INTERNATIONAL MASTER 1 (COMP SC.) PROJECT
UNIVERSITY OF NICE SOPHIA ANTIPOLIS

Analysing the Twitter social graph: Whom can we trust?

Evgeny Morozov[*], **Mourjo Sen**[+]
Email: [*]pgrad@realityexists.net, [+]sen.mourjo@gmail.com

Under the guidance of
***Arnaud Legout** and Maksym Gabielkov*
Email: { **arnaud.legout**, maksym.gabielkov } @inria.fr

Abstract

Twitter is a popular microblogging service that enables users to send short messages ("tweets"), most of which are public. Users can follow each other to receive these tweets, which creates a directed graph of followers. Therefore Twitter can be considered as a medium of information broadcast – where some people produce new information, and others consume it. Much like any other information source, the question of how trusted an information source is (users, in this case) is of primal importance. In this project, we study the Twitter social graph to look for ways to identify users who can be "trusted", that is, users who are who they say they are and whose posts can be considered genuine. We do this by analysing characteristics of the users and their position on the social graph. There is a set of users who have been verified to be genuine by Twitter, which is typically done for celebrities and notable businesses. We analyse other users' trust scores based on the assumption that verified users can be trusted. A key part of the work is to establish objective metrics that can serve as proxy for the subjective notion of "trust". We first worked on a smaller data set of Twitter in 2009 and explored different metrics and found the metric of "trust score" to be the most efficient in classifying users as trusted. We then applied this method on the more recent and larger dataset of 2012 to support our claim that the trust score is a good metric in estimating how trusted a user is.

April – June 2014

Table of Contents

1 Introduction

Twitter is a very popular online social networking and microblogging service that enables users to send and read short 140-character text messages, called "tweets" [1]. Most of these messages are public and can be viewed without even creating a Twitter account. Twitter has more than 500 million users (as counted in 2012) and about 255 million monthly active users [2].

On Twitter, users can "follow" one another to automatically receive any tweets posted by the user being followed. The follow relationship is not symmetric and therefore a person following another person does not mean that the converse is true. In other words, the links on the Twitter social graph are directed. So, Twitter is also considered as a news medium along with a social networking medium [3]. This makes Twitter different from other social networking sites like Facebook, which primarily uses undirected links between users resulting in the symmetric relationship of "friendship". Of course, symmetric relationship can be achieved on Twitter when two users follow each other. Therefore, Twitter is more general and is closer to how information propagates in real life [4]. But it has been found that following is mostly not reciprocated [3]. Rather, there are a few users who have many followers and can therefore directly reach a large audience.

1.1 Goal and motivation

There are about 500 million tweets published every day [2]. It is easily possible for someone to pretend to be someone else (impersonation) on the internet and/or publish fraudulent information with malicious intent or for unethical professional gain, or simply to create rumours [5]. Therefore, it is very important for any information source (here users publishing tweets) to be proved authentic and trustworthy. With more than 500 million users on Twitter, it is almost impossible to manually verify the identity of every user who signs up on Twitter and it is even more difficult to keep track of users who tend to spread information of questionable authenticity. The goal of this project is to find a mechanism to *automatically* analyse the Twitter graph and find users who can be trusted. Once there is a metric to measure how trusted a person is, it will enable consumers to decide whom to listen to (follow) and whom to avoid.

Some users on Twitter have been manually verified by Twitter following administrative processes of cross-checking their profile with their identity. According to Twitter, "Verification is currently used to establish authenticity of identities of key individuals and brands on Twitter" [6]. But only a very small fraction (0.006%) of users is verified by Twitter. Intuitively, we expected Twitter to have more than only 0.006% trusted users. Our main motivation was to explore the notion of trust and authenticity for *unverified* users on Twitter.

1.2 Methodology and objectives

Though "trust" is a subjective topic and is open to interpretation, for this project, we define being trusted as the likelihood of being who one claims to be, that is judging by a person's position on the Twitter graph, how probable is it that he is who he claims to be. To estimate this, we use a metric called “trust score” based on the number of verified followers a user may have. The reason for our choice of this metric as a measure of trust is defined later in this document.

The main objective of this project is to use the trust score to identify the unverified users who can be trusted. But before we can claim users to be trusted by using our method, we have to show that our method is in fact reliable and efficient in identifying characteristics that are expected to be found in trusted users. We also want to show that our metric for trust is unique in identifying trusted users and cannot be substituted by other metrics that exist. Finally, we want to give reasons why our method can also be used in estimating trust on social networks in general.

1.3 Paper organisation

The document is divided into six following sections explaining different topics. Section 2 gives the details about the trust score – how it is calculated and why it is a measure of trust. In section 3, we describe the development process and how we arrived at the trust score. In this section we also mention the other metrics we explored and details about the datasets we worked with. In section 4, we talk about the results obtained from out method. Section 5 explains how the trust score is evaluated, and gives justification to our choice of it as a measure of trust. In section 6, we talk about the challenges faced in the project. In section 7, we conclude by discussing about future work.

2 Our method: The trust score

The goal of this project is to define a trust score for every user such that it indicates how trusted that user is. The trust score is primarily important for users who have not been verified by Twitter as it gives an idea of how authentic a user’s profile is likely to be. But before we describe how the trust score is calculated, we need to discuss the different types of users we have used in this project.

2.1 Categories of users

For the development of this project, we divided the users of Twitter into three groups, as follows:

1. Verified users: The users who have been manually verified by Twitter constitute 0.006% of the total population of users in Twitter. Verification is only initiated by Twitter and it cannot be requested by a user. It is typically performed for celebrities and public figures, including notable businesses. We consider a verified user to be trusted (relative to an unverified one), since to earn verified status, they must use their real-life identity and be sufficiently well-known by the general public for Twitter to consider that there is a risk of someone impersonating them. As stated by Forbes.com, several businesses have been impersonated, with imposters posting tweets that threaten to malign their business’s reputation [7]. The verified status is thus a mechanism to uphold the integrity of the user’s identity. Therefore we consider a verified user to be trusted.

2. Expert users: Expert users have been collected by Naveen Sharma et al [8] and by their study they have shown that these users have a high influence on Twitter because they are related to a certain domain where they are popular. Amongst these users, there are users who are experts on very niche topics, such as robotic space exploration, and stem cells. They have a considerable impact on the Twitter social graph. The list of experts has been obtained from the "who-is-who" service by Naveen Sharma et al [8] at http://twitter-app.mpi-sws.org/who-is-who/. The set of expert users, however, includes verified users as well. 0.91% of expert users are verified. So there are some users who fall in both the categories of experts and verified users. But for this project, we only consider the unverified experts as a set and unless explicitly mentioned, henceforth, all references made to experts mean unverified experts. We now have our second distinct category of users, called "experts".
3. Other users: Experts and verified users constitute less than 0.6% of the users. That leaves us with more than 99% of the users, and we classify them as the "other" category. This category consists of users who are neither verified nor experts.

The different classes of users are shown in Figure 1. It should be noted that there is no user who does not fall in any of the above three categories. Using the three categories of users we just described, the goal of this project is to find a fourth category of users, called **trusted users**, which would be an intersection of the three users described above. The most interesting subset of the trusted users is the set of users who fall in the "other users" category. That is the purpose of this project – to evaluate how trusted users are who have not been considered by other methods.

It is important to note that by our method, there may inadvertently be a small fraction of verified users who are not considered to be trusted by our method, thereby apparently contradicting the point we made above (verified users being trusted). But the goal of the project is not to re-verify users who have already been verified, but to apply it to the users who have not been verified. Our method uses the verified users' class to determine how trusted the unverified users are, and it is not meant to be applied to cross-check the trust of verified users.

