Group Members: Michael Kinkoph, Morgan Soladine, Josh Sefton

Mentor: Hani Dawoud

I399 Final Group Paper

May 2nd 2012

**Abstract**

 Twitter has become one of the quickest ways to receive up to date news information about anything and everything going on in the world. Whether this information that is being spread around the world so quickly is true or false or not is very important to society. The consumption of false information could become a major problem if it affects people in a harmful way. Our focus was to find out why people follow parody accounts? Also, if users do follow parody accounts whether they know it is a fake account or not. In order to find out more about how user’s interact and use Twitter we first sent out a survey asking several questions on parody accounts and why people follow them. This data set was more qualitative and rather small due to unforeseen problems in our survey questions. Next, as a group we wrote a Java program that connected with Twitter’s API to find all of the followers for a certain group of people’s real account and parody account. Once we found all of these accounts we were able to find all of the common followers for a certain user. Lastly, we experimented with making our own parody account of a famous swimsuit model. This deemed to be our most useless research method but served as a good way to learn more about parody accounts through self-experimentation. After combining our survey and results from the Twitter API we came up with multiple conclusions.

First, we found that college students are more aware of parody accounts than any other generation. Second, we discovered a large amount of people that follow parody accounts do not follow the real verified user. It is hard to distinguish why people do follow parody accounts instead of the real accounts. We concluded that they follow parody accounts for humor, satire, or they have been fooled by the parody account into thinking the account is ran by a celebrity. In order to stop people from being fooled we proposed that Twitter should set up a recognition system for parody accounts similar to that of verified accounts. As social media becomes more pervasive in the way we consume information it is necessary to for us to make sure this information is true. If the information being received from social media is in fact not true, people need to be aware of this.

**Twitter: Real or Fake?**

Social Media has become more and more pervasive over the last ten years. The way we use, share and learn information has changed completely. Social media has allowed news to spread faster than ever to a mass amount of people. This has allowed people to be more knowledgeable about day-to- day news whether it is sports, politics or world news. Imagine a time when you had to wait until the six o’clock news to hear about breaking news. Imagine a time when you had to wait to hear news via radio at night or even wait until you received a newspaper the next morning. Social media and other types of technological innovations have allowed people to receive news via cellphone or Internet simultaneously of when breaking news happens. For example another project, which was completed within this class, showed that people were alerted of an earthquake in a nearby city before it even reached their city via Twitter. Another example is from a personal experience of mine with March Madness Indiana University Basketball games. I was on my way home from Florida and could not find the game on any radio stations. As an alternative, I used Twitter and following certain accounts allowed me to acknowledge updates of the game every minute or so. I was able to keep up with the game by simply refreshing my Twitter timeline every couple of minutes.

Twitter is a key mean to spreading information and knowledge. It is vital to know to what extent people do consume false information. The main point of our research was to figure out why users create and follow parody accounts. A parody account is an imitation of the style of a particular writer, artist or genre with deliberate exaggeration for comic effect [6]. Other examples of parodies are songs made by artist in attempt to intimidate or remix another song. A popular newspaper and website, “The Onion,” portrays itself as “America’s finest news source;” yet, they never release an article that is true. They often cross the line with these parodies in an attempt to make fun of a certain person or current event that is happening in the world. People often relate these to the “Colbert Report” and “Daily Show” with Jon Stewart. These shows serve to make fun of politicians and other worldly figures. They both have books published and have comedy tours that go around the world just by simply using parodies as a comedic tool. Lastly, “Saturday Night Live” has actors who dress up and act like celebrities or politicians and mock them by the way they talk or act. This show has been around forever and continues to be a hit simply by using parody to make fun of people. We chose to focus on Twitter parody accounts since this is a new type of parody that is gaining a lot of attention from the public. What is the significance of parody accounts to have hundreds of thousands of followers but fail to tweet even once a month? We hope to answer this question and figure out how much false information is consumed by Twitter users.

