

# The Usefulness of Twitter for Open Source Developers as A Feedback Tool for The Success of Their Projects

Ivor Van Der Schalk, Zaki Abdurrahman Koesoemahardja, and Slinger Jansen

Department of Computing and Information Sciences  
Utrecht University

{i.l.vanderschalk,z.a.koesoemahardja,slinger.jansen}@uu.nl

**Abstract.** This paper analyzes correlations between success indicators of Open Source projects and Twitter posts containing emotional signals on such projects. Within a timeframe of two years (February 1st 2014 - January 31st 2016), 61,570 Twitter posts containing names of Open Source projects were collected. These posts were classified into positive or negative signals based on their content. For instance, posts that contain the terms "happy", "love", "fun", "good", "bad", "sad" and "unhappy" represent positive or negative emotional signals. The purpose of this research is to find whether or not people express their feelings about Open Source projects on Twitter and to determine if these feelings are an indication of how successful an Open Source project is. This is interesting because this enables Open Source developers to use Twitter to get reliable feedback from their users about their projects. Among the aspects that are explored in this research are the number of tweets containing positive signals, number of tweets containing negative signals and the number of downloads of Open Source projects.

**Keywords:** Twitter, Sentiment Classification, Open Source Project

## 1 Introduction

In recent years, there has been a growth of interest among developers in Open Source projects. Developers see that Open Source projects have economic benefits. Another reason for developers to work on Open Source projects is because it gives a good learning opportunity (Lee, Kim & Gupta, 2009). Crowston, Annabi, and Howison (2003) state that the term of Open Source project is used to cover a project that is developed under some sort of "Open Source" license. This means that the code in an Open Source project can be reused in other Open Source projects.

These days, the way that people communicate, collaborate, and share information has been changed by Social Media. Social Media help users to make new links, facilitate discussion, and maintain relationships (Tsay, Dabbish & Herbsleb, 2012). Social Media has been used by Open Source project developers

to improve tools in their development environments. Additionally, Open Source project developers use a variety of Social Media tools to communicate with other developers, to have information about new technologies, to learn from users, and to communicate with users. The use of Social Media is not surprising because of the change in paradigm on how users communicate and work on the Internet. Twitter is one example of a Social Media application that is used by Open Source developers (Storey et al., 2010).

In other research, Twitter has been used to analyse sentiments expressed in tweets to find a relation with the Bitcoin price (Kaminski & Gloor, 2014). In their research they concluded that Twitter acts as a "mirror" for the Bitcoin price. This makes one wonder if Twitter can also be used as a "mirror" for the success of an Open Source project.

In brief, the goal of this research is to find out whether or not Open Source users express their feelings about Open Source projects on Twitter. This is interesting because if this is the case then Open Source developers could use Twitter as a reliable source to get feedback from their users. So far little to no research has been done in this area. Jansen, Finkelstein and Brinkkemper (2009) explain that there are four types of software ecosystems: Market, Technology, Platform, and Firm. Our research can be put in the third type of software ecosystems. The Open Source projects we investigate are part of the Sourceforge platform. This is a platform that offers Open Source developers the opportunity to collaborate and share code with each other. The main research question of this research is presented as follows: "Can Twitter be used as a feedback tool by Open Source developers to determine the success of their projects?".

The remainder of the paper is structured as follows. In the following section we discuss the research method and provide insights about the process of the data collection. Section 3 contains the analysis and results on the collected data. In section 4 the discussion is given and we conclude this paper with the conclusion in section 5.

## 2 Research Method

In this section the research method is given. The first part of the research method consists of a literature study to discover the characteristics of the success of an Open Source project. The literature study is also conducted to find out how to classify Twitter posts into positive and negative signals. The second part of the research method describes the process of collecting the data related to Open Source projects and Twitter.