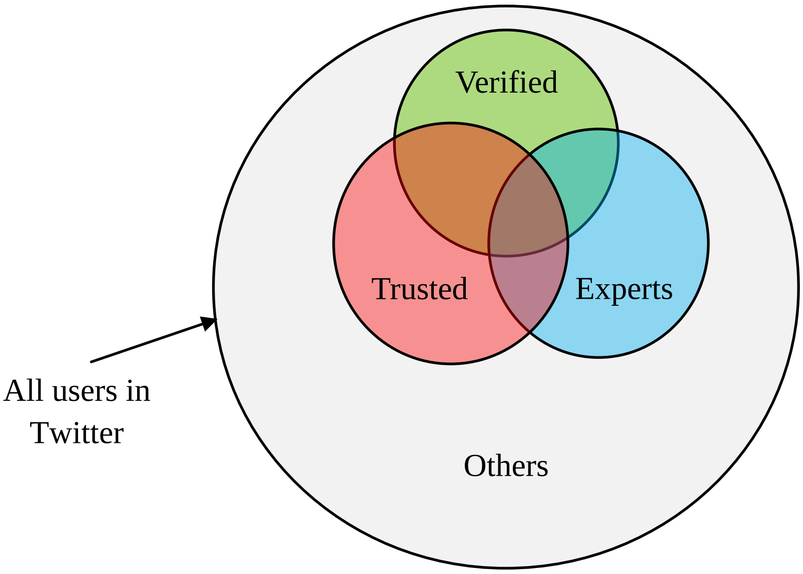


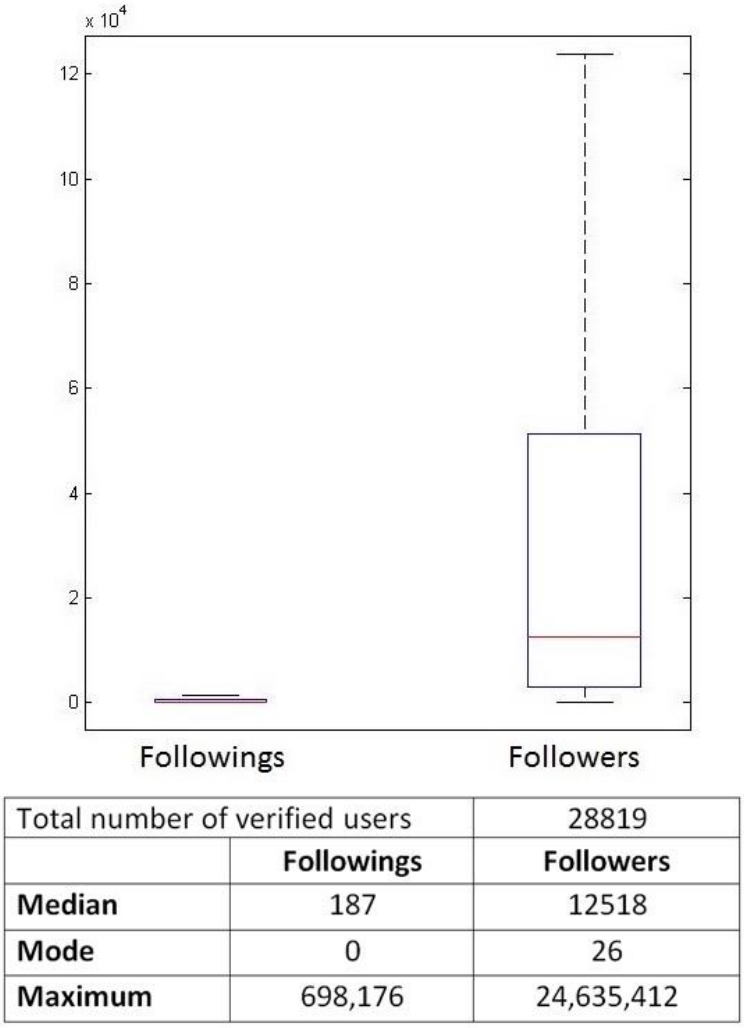
Figure 1: Venn diagram showing the different classes of users (the diagram is not to the scale of the number of users in each category)

2.2 Why the trust score is indicative of trust

As we stated in the previous subsection, only 0.006% users are verified on Twitter and they are considered to be trusted. And since verified users are guaranteed to be of authentic identity, therefore, if an unverified user has many verified followers, it suggests that this user is perhaps more trusted than users who have no verified follower at all. This is basically how we calculate the trust score.

When a user A follows another user B, it implies a level of trust that A has in B. Generally, people follow those who they personally know or are public figures, but no *legitimate* user is likely to follow a known spammer or a user known to be impersonating someone else. The quality of followers thus provides a measure of trust. But the question is how to figure out which follower is *legitimate* or, how to estimate the quality of the followers of a user. Since we know for certain that verified users are trusted, we can estimate the quality of followers by the number of verified followers a user has. Trusted users confer trust onto those they follow and in our case, verified users are considered to be trusted, and so users with a high number of verified followers are also expected to be trusted.

Verified users (0.006%) are mostly celebrities and famous personalities. They tend to follow people selectively, although they often have a large number of people following them. This is exhibited by the boxplot in Figure 2 depicting the number of followings and followers of verified users. We use this selective following characteristic of verified users as a checkpoint for estimating how trusted other users in the system are. Figure 2 also shows that the median of the number of followers of verified users is much larger than the median of the number of followings of verified users, adding more support to our claim. Since verified users follow back only a small number of users, we can postulate that when any user has a high number of verified follower count, it means that the user is more trusted than one with fewer or no verified followers. This postulation is justified by the fact that if we only consider only verified users to be trusted, a person with a high number of verified followers shows that that person's activities (tweets) are followed by many trusted users, indicating that he himself must also be trustworthy.



Total number of verified users		28819
	Followings	**Followers**
Median	187	12518
Mode	0	26
Maximum	698,176	24,635,412

Figure 2: Verified users are followed by more people than they follow. "Followings" are users whom verified users follow. "Followers" are users who follow verified users. The boxplot (outliers are not shown) shows that verified users tend to follow restrictively although they have a large number of followers. This is also shown by the median and maximum values of the number of followings and the number followers of verified users. Therefore, verified users follow selectively, thereby unintentionally providing a checkpoint on the trust of other users in the Twitter graph.

2.3 How the trust score is calculated

The trust score is essentially the number of verified followers that a user has. The score goes from 0 (no verified follower) to 400 (greater than or equal to 400 verified followers). As shown below, the trust score S_u for a user u is calculated using the number of verified followers f_u^v that the user u has:

$$S_u = \min\ (\ f_u^v, 400\ )$$

The score is given the upper bound because only 0.0004% of the users have more than 400 verified followers, and many of them are verified. Figure 3 shows how the score is calculated in the Twitter graph. We first count the number of verified followers a user has, and then put the upper bound of 400.

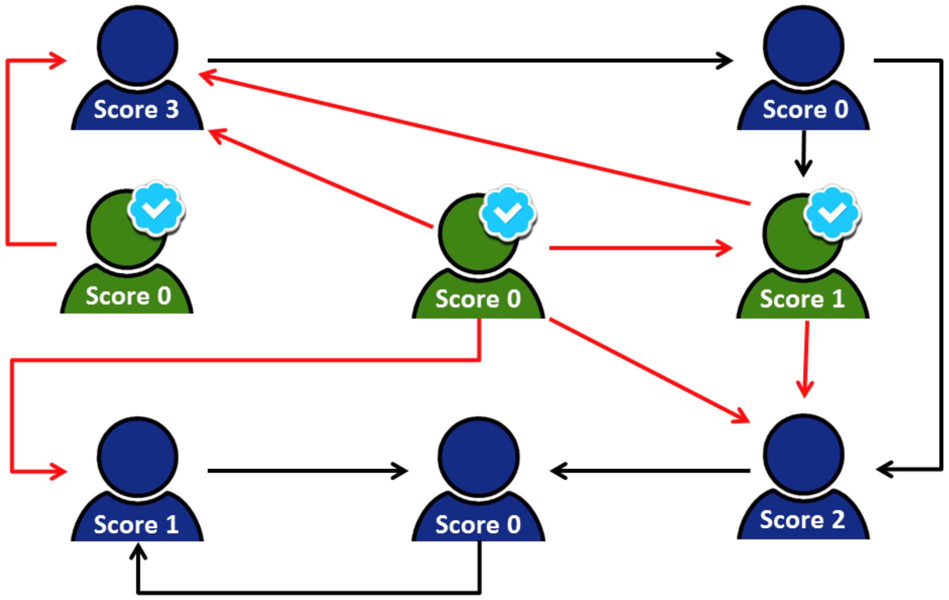


Figure 3: How the trust score is calculated - the red arcs show who the verified users (in green) follow. The score is calculated by the number of verified followers that a user has.

3 Evolution of the trust score

3.1 Initial work on the 2009 dataset

We started our work on the 2009 dataset, which was collected by Kwak et al [3]. This dataset of Twitter was collected in 2009 and it has 41.6 million users. This dataset was used in the initial stages when we were still investigating the best metric to measure trust of a user. The different classes of users in this data set are given below:

Type of user	Quantity	Percentage
Verified	13,016	0.03%
Expert	1,455,720	3.63%
Other	38,634,545	96.34%
Total (all users)	40,103,281	100%

Please note that all figures and data given outside of this section 3.1 refer to the 2012 dataset and not the 2009 dataset. We only mention the 2009 dataset with regard to the development process and how we arrived at the trust score. The final result was carried out on the 2012 dataset.

3.1.1 Preliminary data processing on the 2009 dataset

We wrote a program in C# to do most of the data processing and output the results to a comma-separated-values (CSV) file. This file was then loaded into Matlab to plot charts and perform any statistical analysis. Excel was also used for some tasks. While Matlab theoretically has the ability to load the raw data and process it, we found that it is far too slow and consumes far too much memory to be of any practical use. This made it necessary to write a separate program to work with the raw data. The program had to be both fast and memory-efficient. C# was our language of choice, because it has a large, useful standard library (.NET framework) and no memory overhead for primitive types.