Twitter was created in 2006 and is already one of the largest social media sites in the world. Although Twitter was founded in the United States, in San Francisco, it spread very rapidly throughout the entire world. Twitter currently has over twenty different languages and you can simply set your language preference when you create your Twitter account. You can also change your language preference to almost any language you desire by simply selecting it in your settings of your account at any time you want. This is one reason that contributes to Twitters success; it is a known fact that almost every single country in the world has Twitter users. A couple other key features Twitter has to offer include Twitter on the go, for all the smart phone users out there, and Twitter for SMS text messaging, which works for most of the mobile phones that are on the market today. Twitter on the go is used through the free Twitter application that you can download on iPhones, iPads, or any other phone from which you can download apps on. If you do not have a smart phone, Twitter allows most of the other mobile phones out there today to connect with other Twitter users around the world via SMS text messaging. It is features like these that contribute to the use of Twitter all over the world. Tons of research has been completed on Twitter due to its simplicity and short range of uses. Although it lacks variety, it is very powerful due to the way information is shared and created on the website. For those unfamiliar with Twitter it can be easily compared to Facebook. However, Twitter does not have events (such as birthdays), chat, or pictures. Pictures can be shared using Flickr, Twitpic, and Instagram but are not saved on the users Twitter profile. Basically, Twitter is more of just a way to update your Facebook status. But rather than updating it maybe only once a day, Twitter marketed itself to tweet multiple times a day in an attempt to update people where you are at all times. Twitter is spam free except a few advertisements and people can only tweet 140 characters each time. The website itself is very easy to do research on massive amounts of people. Most research that is conducted on Twitter is done on verified accounts.

Verified accounts go through a strenuous process in order to become verified. First, the user must have a public account. Public accounts allow other users to re-tweet, see followers and see those who are followed by the user. Since these accounts are public it allows any researcher to connect with Twitter’s API to find massive amounts of data on public users. API stands for application programming interface, API’s allow for programming languages to connect with a website. This involves crawling through the HTML, parsing data and allows for dynamically posted content. Dynamically posted content means an application like “Tweet deck” can post a status on someone’s Facebook, Twitter and MySpace all at the same time. The way we connected with Twitter’s API was by crawling through the HTML by writing a Java program to find common followers. We will go into more detail about that once we reach the research methodology. Continuing on with the verification process, once a user has a public account they must actively tweet from this account in order to be considered for a verified account. Next, the user must establish a large following base; usually this has to range above 20,000 followers. Once this happens, Twitter representatives will contact the user and make sure they are in fact who they claim to be. For example, Lady Gaga would create a public account, attain at least 20,000 users and actively tweet, only then would representatives from Twitter contact her to verify her account. Once an account has been all the way through the verification process, they then receive a blue checkmark next to their name. This denotes to all of their followers that this is actually the real celebrity. This is the only way Twitter recognizes celebrities for being who they say they are; hence as to why so many people are fooled by parody accounts. However, parody accounts will never be verified no matter how many followers they receive. As there are strict guidelines with the real account verification process, there are many rules behind parody accounts. Twitter’s support team states:

 “Twitter provides a platform for its users to share and receive a wide range of ideas and content, and we greatly value and respect our users' expression. Because of these principles, we do not actively monitor users' content and will not edit or remove user content, except in cases of violations of our Terms of Service.” [4].

