate since another study performed by Raymer (2015) has the same hypothesis and same conclusion. In Raymer's study (2015) she had multiple hypotheses that she wished to test, but the one that stood out the most was, "Female students spend more time on Facebook compared to male students, being more vulnerable to a lower self-esteem" (Raymer, 2015). Although the previous paper did not include gender, Pantic conclude that lower self-esteem users are more likely to spend more time on social

media. In Raymer's conclusion, the same outcomes were present. She quotes "On a daily basis, females do spend more overall time in minutes on social networking site than males ... and they tend to spend more time uploading photos, posting statuses, and viewing others' profiles ... and are also more vulnerable to attacks about their appearances ... promoting a lower self-esteem" (Raymer, 2015).

To further this idea of individuals with depression and lower-self-esteem spending more time on social media self-promoting themselves, a study conducted by Choudhury, and Gamon (2011) calculate depression stages amongst social media users by studying their languages. Their study included a set of twitter users who reported being diagnosed with clinical depression. These user's twitter was monitored and observed for over a year to locate signs of depression by measurements of: user engagement and emotion, egocentric social graph, linguistic style, depressive language use, and mention of antidepressant medications. After acquiring their data, and performing measurements by relying on multiple online tools they finally reached a conclusion. In the conclusion the authors wrote, "Our findings indicate, for instance, that individuals with depression show lowered social activity, greater negative emotion, high self-attentional focus, and heightened expression of religious thoughts" (Choudhury, and Gamon, 2011).

In all three of the papers a similar hypothesis was created and tested, leaving a conclusion that was almost identical in all three papers. The hypothesis was to understand the correlation between depressed and low self-esteem social media users and the amount of time they spend self-promoting and writing tweets and statuses compared to non-depressed individuals. In all three of the papers, the conclusion repeated that lower self-esteem members on social media are expected to provide more tweets and statuses. However, Raymer's (2015) conclusion took a turn and included gender into the equation. Raymer

(2015) explained that females are more likely to spend time on social media being vulnerable to attacks from other social media users. She even included the perception of the ideal women into the conversation. By spending more time on social media observing and being exposed to the concept of beauty can lower many female's confidence and self-esteem. The discussion of the hypothesis and outcomes in Pantic (2014), Raymer (2015), and Choudhury & Gamon (2011) 's papers will also be tested and furthered in this paper. When performing this study on my twitter corpus I should expect a similar result as to the following papers.

4 Goal and Hypothesis

The goal of this study is to understand and see how teenage development can also be displayed through their social media tweets. To test the following statement, multiple small scaled studies will be attempted with unique hypotheses. The first study will include linguistic aspects to check each user's spelling development, bad word usage, and acronym usage. By diving the user's tweets based on a unique year, one can perceive if the following topics would appear frequent or less frequent as the user continues to age and mature. In my hypothesis for this portion of the study I predict that spelling development should increase, and bad word and acronym usage should decrease. Many would agree that as we age we acquire proper knowledge that would correct spelling errors, and decrease our bad mouths in a publicly viewable social media. In the second small scale study, I'll observe the emotions left behind by the tweets of users and see how they change or stay the same throughout the user's progression in age. In this portion of the paper, users tweet near holidays such as Christmas, New Years, and Thanksgiving will be monitored to detect if holidays play a role in mood changes. Along with this, I'll also be taking the hypothesis from the studies of Pantic (2014), Raymer (2015), and Choudhury & Gamon (2011)

which can be summarized as: social media users who struggle with low-self-esteem, depression and those who spread negative tweets are more likely to spend ample more time on social media applications. My hypothesis for the second study is that near holidays the majority of tweets will provide a positive trend. Also for those who provide negative and depression based tweets will be more likely to spend their precious time on twitter.

Finally, in the third study, I will record the corpus's statistics such as the number of tweets, re-tweets, pictures, personal reference, 2'nd person references to determine how aging plays a role in twitter statistics. In this section, I'll also be furthering Pantic's (2014) study in which she specifies that low-self-esteem social media members are more likely to provoke self-promotions. To hypothesize my final study, I predict that as the teenager continues to age, their twitter statistics should also increase. It would be reasonable to think that as adults there would be more topics, more problems, and more events to share on twitter increasing the duration of the time they spend on it. Overall, by conducting these three studies I can attempt to prove that social media member's development as humans can also be observed through their twitter pages.