The first task was to store the data in a more compact format, so that it could be loaded more efficiently. The original data contains pairs of numeric Twitter user IDs in text format, for example, "12 13" to indicate that user 13 follows user 12. This file was about 24.3 GB in size. We converted this to a binary representation of the adjacency list, storing each ID as a 32-bit integer, and grouped the list by the user being followed. That is, we stored the ID of each user being followed only once, together with the IDs of all their followers, rather than the raw list of (followed, follower) pairs. This format was far more efficient: the resulting file was about 5.8 GB in size and took 80-90 seconds to read. The lack of memory overhead for primitive types in C# was important here, so that a 32-bit integer actually took only 32 bits of memory. This allowed us to read the entire file into memory on a machine with 16GB of RAM. Additionally, our binary file distinguished between verified and unverified users to avoid the overhead of reading the list of verified users from a separate text file every time and checking each user against this list. (Verified user IDs were internally represented as negative integers, so that the verified "flag" took no additional memory, but this was hidden from user code by a layer of abstraction).

3.1.2 Metrics explored on the 2009 dataset

Once the data was processed and efficiently stored, we explored different metrics that we thought would be helpful in estimating trust. We will now illustrate those metrics, how they were calculated and why they were not taken to be the best metric.

3.1.2.1 Number of followers

We computed the average number of followers for verified and unverified users. While verified users are often thought of as "celebrities", not every verified user has many followers. Some have fewer than a hundred [9] and, indeed, Twitter explicitly states that the number of followers is not one of the criteria that determine who gets verified [6]. Nevertheless, we found that the average verified account does have many more followers than the average unverified one: 20,284.4 vs. 28.9 in the 2009 dataset. This is a 700-fold difference, which confirms that verified users are, as a group, not only trusted, but influential: they can directly reach a large audience with their tweets.

3.1.2.2 Degree of trust

When one user follows another, this implies some level of trust: many people follow people they personally know or famous personalities/businesses, but no genuine user is likely to follow a spammer or user known to be impersonating a celebrity. The number of followers is, therefore, one possible measure of trust (as well as influence). However, spammers and impersonators may well follow each other or have dummy accounts following them in order to give the false impression that they are real, important people. Therefore the level of trust of the followers is also important: the more trust a user has the more trust they confer onto those they follow.

We treat the verified users as implicitly trusted, as outlined above. We further assume that if a verified user A follows an unverified user B, it indicates some degree of trust in user B – they are more likely to be personally known to someone trusted than the average user. User B is therefore more trusted than the average unverified user (since they may be known to a verified user), but still less trusted than a verified user (since they are probably not a celebrity themselves). An unverified user C, who is followed by B, is less trusted than B and more trusted than an average user because he has a follower who is followed by a verified user. We generalise this idea into *degrees of trust* (where lower degree means "more trusted").

The degree of trust is defined as the minimum distance of a user from a verified user in the Twitter graph. A verified user is assigned degree 0. A user who is not verified, but has at least one verified follower is assigned degree 1. A user that is not verified and has no verified followers, but has at least one degree 1 follower is assigned degree 2, and so on. There is also a class of "no degree" users meaning there is no path leading from any verified follower to these users. There is also another group of users who have no followers at all, and thus there exists no path from verified users to these users either.

The number of users of each degree (in the 2009 data set) is shown in the Figure 4.

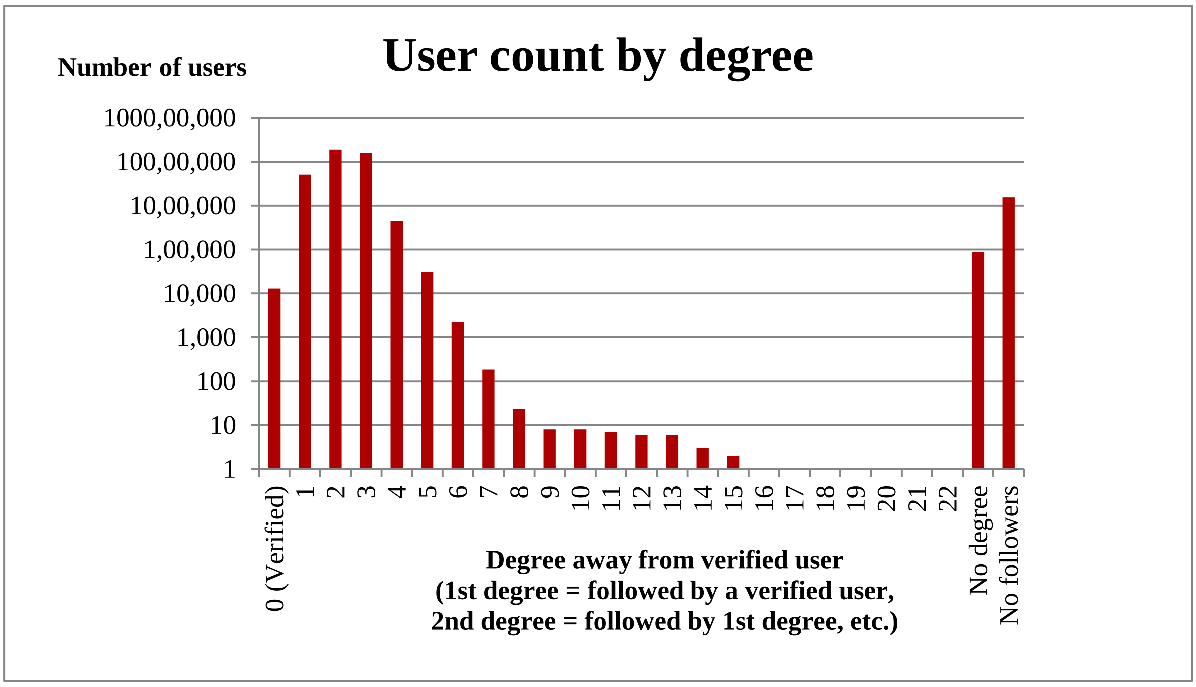


Figure 4: User count by degree (log scale)

Although degrees go up to 22 (that is, there is one degree 22 user, but no degree 23 users), 95% of all users are in degrees 1-3 and 1% are in degree 4. Verified users make up only 0.03% of all users. 3.7% have no followers at all and 0.2% have no degree (though they do have followers).

Additionally, since 12.2% of users are in degree 1 (have at least one verified follower), it appears that having *at least one* verified follower is not very rare, so it may not imply as much trust as we first thought. This finding prompted us to look for an alternative way to determine which unverified users can be trusted. Moreover, the degrees 1-3 contain 95% of the users and so it is difficult to estimate trust as most of the users are clustered into only three classes.

3.1.2.3 Degree of trust and influence

Next, we considered the relationship between a user's degree of trust and their influence, as measured by their total number of followers (Figure 5).

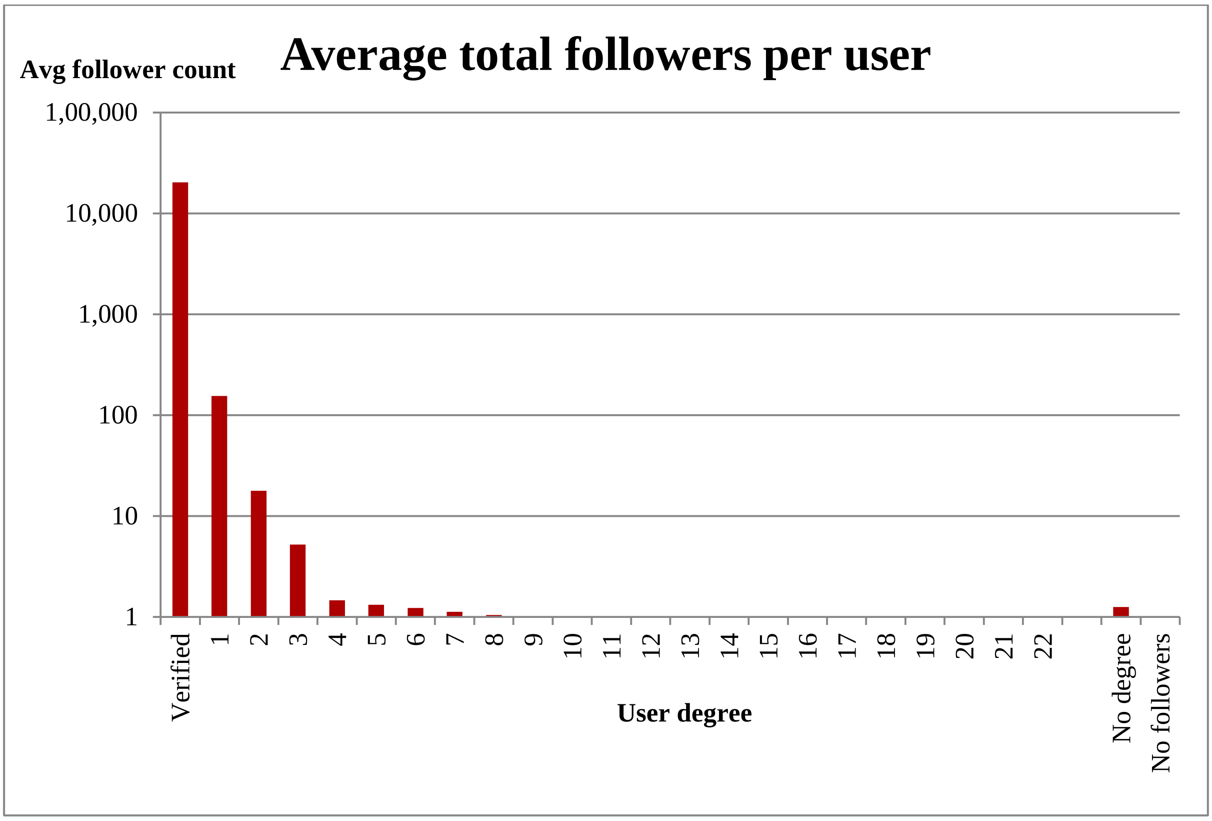


Figure 5: Average total followers per user by degree of trust (log scale)