 The support team also requires that parody accounts must denote in their name and username “not” or “fake” so they do not fool people. Twitter support also restricts impersonation of another person in an attempt to fool or deceive people. However, through our research of parody accounts, we found that many accounts do not follow these guidelines. For example, Jay-Z has a real account and a parody account. His real account is “@S\_C\_” and has 1,700,000 followers, while his parody account is “@Jay z” and has 700,000 followers. This is one of the smallest ratios of celebrities we found of real followers compared to parody followers. Our reasoning behind this is that the parody account is his actual stage name and people are confused by this. Although this seems like a violation of Twitter policy it is not because Jay-Z’s real rap name contains a dash, while the parody account name does not. Other Twitter policies include trademarks and other name violations as well. This relates well to Twitter’s name squatting policy. Name squatting is the act of trying to prohibit a celebrity that does not already have a Twitter account to take possible names. If the user squats on a name and shows no activity for six months Twitter support holds the right to deactivate the account [2]. For example, I could possibly make an account for an up and coming CEO of a company that does not yet have a Twitter. People attempt to do this in order to gain a massive amount of followers very quickly. Or they have hope that maybe this CEO will pay them for that Twitter name, but really that would be breaking the law. Therefore, Twitter attempts to squash anyone’s account that seems as though they are trying to “squat” on an account name. Twitter support states that: “We reserve the right to reclaim user names on behalf of businesses or individuals that hold legal claim or trademark on those user names. Accounts using business names and/or logos to mislead others will be permanently suspended.” [3]. One example of Twitter taking action on a parody accounts involved Steve Jobs. The parody account was mimicking Steve Jobs, CEO of Apple at the time, the account accumulated 450,000 followers before being shut down, Twitter support allowed the account to be reactivated after the user changed the name from “@CeoSteveJobs” to “@FakeCeoSteve” [7]. This is one of many examples where Twitter support has had to step in to take care of people attempting to impersonate another person. Ultimately we hope to find out why people attempt to impersonate certain celebrities and the effect it has on people who do not know that these Twitter accounts are not the actual person.

As stated earlier our first form of research methodology was making a survey and distributing it to as many people as possible. Our methods of distribution included another social networking site, Facebook, OnCourse and e-mail. All three researchers created an event and sent the survey out to as many pupils, friends, family members and co-workers. This was the main way we ended up getting people to complete the survey. In the end, we received around fifty results from the survey using a website called Survey Monkey. The results were very consistent and the idea of having open ended answers at the end was definitely helpful in giving us qualitative feedback. When handling surveys for research we will continue to use Survey Monkey. The survey is very easy to manage and once you have enough results, Survey Monkey organizes these results into nice structured data with the use of charts and graphs. Unfortunately, some of our questions were rather vague and gave us a variety of different answers. We learned you should make the survey as simple as possible and if the question is not an open ended answer that it should be able to be answered with “yes” or “no.”

Another problem with our survey is the population of people that we surveyed and the amount of time we had to actually survey people. First, we had a small data set, it is very difficult to distribute a survey and get people to actually take the survey and answer with seriousness. The people we sent the survey to were mainly college students which gave us a bias result. However, we were able to make certain conclusions on how college students react to social media because of this. For our next survey we plan to use other websites and try to give people an incentive to take our survey. In order to find a larger dataset with more of a variety in our population we plan to find different ways to distribute the survey.

* The second form of research we used was a more technical route that served to be much more helpful in the long run. We used automatic data mining, which is the process of discovering new patterns from large data sets involving methods of statistics and database systems. This process consists of discovering new patterns from large data sets involving methods of artificial intelligence, machine learning, statistics and database systems. Our overall goal of the automatic data mining process was to extract knowledge from a data set in a human-understandable structure. Data mining is a rather new way to find a large amount of results by usually connecting with a website’s API. Our mentor, Hani, helped write a Java program that involved crawling through the HTML of the website and finding followers of certain celebrities. Remember, we are able to crawl through the HTML by connecting with Twitter’s API to find all of these followers because all of the accounts we dealt with were public. As a group we selected ten musical artists that had a real account and parody account. These artists are also very popular right now. This was a good group to choose because they have a large amount of followers on Twitter and seem to influence people on a day-to-day basis. All of the real accounts that were crawled through are verified on Twitter. All of the parody accounts that were crawled through have at least 30,000 followers and of course have a public account.
* The parody accounts chosen were all an attempt to impersonate the counterpart real account. As stated earlier, Java was used to connect with Twitter’s API to find all of the corresponding followers for the ten celebrities chosen. Two different Java programs were written. First, a Java program was written to find all of the followers of the real account and the parody account. There were certain limitations in this process due to constraints with Twitter’s API. Twitter only allows 140 requests of 5,000 followers per hour, meaning that our program could only retrieve 700,000 followers per hour. However, during the research process we found that some significant researchers are given a white flag. This white flag enables them to bypass this limitation and find make as many requests as they would like at any given time. The point of the Twitter API restricting this is because there would be no point in even having the Twitter API if there weren’t restrictions on it. This became a problem due to the fact some celebrities have over ten million followers which took up to a day just to retrieve one celebrities followers. This constraint mixed with our busy schedules and not always being able to run the program prohibited us from coming up with a bigger set than ten celebrities. Once the program ran once and had not retrieved all of the followers it would return a cursor number. The next time the program was ran this cursor number would tell the Twitter API where to start crawling again. This way the Java program would not find the same followers each time but only new followers the program had yet to find. Each time the program ran, all of the followers found were copied and pasted into a text file. You would think the program would find the actual Twitter names of the followers pertaining to the certain celebrity. However, all social networking sites give each account name a user identification number. This makes it easier to put into a database and differentiates identical account names. Therefore, the text files contained all of the identification numbers of the accounts that follow the certain celebrities.