4.1 Data collection

To begin, the data for this paper was collected using twitter's API library. By accessing a consumer and an action token, I was able to write a quick code in the programming language R. With the help of the twitterR and ROAuth library, I began to setup a twitter authorization that allowed me to grab specific user tweets. Finally, by specifying the username, and the number of desired tweets, the console would display an array of the tweets, as seen in Figure 1.

```
> getTweets<-userTimeline("realDonaldTrump", n=200)
> getTweets
[[1]]
[1] "realDonaldTrump: The White House is running very smoothly

[[2]]
[1] "realDonaldTrump: The Fake News is going crazy making up fa

[[3]]
[1] "realDonaldTrump: During Small Business Week, we celebrate

[[4]]
[1] "realDonaldTrump: The migrant 'caravan' that is openly defy
```

Figure 1: Data retrieve

The API library also allowed dates in the parameter along with username, and quantity which gave me access to study the tweets at a specific event. The tweets were separated by a unique year, and exported to a destination file. Since twitter's API only allows 3200 tweets to be extracted per user, I had to obey their regulations and work with 4 to 5 years' worth of data for the teenage accounts I had chosen. After cleaning up the data by removing unwanted symbols, I was left with my final corpus. As for selecting the user accounts for my corpus, I decided to choose random teenagers who provided their birthdates in their description to easily group them as teens. The final corpus looked similar to Figure 2.

```
[April 30, 2018]
realDonaldTrump: The Fake News is going crazy making up f
[April 30, 2018]
realDonaldTrump: During Small Business Week, we celebrate
[April 30, 2018]
realDonaldTrump: The migrant 'caravan' that is openly defy
[April 30, 2018]
realDonaldTrump: I recently had a terrific meeting with a bip
[April 30, 2018]
RT: "realDonaldTrump: Numerous countries are being consid
```

Figure 2: Final corpus

4.2 Study methodology

For the first study to observe the changes of linguistic aspects in the users tweets through 4 to 5 years of data, a small word comparison application in python did the trick. To detect the changes in grammar development, I collected some of the most common misspelled words from Oxford dictionary and stored them as a list inside a python file. By writing a quick code that looped through every line of the tweet file, I was able to check the frequency of how many times the misspelled words appeared in the tweet files. After

performing this technique on 4 to 5 files for each user, I was able to keep a record of how many misspelled words occurred throughout the user's dataset.

Next, to check how bad word usage increased or decreased throughout each user's data, I used the same python code as to the one used for grammar check by replacing the list from misspelled word to bad words. To test the usage of acronym such as lol, wyd, and atm, the same methodology was used. After running the code for each user file, the output looked similar to Figure 3.

```
>>> abbreviation('levil.txt')
Out of the total word: 10262
# of acronyms found are: 393
>>>
RESTART: C:\Users\fresh\Desktop\
e.py
>>> find_badword('levil.txt')
Out of the total : 10262
# of bad words found are: 39
```

Figure 3: Code results

For the second portion of the study, to determine user emotions, I acquire help from an online tool known as Meaning cloud that helped me perform sentiment analysis on each user's dataset. By inputting a file, the tool would compare the data to thousands of positive, negative, and neutral words with each word having its own priority. Overall the tool provided a simple positive, negative, or natural output to specify the type of tweets a user usually provides. In this application, context of the tweets is not concerned. With the output, one can suggest the general mood and emotion of each user from their tweets. With Meaning cloud, I was also able to test my hypothesis of whether holidays such as Christmas, New Years, and Thanksgiving provoke a positive trend in tweets. Also, to test the hypothesis of those providing negative and depression based tweets spending more time tweeting, retweeting, and uploading pictures was assisted with this online tool and the statistical data collected from the third study.

For the third study to calculate statistical data about the user's tweet such as the number of tweets, retweets, image posting, 1st person

references such as I, me, we, us, myself ...ext, and 2nd person references such as you, us, he, she, they, @...ext, was calculated by coding up a frequency function in python. The code would keep a frequency track of the searched words. By using the function, I can determine if the statistics about the user's tweeting increases or decreases as their age increases.

5 Results

After performing the three studies and collecting my results, in this section I will share my findings. To refresh, for the linguistic aspects of the first study I hypothesized that spelling development should increase, and bad word and acronym usage should decrease as the twitter user continues to mature. After using my resources and collecting the output data of the study, my hypothesis was true for the spelling errors decreasing as the users continues to age. In Figure 4, one can observe that compared to the first year's spelling errors (marked in blue) the years after follows a downward trend for majority of the users. Keep in mind, the target geographic of these tweeters are between ages 16 to 20.