There is an obvious correlation between the two: verified users have on average 20,842 followers each, degree 1 users have 155, degree 2 have 17.7, degree 3 have 5.2 and degree 4 onwards have around 1 follower each. Clearly, the “closer” someone is to a verified user, the more followers they tend to have. This is as expected: verified followers tend to be celebrities and therefore have a very high influence. If a user can get a verified user to follow them they can also get many others to follow them, but not nearly as many as it would be if they were a celebrity themselves. If the “best” follower they can get is degree 1, then they are even less likely to be influential, and so on. Even through this metric, the notion of trust is not very apparent.

3.1.2.4 Ratio of verified followers to total number of followers

We realised that judging a user's trust by taking minimum distance from verified users is not a good estimator of trust because users having one verified follower and ten verified followers are given the same degree of trust. Therefore we computed the ratio R_1 of verified followers to total number of followers for all users, in the hope of bringing in the factor of the number of verified followers.

$$R_1 = \frac{Number\ of\ verified\ followers}{Total\ number\ of\ followers}$$

We computed R_1 for the following classes of users: verified users, users of trust degree 1, experts, and other users. To compare, we used a CDF (see Figure 6). We expected a high value for the ratio for verified users.

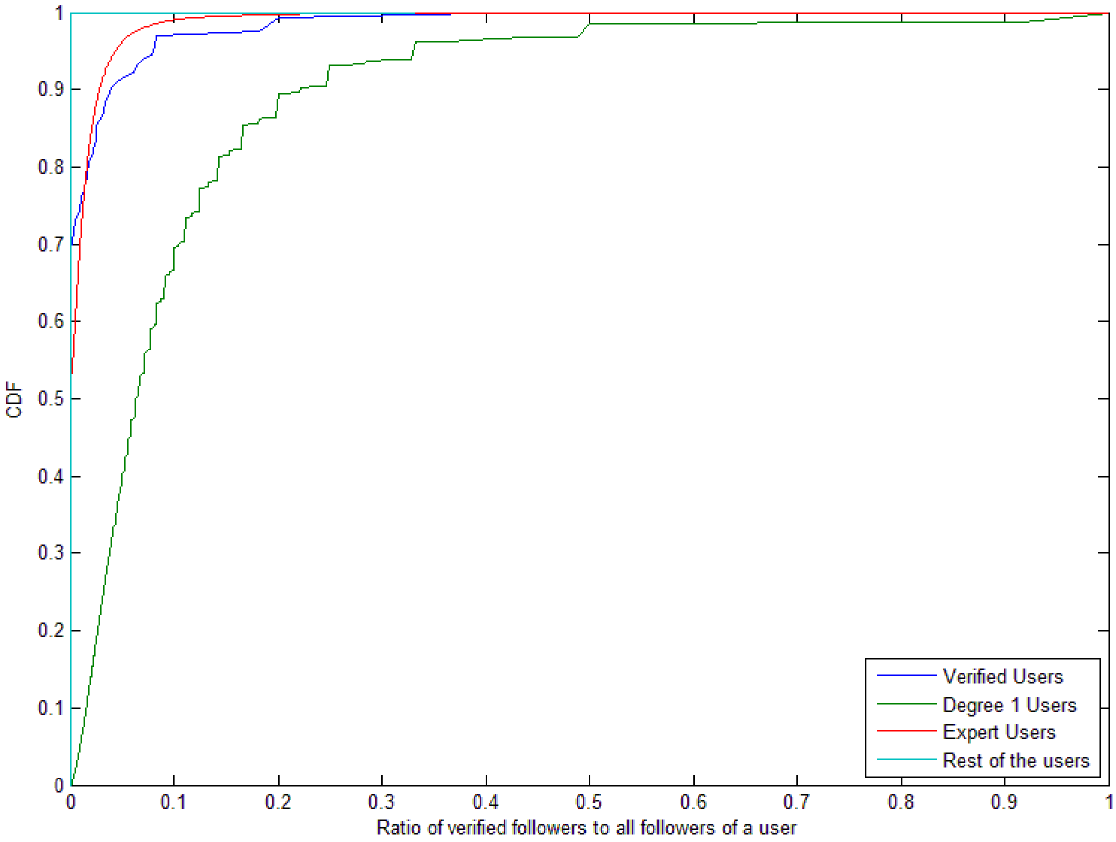


Figure 6: Ratio R_1 of the number of verified followers to the total number of followers for all users by category: verified users, users of trust degree 1, experts, and other users

The most striking feature of this plot is that it shows that users having trust degree 1 have a higher value than verified users. After investigation, we realised that this is observed since the ratio is normalised by the total number of followers. The verified users have a much higher number of followers than degree 1 users, and although the verified users also have a higher number of verified followers, the total followers is far too great and it drowns the number of verified follower count giving a higher value of the ratio to users of degree 1.

3.1.2.5 Ratio of the number of verified followers to total number of verified users in the system

As we saw that R_1 was unsatisfactory because of the normalising effect of the total number of followers in the denominator, we tried a second ratio R_2 which is defined as the ratio of the number of verified followers to total number of verified users in the system:

$$R_2 = \frac{Number\ of\ verified\ followers}{Total\ number\ of\ verified\ followers\ in\ Twitter}$$

We then plotted the CDF of R_2 for all users by category given in Figure 7.

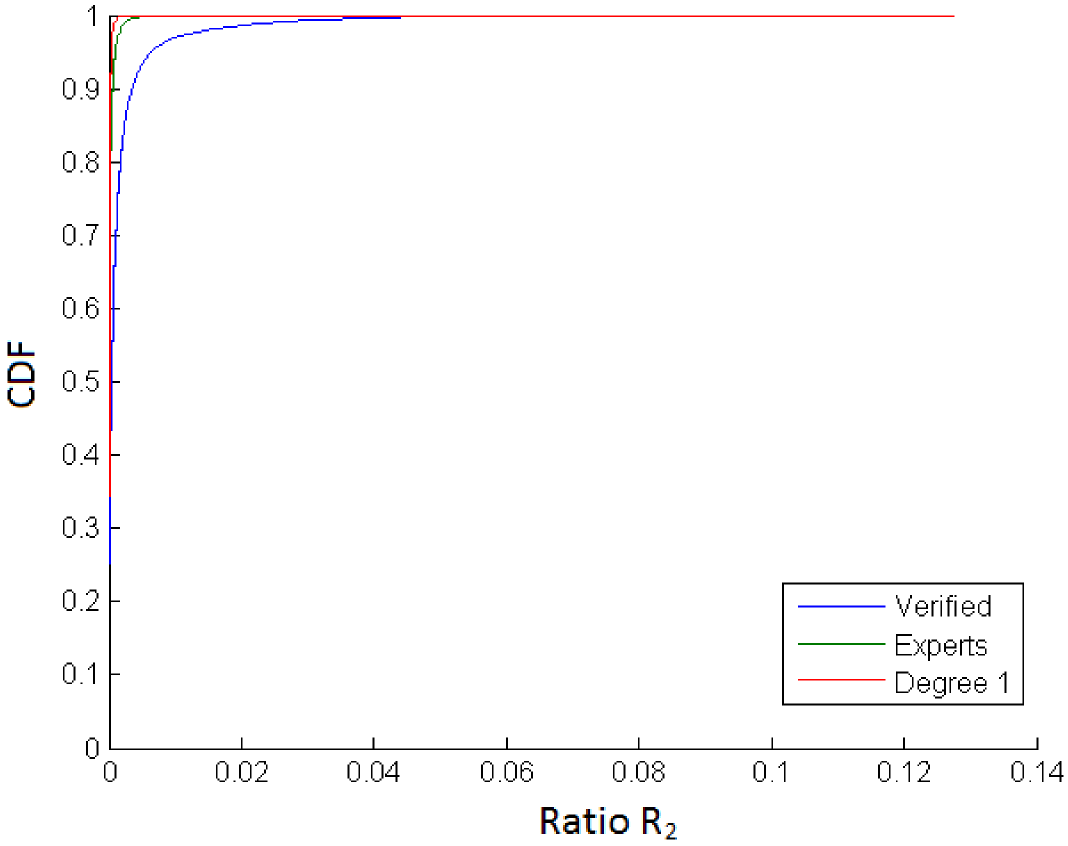


Figure 7: Ratio of the number of verified followers to the total number of verified users in the system for all users by category: verified users, users of trust degree 1, experts

R_2 upholds the notion of trust as verified users have a higher value than experts. But the problem with this ratio was that being a ratio, the value of R_2 was always less than 1 and it was difficult to find the different trust levels since R_2 was a continuous metric. We would rather prefer to have a discrete metric.