From our two research methodologies, we concluded a variety of different things. We found from the survey that people in college are very knowledgeable about how Twitter parody accounts work. The main problem with our survey was that all of our participants were college students. However, we were able to make certain conclusions about this specific age group and their involvement with social media. The first question asked in our survey was “Do you follow a parody account on Twitter? Yes or no? If yes, why?” After this question we defined what a parody account was. The survey went on to ask other questions about their knowledge and interaction with parody accounts on Twitter. Our results found that most college students either follow a lot of parody accounts (three or more) or do not follow any at all. Sixty percent of our participants responded to following more than three parody accounts or none at all. All of the people that do follow parody accounts seemed to know that this was an actual fake account. Also, other users that did not even follow parody accounts knew about the verification process with Twitter celebrity users. When participants were asked why they follow parody accounts, the main idea was for humor or to mimic certain people by tweeting funny quotes or pictures. About seventy-five percent of our participants noted that they do not receive any extra information that allows them to learn more about the real celebrity from following their parody account. Once again, our survey was conducted over a short period of time and did not return as much data as hoped. Our survey set was mainly college students which gave us a bias result of information. College students are probably more aware of social media sites than most Twitter users. They may use social media sites to connect with people rather than use it for news which was out of the scope of our research. Even though our dataset was rather bias it allowed us to make conclusions on a certain age group. Also, when comparing this data to the data mining results we were able to make other assumptions.

 The data mining results returned contrasting results compared to the survey. The survey

results found what we were hoping for because it meant that people were knowledgeable about parody accounts and not consuming false information. However, the data mining results showed that most people that do follow the parody account of a celebrity do not follow the counterpart real account. Out of the ten musical artists chosen, we found common followers of only one artist that had over fifty percent of people that followed the real account compared to the parody account. This artist was Wiz Khalifa and we related his strong follower population to both the real and parody account to how committed his fan base is to him. Also, we assumed that this parody account tweeted humorous things or inspirational quotes. Another thing was that the parody account contains his name with just minor spelling differences. This relates to misleading people into thinking it really is him, therefore they are consuming false information. Surprisingly enough Twitter has yet to deactivate this account, but some action may be taken in the near future. As for the other artists that received less than fifty percent of a following rate on the real account compared to the parody account, we can assume the following. Unfortunately a lot of people seem to be fooled by these parody accounts and are consuming false information. Nevertheless, according to our survey a lot of these people may be following this account simply for humorous reasons, or at least we hope so.

 According to an article by Matt Carmichael, less than fifty percent of Twitter users are between the ages of 18-29 [8]. Therefore, older age groups are less knowledgeable about parody accounts and social media in general. This may be the cause of this low ratio of parody account followers to real account followers. We hope that most people choose to follow these parody accounts simply for humor and may not have any interest in the actual musical artist at all. According to our survey this runs true because most people are very aware of the functions of parody accounts. Besides finding out that people are unaware they are following real accounts or parody accounts, we concluded they might not necessarily consume false information. A lot of times parody accounts are not a parody at all and give out other information. We decided parody accounts are not necessarily a bad thing. Nonetheless, if people are not aware the information is false they are receiving this could become a major issue. Twitter can spread information to a large amount of people in a very quick manner. If this information is false, then we can only imagine the consequences and turmoil that will have to be dealt with for serious events taking place.