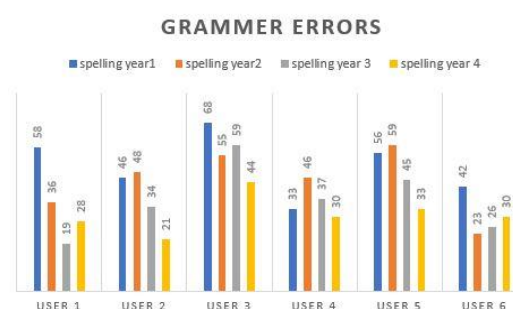


Figure 4: spelling error result

To be more precise, out of the 25-teenage user dataset, 19 of them showed a negative trend meaning that as age increases the development of spelling also increases. Along with studying spelling errors, bad word and acronym usage was also explored. I hypothesized that bad word and acronym usage should decrease as the user advances in age. After the study unfortunately, my hypothesis for both the bad word and acronym usage came out false. As

seen in Figure 5 and Figure 6, the usage of bad words and acronym usage actually increases, opposite from what I had expected.

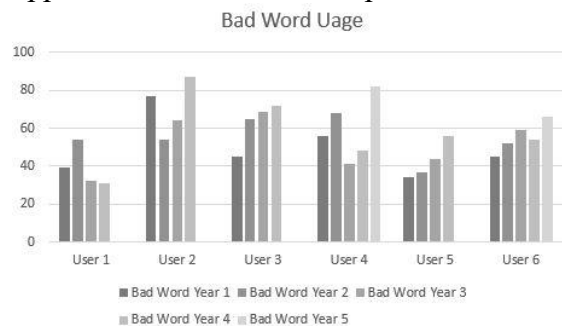


Figure 5: Bad word usage results

For bad word usage, the number increases for every subject except for 4 users out of 25. From this output, we can state that as teenagers between 16-20 enter adulthood they are more likely to rely on bad word usage. As for the acronym usage such as lol, wyd, and atm also showed an increasing trend for majority of the users. To be precise, out of the 25 users, 16 of them showed an increasing direction. A sample of the study data can be viewed on Figure 6. An interpretation of the outputs can be viewed in step 6-result interpretation.

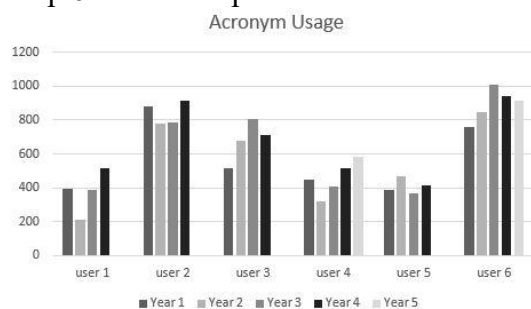


Figure 6: Acronym usage result

For the second study that dealt with emotions found on my dataset, I decided to use an online tool that helped me perform sentiment analysis called Meaning cloud. By providing an input file of each year for the user's tweets, the software generated an output of either positive, negative, or neutral emotion. The outputs from Meaning cloud classify the groups of words found in a tweet to be either happy, sad, or neutral. By determining one of the groups in which the user's tweets fell into, we can assume that as their overall emotion

found from twitter. For example, if a twitter member usually speaks about positive topics in their tweets that includes words such as happy, love, excited, blessed and much more, then we can assume that in general that user falls into a positive emotion category. For this study I hypothesize that user emotions near holidays would show a positive trend. My expectations and hypothesis was proved true by using Meaning cloud's sentiment analysis tool to determine each user's emotion types by providing a file of tweets near holidays. By dividing my files according to holiday events and uploading them to this software, I was able to clarify that near holidays most of the tweets indeed show a positive trend.

My second hypothesis for this study stated that social media users whose emotions fall into the negative category are more likely to spend ample time on twitter tweeting, retweeting, and posting pictures. With the help of the calculated statistics from the third study, I was able to conclude that social media users whose emotions fall into the negative category do spend more time tweeting, retweeting and posting pictures. Low-self-esteem, depressed, and those who spread negative tweets are more likely to spend ample more time on social media applications can be seen on Figure 7. Although Figure 7 only shows 6 users, out of the total 12 negative emotion based users, 9 of them showed having more tweets, retweets, and image posting than those who were in the positive category.

User	Tweets Total	Re-Tweets Total	Image Upload Total	Category
user 1	2540	1250	32	Positive
user 2	3001	2800	98	Negative
user 3	1084	3101	22	Positive
user 4	990	1978	12	Negative
user 5	3200	812	112	Negative
user 6	2456	1359	57	Negative

Figure 7: Tweet Statistics

For the thirds study, before calculating the statistics about each user's number of tweets, retweets, and image posting, I hypothesized that as the user continue to age their twitter statistics should also increase. After gathering my data, I was able to verify this hypothesis.