3.1.2.6 Number of verified followers

To overcome the problems of R_2, we used this intermediate step before we went on to calculate the trust score. Here we use the number of verified followers as an indicator of trust, much like the trust score. In this step we see how the number of verified users varies across unverified users. The aim was to investigate if there is any sharp change, which would denote two or more types of populations among the unverified users. However no such sharp change was noticed, as seen in Figure 8. Please note that the X-axis is in log scale and there is an exponential-like decay.

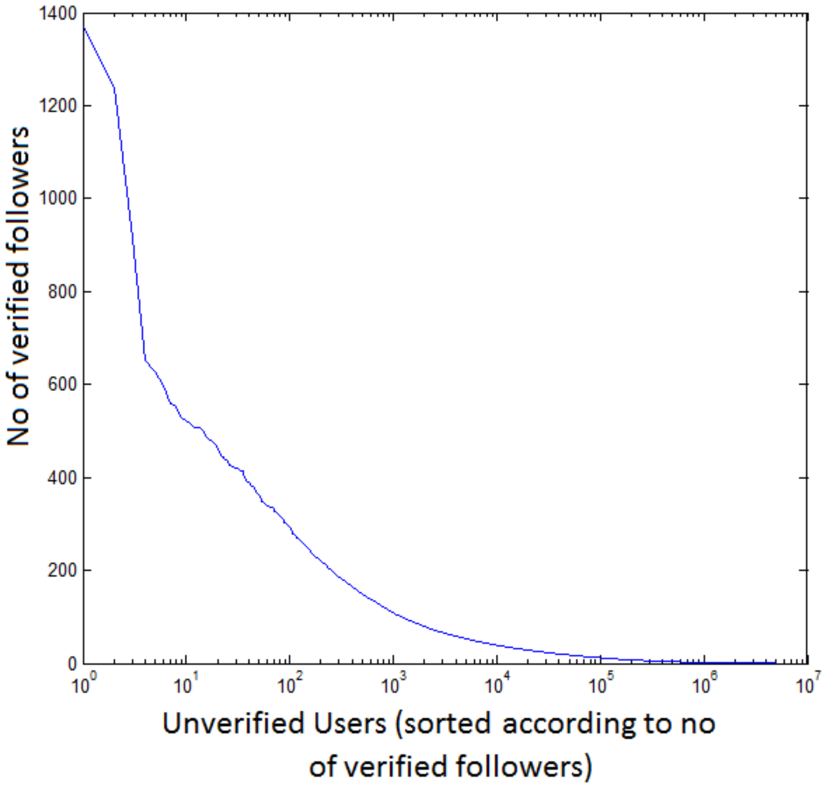


Figure 8: Number of verified followers for unverified users. X-axis is in log scale.

3.1.2.7 The trust score (2009 dataset)

With the above plots we are sure that the number of verified followers is a good metric to estimate trust because it should be able to differentiate different classes of users, in terms of trust. Also we are sure that the trust score falls more or less uniformly exponentially, and that shows that having a high trust score is rigorous and not everyone would get a high score. The trust score also has discrete levels of trust values (from 0 to 400) and that makes it easy to measure and compare. The trust score thus overcomes all of the disadvantages of the previously mentioned metrics, and can be used to measure trust.

We plotted a CDF of the score for the different classes of users in the 2009 data set in Figure 9. As expected, we see that verified users have a higher score than experts and experts have a higher score than random users. This puts faith in the trust score since it can uphold the different classes of users, in terms of trust. It also shows that verified users have the highest trust score, and that is so because verified users have been manually verified by Twitter and therefore automatically measured trust value of users cannot surpass the trust of verified users.

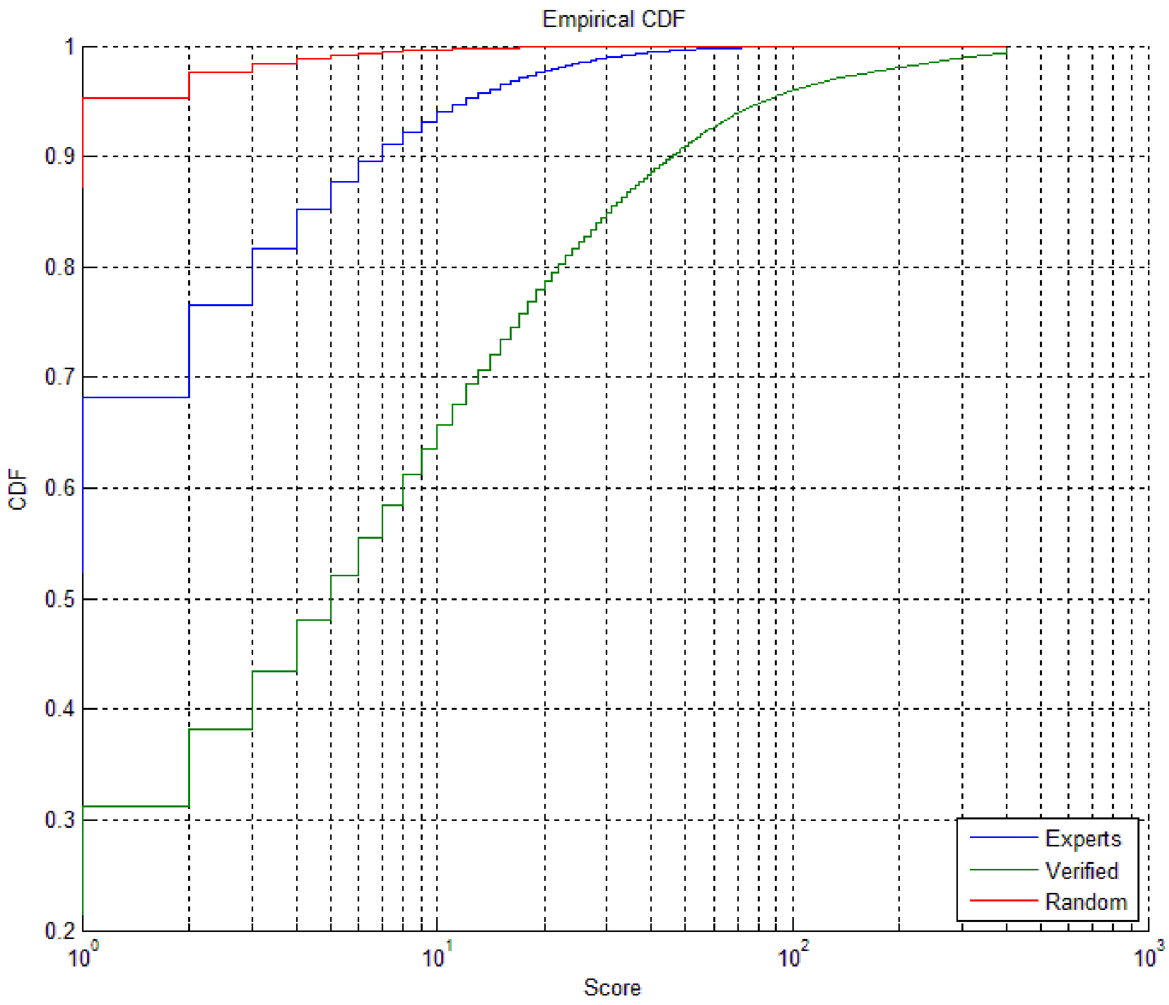


Figure 9: CDF of score (X-axis in log scale) for verified users, experts and randomly chosen users. This figure shows that the score of verified users is higher than that of experts, who in turn have higher score than randomly chosen users. This shows that the trust score is a good metric for measuring trust.

3.2 The 2012 dataset and the final analysis

The 2012 dataset was collected by M. Gabielkov et al [4] and it is ten times as large as the 2009 dataset with 505.4 million users. There are many reasons why the 2012 dataset is better than the 2009 dataset that we worked with in the first half of the project. We illustrate them below:

- The 2009 data set (along with other previous data sets) was not exhaustive/complete and there were subtle properties that were not visible in the 2009 dataset [4].
- There was a major change in the Twitter graph in 2009-2010, when a large number of celebrities joined Twitter and its popularity started increasing manifold. As mentioned before, celebrities are not interested in following too many people, but they have a high number of followers. This changed the properties of the Twitter graph in 2009-2010. These changes were not present in the 2009 dataset [4].
- There has been a tenfold increase in the size of Twitter since 2009. Twitter today is one of the most popular social networking sites on the internet [10].