### 2.1 Literature Study

In order to explain the context of this research, a literature review about Twitter signals and the success of Open Source projects is carried out. Twitter is a Social Media and microblog that enables users to post short messages that are restricted to 140 characters. Due to the short messages in Twitter, people like

to use acronyms, emoticons, and other characters that show special meanings (Agarwal et al., 2011). Zhang, Fuehres and Gloor (2010) state that the main topic in Twitter can be concluded from one or two keywords because of the quick and short messages. Agarwal et al., (2011) use an emoticon dictionary and acronym dictionary (table 1) to classify tweets into positive or negative signals. Tweets that have positive emoticons like ":)" or ":-)" are defined as positive tweets and tweets that have negative emoticons like ":(" or ":-(" are defined as negative tweets. Kaminski and Gloor (2014) use emotional signals like "happy", "love", "fun", "good", "bad", "sad" and "unhappy" (table 2) to classify tweets into positive or negative signals.

From the survey of Black, Harrison, and Baldwin (2010) about Social Media use in software system development, they found that 91 percent of 31 respondents used Social Media to communicate with their colleagues. Twitter and Instant Messaging were found to be the most popular media. Social Media was also used by the respondents to share new ideas, source code, specification, and design information.

**Table 1.** Positive and negative emoticon signals

Emoticons	Signals
:) :-) :o) :] :3 :c) :D C:	Positive
:( :-(: :c :[ D8 D; D= DX V.V	Negative

**Table 2.** Positive and negative emotional signals

Emotional Words	Signals
Feel, happy, great, love, awesome, lucky, good	Positive
Sad, bad, upset, unhappy, nervous	Negative

An Open Source project is an example of user-driven innovation. Such a project starts from a need of an individual or group of developers to have new or additional application functionalities. An Open Source project is successful when it is used by a large number of users or when large number of developers work together in the project (Comino, Manenti & Parisi, 2007). It is important to measure the success of Open Source projects because it can be useful for the developers in evaluating their projects. Crowston, Annabi, and Howison (2003) show seven indicators that can be used to measure the success of an Open Source project: system and information quality, user satisfaction, use, individual and organizational impacts, project output, process, and outcomes for project members. In this research we use two indicators that are shown in Table 3. The number of downloads is available as public data for every Open Source project that is active on Sourceforge. The user ratings are also available for every Open

Source project on Sourceforge, but usually only the more popular Open Source projects tend to have a decent number of user ratings available.

**Table 3.** The indicators that are used to measure the success.

Measure of Success	Indicators
User satisfaction	User ratings
Use	Number of downloads

## 2.2 Data Gathering

In this section it is explained how the data for this research is collected. The tool Chrome Web scraper is used to scrape the number of downloads and user ratings of Open Source projects from Sourceforge.

To scrape the tweets of Open Source projects from Twitter and classify them as positive or negative signals we use a self developed web scraping tool. A tweet is classified as a positive signal when it contains one of the following terms: ":", ":-)", ":o)", ":]", ":3", ":c)", ":D", "C:", "Feel", "happy", "great", "love", "awesome", "lucky", "good". A tweet is classified as a negative signal when it contains one of the following characters: ":((", ":-(", ":c", ":[", "D8", "D;", "D=", "DX", "V.V", "sad", "bad", "upset", "unhappy", "nervous".



**Fig. 1.** Twitter signals example.

An example of three tweets about the Open Source project 'Filezilla' can be found in Figure 1. The first tweet contains the term "good" and can therefore be classified as a positive signal. The second tweet contains the term "bad" and can therefore be classified as a negative signal.

The scraper is built with PHP code by using a PHP DOM inspector library for accessing the content of HTML elements. The input of the scraper is the

source of a Twitter page containing all the tweets in a given period and a given keyword. The scraper extracts the content of every tweet. Once all the content is extracted the scraper checks for each tweet if it contains positive or negative terms. The output of the scraper is a list of tweets in which is indicated whether the tweet is a positive or negative signal.

Below is a short description of the data sources from which data is collected for the statistical analysis of this research:

1. Sourceforge (<http://sourceforge.net>) freely provides detailed about Open Source project statistic:
  - Number of downloads
  - User ratings
2. Twitter (<https://twitter.com/>) shows the tweets for each Open Source project.