As seen on Figure 8, twitter user's number of tweets, re-tweets, and image posting does indeed increase as they continue to age. As for the second hypothesis in this study which explained that twitter members who fall into the negative emotion category are more likely to provoke self-promotions, was also proved true by comparing all the user tweets with the calculated first-person reference data. The results showed that negative emotion users are more likely to use words such as I, me, myself, we, us ... ext, in their tweets more than those who provided a positive trend in their tweets. Out of the 12 negative emotion category users, 10 of them showed having more personal references when compared to the 13 positive emotion category users.

User	Tweet Year 1	Tweet Year 2	Tweet Year 3	Tweet year 4	Tweet Year 5
user1	780	740	880	950	
user2	450	700	658	784	
user3	710	798	880	781	
user4	554	576	784	900	887
user5	312	641	218	477	
user6	571	410	559	634	687

Figure 8: Number of tweets per year

6 Result Interpretation

The first study of testing the user's spelling development, bad word and acronym usage resulted in spelling errors decreasing while bad word and acronym usage increasing. To interpret this output, one can assume that due to education and maturity, something like spelling errors are likely to decrease. But my hypothesis for bad word and acronym usage came out false which was surprising. I assumed that as twitter users develop from teenagers to early adults unmatured word usage would decrease but the opposite occurred. To understand this result I can assume that as we develop, bad word usage comes more and more naturally to a point where it's almost too easy to spread them on any platform. Also in my opinion, swearing provides more emphases on a sentence which can be used to convince a younger audience much quicker. For the second study to detect emotions found within my dataset, I performed sentiment analysis to group users according to

positive, negative or neutral tweets. By categorizing each user to a specific group, I proved that negative emotions category users are more likely to spend time on twitter tweeting, re-tweeting and sharing pictures. Along with this I also discovered that negative emotion category users (depressed) are more likely to invoke in tweets that includes multiple 1st person references. Both studies provide a reasonable result as they match the outcomes of Pantic (2014), Raymer (2015), and Choudhury & Gamon (2011) 's studies. To interpret this study, all of these authors would agree that depressed, low-self-esteem, and users that fall into the negative category of tweets seek more attention and drama compared to positive and neutral category users. Choudhury & Gamon (2011) mention in their study that, "Our findings indicate, for instance, that individuals with depression show lowered social activity, greater negative emotion, and high self-attentional focus". (Choudhury & Gamon 2011).

Finally, the third study resulted with proving my hypothesis that as the user continue to age their twitter statistics should also increase. My interpretation of this result is that as we reach adulthood there are much more responsibilities, events, gossip, drama, and problems to deal with, and as mentioned before in a time where sharing something personal on social media has become so easy, for many the number of tweets should definitely see a rise.

7 Further Directions

Although this was just a small glimpse of a larger study, there are many directions that can be furthered with this study as an introduction. Some topics that can be covered in the bigger study could include how privacy distribution can increase or decrease as the social media user continues to age? Another potentially topic could be how relationships between friends and family develop or fall as the social media member continues to mature. There are multiple studies that can be furthered by

understanding that as social media users continue to use these applications, they leave a trail and a route that can be studied to understand human progression.

8 Conclusion

In this paper, I began with a general topic statement that as we humans develop, we leave a trail of our development on the social media's we use. To study this statement, I gathered a twitter corpus of 25 teenage users that included tweets from 4-5 years. After cleaning up the corpus, I performed three studies. The first study attempted to measure the user's change in spelling development, bad word and acronym usage. The result of this study explained that as my corpus continued to age, their spelling errors decreased while bad words and acronym usage increased. The second study measured the corpus's emotions. By using an online tool, I was able to determine the user's emotions from all their tweets. With this data, I conclude that depressed and low-self-esteem tweeters are more likely to spend time tweeting, retweeting, and uploading pictures, along with seeking attention by having more personal promotions with 1st person references. Finally, the third study calculated each user's twitter statistics to conclude that as my corpus aged, their tweets, retweets, and image posting increased.

To conclude, this study was an introduction to prove that as teenagers develop, their growth can be observed through their most used social media accounts. Ample research is still required to fully prove this statement, but others are more than welcome to fully verify this topic.

9 References

Amanda Lenhart. 2015. *Teens, Social Media % Technology Overview 2015*. Pew Research Center Internet & Technology.

Kristine R. 2015. *The effects of social media sites on self-esteem*. Rowan university, Rowan digital works.

Igor Pantic. 2014. *Online Social Networking and Mental Health*. Cyberpsychology, behavior and Social Networking.

Mumun De Choudhury and Michael Gamon. 2011. *Predicting Depression via Social Media*. Microsoft Research Lab.