The 2012 dataset has verified users, experts and other users like the 2009 dataset (all distinct categories) as depicted below.

Type of user	Quantity	Percentage
Verified	28,819	0.006%
Expert	2,889,063	0.572%
Other	502,492,882	99.422%
Total (all users)	505,410,764	100%

The goal was to inspect the trust metrics (mentioned in the previous section) on the 2012 dataset but because of its huge size, we did not work on it until we were sure what method we wanted to apply. After we deduced that the trust score is the best metric for measuring trust, we applied it on the 2012 dataset. But the raw graph (adjacency list) was too huge and required some data processing for making it efficiently accessible by our algorithms.

3.2.1 Preliminary data processing on the 2012 dataset

Unlike the 2009 data set, we had the adjacency list of the 2012 dataset from the beginning. This reduced the amount of data processing required but it was still not ready for running our algorithms. The adjacency list only included the users who had at least one follower, i.e., users with no followers were not in the list. Although it only accounted for 57.63% of the users (users with at least one follower), the size of the adjacency list was a staggering 214 GB. It was evidently far too big to work with, on our laptops. We parsed the adjacency list into a new file (called the properties file) with the user ID, number of followers of that user and the number of verified followers of that user. This was calculated while parsing the adjacency list, and it took almost 9 hours to complete on a regular laptop computer. The size of the properties file was about 4.12 GB. The properties file was further subdivided into other smaller files for verified users, experts and others. For the rest of the work on the 2012 data set, we used the properties file to do the computations. While working with the properties file, we loaded into memory (as a hash map) the most frequently accessed data (verified accounts and expert accounts). We then used Matlab for plotting the data.

3.2.2 The trust score (2012 dataset)

As with the 2009 dataset, trust score has discrete levels of trust values (from 0 to 400). We plotted a CDF of the score for the different classes of users in the 2009 data set in Figure 10. As expected, we see that verified users have a higher score than experts and experts have a higher score than randomly chosen users.

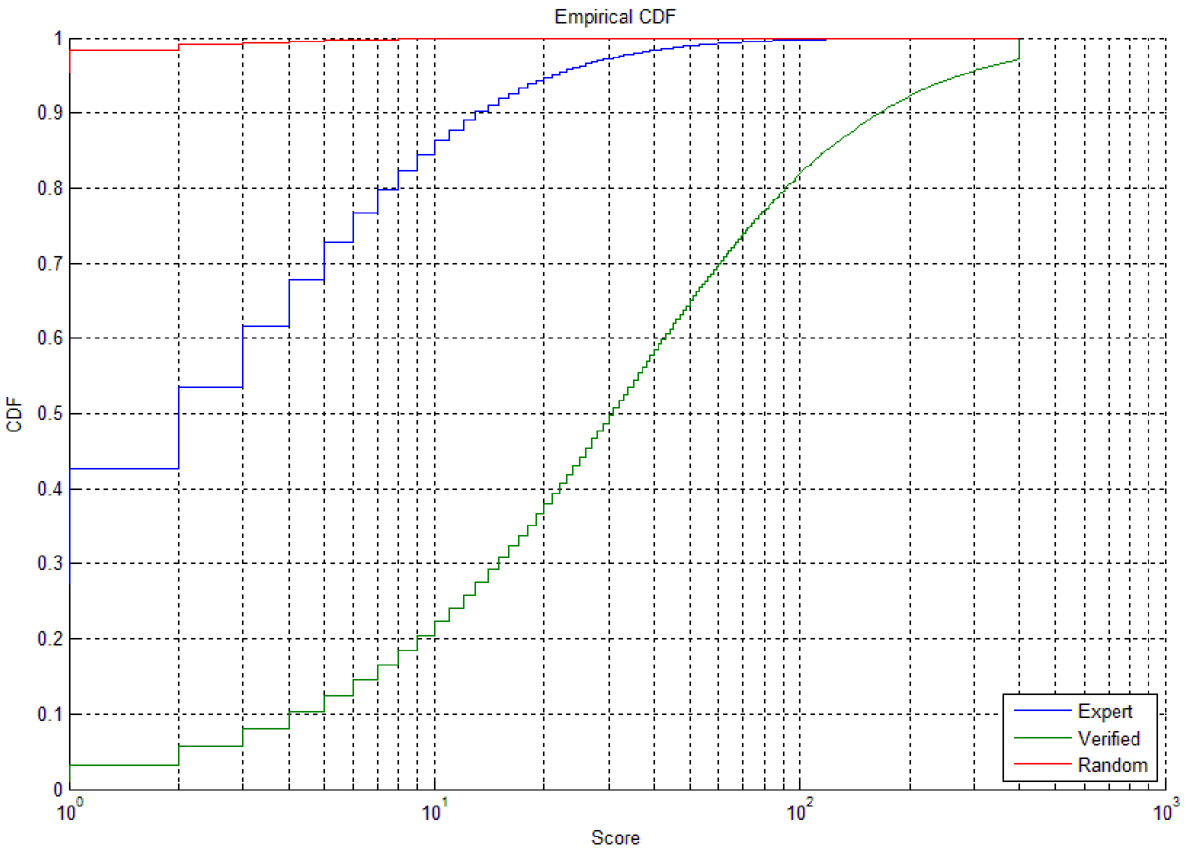


Figure 10: CDF of score (X-axis in log scale) for verified users, experts and randomly chosen users. This figure shows that much like the 2009 dataset, the score of verified users is higher than that of experts, who in turn have higher score than randomly chosen users. This shows that the trust score is a good metric for measuring trust.

4 Results: The trusted category of users

Now that we have seen that the trust score takes distinct values from 0 to 400, the next question is how to draw the line between who can be trusted and who can't. We know that the higher the trust score, the more trusted the person. But we still cannot categorise users into the fourth category of users (trusted users) explained in section 2.1. Above what threshold value of the score do we classify users to be "trusted"? This is a subjective question, and we try to answer that, by first showing how the number of unverified trusted users[1] varies with increasing threshold values (Figure 11) and then analysing the number of trusted users at some selected threshold levels. Because of the subjective nature of the question, there is no definite answer.

[1] See Figure 1 for a pictorial explanation: the trusted category comprises verified, experts and other users. Here we only take the unverified trusted users since the verified users are already considered to be trusted.

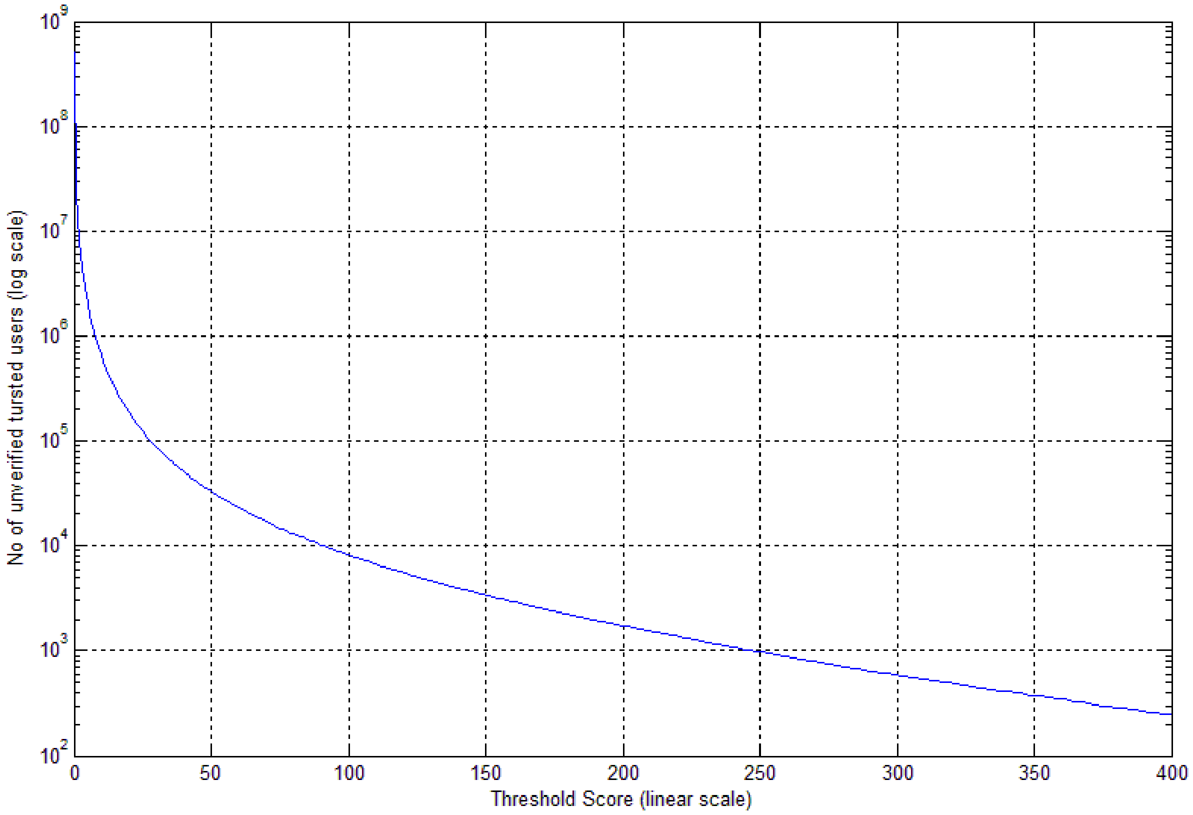


Figure 11: The fall in the number of trusted (unverified) users with increasing threshold (verified users are excluded because they are already considered to be trusted)