During this research data is collected of three Open Source projects that are available in Sourceforge. For the statistical analysis it is important there is a sufficient number of downloads and number of tweets available. A project is therefore selected based on the number of downloads and number of tweets containing the name of the project in Twitter. Based on this criteria FileZilla, WampServer and OpenOffice.org are selected for this research. Table 4 contains a description of the three selected Open Source projects.

**Table 4.** Open Source projects description

Open Source Project	Description
FileZilla	FileZilla is a cross-platform graphical FTP, SFTP, and FTPS file management tool. It helps users quickly move files between computer and web server.
WampServer	WampServer is a web development platform on Windows that allows users to create dynamic web applications.
OpenOffice.org	OpenOffice is an Open Source office productivity software suite. It is a successor project of OpenOffice.org.

Within a timeframe of two years (January 1st 2014 - December 31st 2015), the number of downloads and the user ratings (5 stars rating scale) of each Open Source project was scraped from Sourceforge. After collecting this data, the average user ratings were calculated per month for each project.

In addition to that, for each Open Source project all the tweets containing the name of the Open Source project were scraped from Twitter between February the 1st 2014 and January the 31st 2016. Finally, the tweets were classified into positive and negative signals.

In order to ensure that the data was gathered between the selected date ranges, we used date filters that are provided by Sourceforge and Twitter (Twitter Advanced Search). The average user rating per month was calculated by

counting the total number of stars per month divided by the total number of user reviews per month.

### 2.3 Data Analysis

The analysis is performed by using SPSS. Table 5 summarizes the data set of the three Open Source projects of the year 2014 and 2015 collected from Sourceforge and Twitter.

**Table 5.** Overview of the data of Filezilla, Wampserver and OpenOffice.

Year	Project	Downloads	Average User Ratings	Number of Tweets	Positive Tweets	Negative Tweets
2014	FileZilla	56,709,814	2.2	8,354	503	66
	Wampserver	5,684,480	2	1,253	77	7
	OpenOffice	48,764,619	4.1	20,818	964	172
2015	FileZilla	101,130,524	3.5	9,691	375	88
	Wampserver	5,316,543	2.2	1,130	38	7
	OpenOffice	40,611,428	4	20,324	901	133

For each Open Source project the Pearson correlation has been computed between the number of downloads, number of tweets, number of tweets containing positive signals, number of tweets containing negative signals and the average rating of each month.

It should be noted that the Pearson correlation was computed between the number of downloads in one month (for example January) and the number of tweets, number of tweets containing positive signals, number of tweets containing negative signals and the average user rating in the next month (for example February). The reason behind this can be found in the fact that this will help to find what the effect is of the number of downloads on the other variables. The following section presents the results for each Open Source project.

## 3 Results

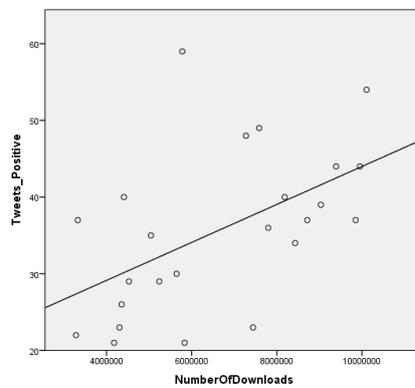
### 3.1 Filezilla Results

This section presents the results of the data related to Filezilla. Figure 2 present the computation of the Pearson correlation between the variables. There is a significant positive correlation between Filezilla's total number of tweets containing positive signals and the number of downloads,  $r = .525$ ,  $p = < .001$ . A visualization of this correlation can be found in Figure 3. There is also a significant correlation,  $r = .719$ ,  $p = < .001$ , between the number of tweets and number of tweets containing negative signals.