Twitter is going to continue to shape the way people absorb information. Whether or not people are absorbing true or false information depends on the user’s knowledge of how to correctly use the website. Our goal was to find out unknowingly how much false information is consumed on Twitter. Unfortunately, it seems as though a large amount of people consume false information from celebrities on a day to day basis. Therefore, we proposed a plan to inform Twitter users better of the difference between parody accounts and real accounts.

An earlier focus on the verification process of Twitter stated that user’s that are verified have a blue checkmark next to their name. First of all, this is not enough recognition for verified users since most people are not even aware of this process. So, Twitter must make a more formal dedication to inform users of verified users. Next, they must differentiate verified accounts and parody accounts. Rather than not doing anything at all to parody accounts there should be some sort of notification that the user is a parody. A different color check marked may work and would be helpful to alert people of the parody. But, also when a request to follow a parody account is initiated there should be some sort of notice that states “This account is a parody, do you still wish to follow them?” For users that are not aware of what a parody account is, the message should also entail a short definition of what a parody account really is. Lastly, for verified accounts there should also be an alert message that states the opposite, something like “This account is a verified user, which means this is the actual real celebrity, do you still wish to follow them?” Of course these are all just ideas and the Twitter support team could probably come up with something more intuitive and user friendly that alerts people of parodies.

 This may not seem like a big deal now; however, if the consumption of false information continues to grow, it could become a major problem. Breaking news, emergencies or small lies on a parody could eventually end up hurting society if enough people are affected. For the future, Twitter makes it easy to analyze and database information. From doing research for only three months, with a group of four people, with about a six-hour commitment per week, we were able to draw heavy conclusions about this. Imagine a group of experienced researchers committing all of their time to a study on Twitter; the results would be very impactful.

Indiana Universities own Johan Bollen struck it big by building an algorithm to predict the stock market judging by people’s mood on Twitter. There are tons of different ways in which data can be analyzed on Twitter. The beauty of it is that Twitter is so simple and easy to find certain information. As social media continues to affect the way news is distributed and shared, it is vital to know that this information is true. In order to do this, social media sites like Twitter must continue to inform users of how the website works. Doing this will prevent users from consuming false information and the website can serve its rightful purpose of informing people of breaking news and events as quickly and truthful as possible.

Acknowledgments

 We would like to thank our mentor, Hani Dawoud, for patiently working with us and helping us out this semester. We are very appreciative that he took time out of his day no matter what he was doing to come help us and make sure we understood everything, which made our research and project a lot less stressful. We would also like to thank our class of Informatics I399 for participating in our research by taking our survey. Lastly, we would like to thank professor Fox, the class AIs, and anyone else who came into class to speak to us. We greatly appreciate your knowledge that you shared with us, which helped us complete our project in a more efficient manner.

* Citations
1. https://support.twitter.com/groups/33-report-a-violation/topics/148-policy-information/articles/18366-impersonation-policy#

2. https://support.twitter.com/groups/33-report-a-violation/topics/148-policy-information/articles/18370-name-squatting-policy#

3. https://support.twitter.com/groups/33-report-a-violation/topics/148-policy-information/articles/18311#

4. https://support.twitter.com/groups/33-report-a-violation/topics/148-policy-information/articles/106373-parody-commentary-and-fan-accounts-policy#

5. <http://platformmagazine.org/2012/03/parody-twitter-accounts-friend-or-foe/> (never used)

6. http://dictionary.com

7. Gross, Doug. "Twitter Suspends Fake Steve Jobs Account." CNN. 02 Mar. 2011. Web. 01 May 2012. <http://articles.cnn.com/2011-03-02/tech/twitter.steve.jobs\_1\_twitter-account-apple-ceo-steve-jobs-parody?\_s=PM:TECH>.

8. Carmichael, Matt. "The Demographics of Social Media." The Demographics of Facebook, LinkedIn, MySpace and Twitter. 16 May 2011. Web. 01 May 2012. <http://adage.com/article/adagestat/demographics-facebook-linkedin-myspace-twitter/227569/>.