We have taken four sample threshold values to analyse the most optimal threshold value for trust. In the table that follows, we see that there are more than 8.2 million trusted users with a threshold of 2, which means that this threshold is not very stringent, and therefore is not the best choice. We see that more than 600 thousand users are trusted at a threshold of 10, and this seems to be a good value because it makes 0.132% of the users trusted, as compared to the 0.006% verified users. This threshold level is neither too restrictive nor stringent. If we take the threshold of 53, we obtain 29 thousand trusted users, which is almost equal to the 28.8 thousand verified users.

Threshold	Number of trusted (unverified) users	Percentage of total users
2	8,256,020	1.634%
5	2,076,566	0.411%
10	668,499	0.132%
53	29,404	0.006%

4.1 The new population

We have now seen how the threshold is used to divide the set of unverified users into trusted and not trusted users. But as shown in Figure 1, the interesting subset of users is the trusted set of users in the "other" category (defined in Section 2.1) of users. The trusted users in the "other" category are neither verified, nor are they experts, but are trusted by our method.

Therefore they constitute **a new population of trusted users that has not been considered before**. In this section we show how the size of this population changes with the threshold.

Threshold	Total number of trusted (unverified) users	Number of trusted users in the "other" category	Percentage of (unverified) trusted users in the "other" category[2]
2	8,256,020	6,596,586	80%
5	2,076,566	1,147,756	55%
10	668,499	221,282	33%
53	29,404	528	2%

The percentage of trusted users in the "other" category drops significantly as the threshold is increased. And this shows that experts are more trusted, which is why they are called influential as well. But at the threshold of 5, there are 1.1 million trusted users from the "other" category, which is quite a significant amount considering that there are 2.8 million expert users and only 28,819 verified users. The number of trusted users from the "other" category is more than one-third of the experts' category in size. Thus we have achieved our goal of finding a new population of significant size of trusted users on Twitter.

5 Evaluation of the trust score

We need to provide an evaluation mechanism to show how useful the trust score is. We provide two methods of evaluating the trust score as a metric. Firstly we see if the trust score can detect different categories of users, who we know are of different trust levels. Secondly we see if the trust score can be predicted by investigating other properties of the Twitter social graph.

5.1 How the trust score detects different categories of users

The CDFs in Figure 9 and Figure 10 have already shown us that the trust score can differentiate between categories of users. We expected to find verified users to have the highest trust, followed by experts and then randomly chosen users. This is exhibited by the CDFs in Figure 9 and Figure 10 and they are indicative of the correctness of the trust score as a metric for measuring trust. Since experts have been studied and have been found [8] to be influential on the Twitter graph, we would expect experts to be like the verified users since verified users are also influential. To explore this, we plotted the number of users in each discrete score value (see Figure 12). We see that the plot for experts is similar to that of the verified users – long tail and slower decay than other users. We also see that the tail of the experts' plot almost merges with the verified users'. From this plot we can infer that experts are more like verified users (albeit less trusted) than other users. Thus the trust score can clearly show similarities and dissimilarities between categories of users.

[2] The rest of the (unverified) trusted users are the experts who are considered to be trusted by the threshold value. See Figure 1 for a pictorial explanation: the trusted category comprises verified, experts and other users. Here we only take the unverified trusted users since the verified users are already considered to be trusted.

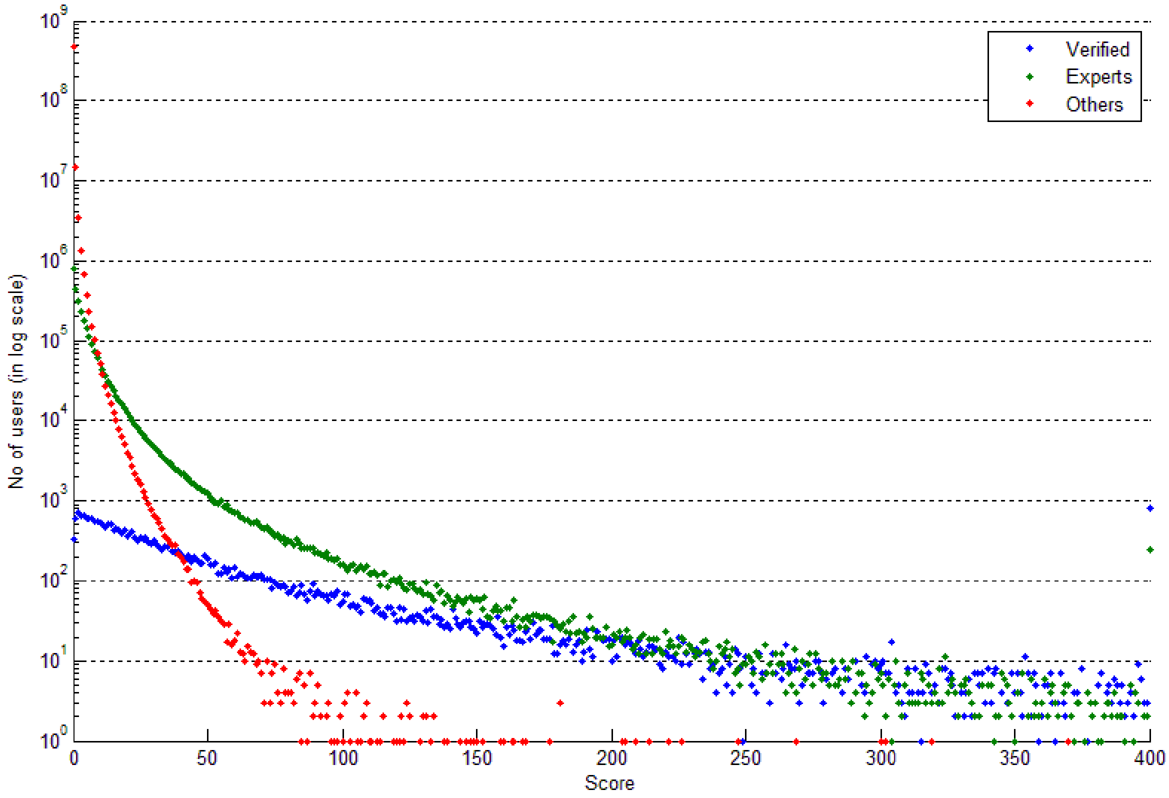


Figure 12: The number of users in each score value from 0 to 400 by user category. The verified users and the experts have a similar curve – long tail (which almost merges in the end) and slower decay than other users. This shows that, according to the trust score, experts are more like verified users than other users.

Derived from the above plot, we now show how the number of trusted users changes by increasing threshold for experts and other users (Figure 13). This was done to reinforce that experts are indeed more trusted than other users, thereby putting confidence in the trust score. Although the “other” users are much more in number, at any threshold level above 6, there are more trusted experts than “other” trusted users. This shows that the score upholds the property of experts being influential and having a large impact on Twitter, which is why experts are more trusted than other users.

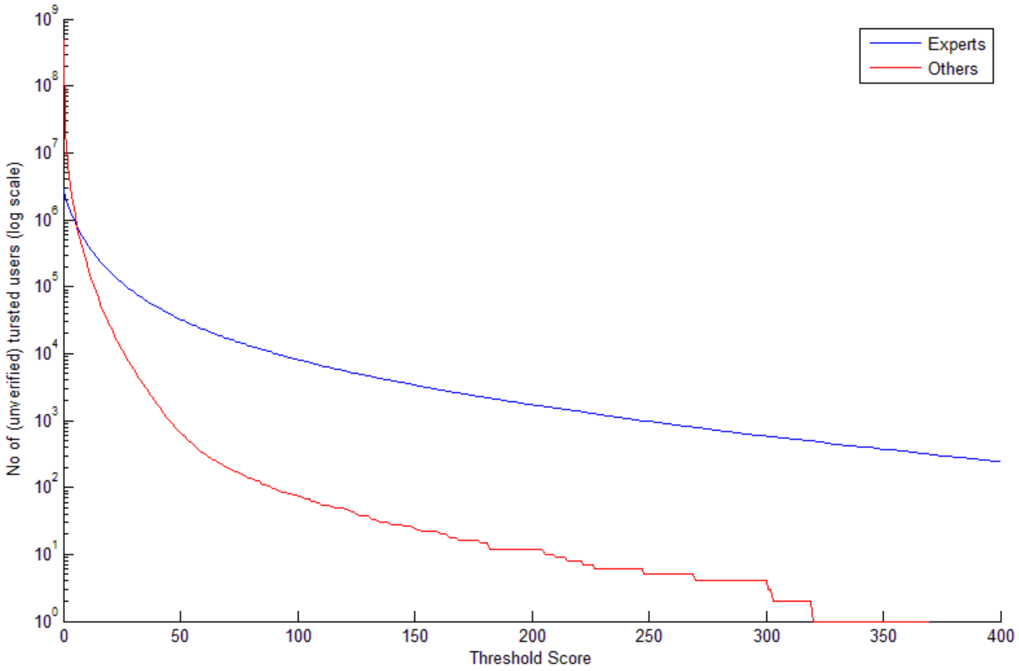


Figure 13: Although there are much more "other" users, at any threshold level above 6, there are more trusted experts than other users, and this shows that the score upholds the property of experts being influential and having a large impact on Twitter, and therefore are more trusted. Please note that verified users are not shown in this plot.

5.2 Correlation with other metrics

We will now investigate if the trust score uniquely predicts trust and whether what it predicts could be achieved by other metrics from the Twitter social graph. We investigate the correlation of the trust score with the total number of followers a user has (see Figure 14). We find that there is no correlation between the trust score and the number of followers. This proves that there is no dependency between trust score and the number of followers. Therefore, the trust score cannot be replaced by the number of followers to obtain the same result. It has been shown by Gabielkov et al [4] that metrics like the number of tweets and retweets are more erratic than the number of followers and therefore we expect no correlation with those metrics either. Therefore we conclude that the trust score is a unique metric which cannot be predicted by other metrics.

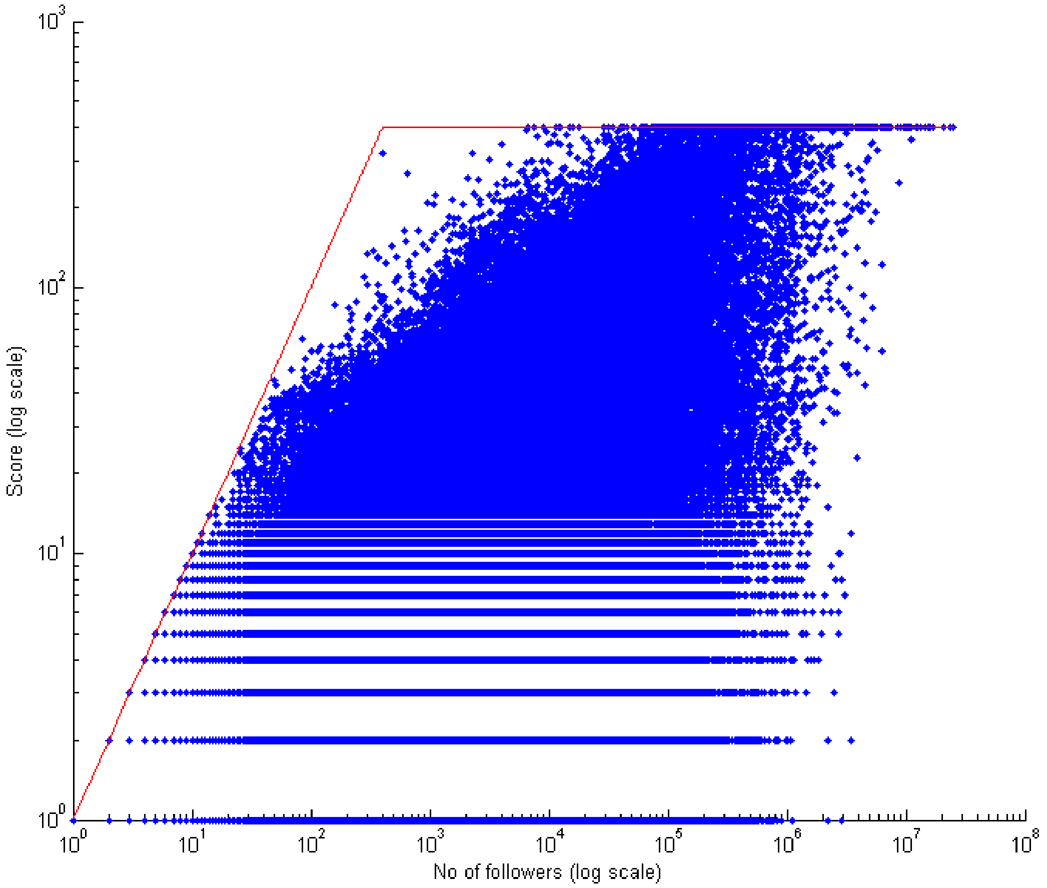


Figure 14: Scatter plot of the trust score vs the total number of followers of all users. There is no correlation, which means that the trust score cannot be predicted using the number of followers of a user. The red line shows the maximum score that a user with a given number of followers can have.

6 Challenges faced

There were some major challenges faced during the development of the project. Firstly, we had to handle a huge amount of data (around 214 GB) in order to analyse the entire Twitter graph. Since we did not have access to more powerful hardware at INRIA, we had to use our laptops with modest computing capabilities. We had to use algorithms that work by reading data sequentially from disk rather than randomly access data in memory. This constraint made the computations quite slow. Some computations required as long as 9 hours to complete. Although computations could be done on the disk, plotting the data using Matlab required the results to be in memory, which often made Matlab slow and unresponsive resulting from the large amount of data to plot. In some cases, we had to take random samples to reduce the size of the data to be plotted.

Secondly, we were searching for a new metric for measuring trust, and during the first half of the project we were unsure of the best metrics. We explored a lot of possible metrics (see section 3.1.2) but we did not know what the resulting best metric would be. As this is a research project, we were not sure if we would even be able to find a good metric for measuring the trust.

7 Conclusion and future work

The goal of this project was to establish a mechanism to determine who can be trusted on a social networking service like Twitter. We have shown that our method is to use the number of verified followers of a user as a metric in determining how trusted a user is. Our technique could be applied to any social network that permits directed arcs and has a set of verified users. We have also shown why it was our choice as a metric and that it can detect the different trust levels of different classes of users. We have also found a new set of users who have not been considered trusted by other metrics.

In the future, we would like to carry out further analyses with the trust score. The trust score can be further studied to explore the "chain of trust". The chain of trust is a phenomenon observed mostly in celebrities - most celebrities follow back only a small subset of their followers. We would like to see whether the trust score can identify this subset. As part of potential future work, we would like to study the new population of trusted users that we found. We would like to see if this population has any special characteristics in their tweeting/retweeting habits, and/or if their profiles are any different from the rest of the "others" group. To sum up, we want to keep working with the trust score to better explain how trust varies across the 500 million users on Twitter.

In conclusion, we would like to thank Prof Arnaud Legout and Maksym Gabielkov for their constant support, guidance and patience during the development of the project.

References

[1] *Wikipedia article on Twitter*. http://en.wikipedia.org/wiki/Twitter

[2] *About Twitter*. https://about.twitter.com/company

[3] H. Kwak, C. Lee, H. Park, and S. Moon. *What is Twitter, a social network or a news media?* In Proc. Of WWW'10, Raleigh, NC, USA, May 2010.

[4] Maksym Gabielkov, Ashwin Rao, and Arnaud Legout. Studying Social Networks at Scale: *Macroscopic Anatomy of the Twitter Social Graph*. To appear in Proc. of ACM SIGMETRICS'14, June 16--20, 2014, Austin, Texas, USA.

[5] Mashable.com: *Eight social media hoaxes you fell for this year*. http://mashable.com/2012/11/05/social-media-hoaxes/

[6] *FAQs about verified accounts*. https://support.twitter.com/articles/119135-faqs-about-verified-accounts

[7] *Can becoming a twitter 'verified user' help your business?* http://www.forbes.com/sites/drewhendricks/2013/09/25/can-becoming-a-twitter-verified-user-help-your-business/

[8] N. Sharma, S. Ghosh, F. Benevenuto, N. Ganguly, K. P. Gummadi. *Inferring Who-is-Who in the Twitter Social Network*. 4th ACM SIGCOMM Workshop On Social Networks (WOSN), Helsinki, Finland, August 2012.

[9] *The Least Popular Verified Twitter Accounts (August 2012).* http://www.buzzfeed.com/katienotopoulos/the-least-popular-verified-twitter-accounts

[10] Mashable.com. *Facebook Is Most Popular Social Network for All Ages; LinkedIn Is Second [STUDY]*. http://mashable.com/2011/11/04/facebook-most-popular-forrester/