		NumberOfDownloads	NumberOfTweets	Tweets_Positive	Tweets_Negative	Average_Rating
NumberOfDownloads	Pearson Correlation	1	.056	.525**	.121	-.340
	Sig. (2-tailed)		.796	.008	.574	.103
	N	24	24	24	24	24
NumberOfTweets	Pearson Correlation	.056	1	.240	.719**	.024
	Sig. (2-tailed)	.796		.258	.000	.910
	N	24	24	24	24	24
Tweets_Positive	Pearson Correlation	.525**	.240	1	.241	-.172
	Sig. (2-tailed)	.008	.258		.257	.421
	N	24	24	24	24	24
Tweets_Negative	Pearson Correlation	.121	.719**	.241	1	-.034
	Sig. (2-tailed)	.574	.000	.257		.876
	N	24	24	24	24	24
Average_Rating	Pearson Correlation	-.340	.024	-.172	-.034	1
	Sig. (2-tailed)	.103	.910	.421	.876	
	N	24	24	24	24	24

\*\* . Correlation is significant at the 0.01 level (2-tailed).

**Fig. 2.** SPSS output of the Pearson correlation between Filezilla’s number of downloads, number of tweets, number of tweets containing positive signals, number of tweets containing negative signals and the average user rating of Filezilla.



**Fig. 3.** A visualization of the correlation between Filezilla’s number of tweets containing positive signals and number of downloads per month.

### 3.2 Wampserver Results

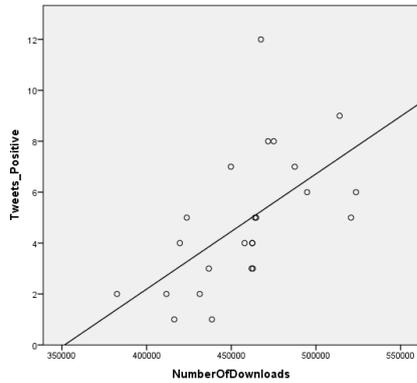
This section presents the analysis and results of the data related to Wampserver. It should be noted that because there was an insufficient number of tweets containing negative signals and user ratings of Wampserver it is not possible to perform a statistical analysis on these variables.

Figure 4 present the computation of the Pearson correlation between the Wampserver’s number of downloads, number of tweets and number of tweets containing positive signals. A significant positive correlation was found between the number of tweets containing positive signals and the number of downloads,  $r = .587$ ,  $p = < .001$ . A visualization of this correlation can be found in Figure 5.

		NumberOfDownloads	NumberOfTweets	Tweets_Positive
NumberOfDownloads	Pearson Correlation	1	.172	.587**
	Sig. (2-tailed)		.422	.003
	N	24	24	24
NumberOfTweets	Pearson Correlation	.172	1	.614**
	Sig. (2-tailed)	.422		.001
	N	24	24	24
Tweets_Positive	Pearson Correlation	.587**	.614**	1
	Sig. (2-tailed)	.003	.001	
	N	24	24	24

\*\* . Correlation is significant at the 0.01 level (2-tailed).

**Fig. 4.** SPSS output of the Pearson correlation between Wampserver’s number of downloads, number of tweets and number of tweets containing positive signals.



**Fig. 5.** A visualization of the correlation between Wampserver’s number of tweets containing positive signals and the number of downloads per month.

### 3.3 OpenOffice Results

This section presents the analysis and results of the data related to OpenOffice. Figure 6 present the computation of the Pearson correlation between the variables. No significant correlation was found between the number of downloads and the other variables.

		NumberOfDownloads	NumberOfTweets	Tweets_Positive	Tweets_Negative	Average_Rating
NumberOfDownloads	Pearson Correlation	1	.281	.112	.047	-.289
	Sig. (2-tailed)		.183	.602	.827	.172
	N	24	24	24	24	24
NumberOfTweets	Pearson Correlation	.281	1	.456*	.220	-.149
	Sig. (2-tailed)	.183		.025	.302	.486
	N	24	24	24	24	24
Tweets_Positive	Pearson Correlation	.112	.456*	1	-.244	-.111
	Sig. (2-tailed)	.602	.025		.250	.607
	N	24	24	24	24	24
Tweets_Negative	Pearson Correlation	.047	.220	-.244	1	-.095
	Sig. (2-tailed)	.827	.302	.250		.659
	N	24	24	24	24	24
Average_Rating	Pearson Correlation	-.289	-.149	-.111	-.095	1
	Sig. (2-tailed)	.172	.486	.607	.659	
	N	24	24	24	24	24

\*. Correlation is significant at the 0.05 level (2-tailed).

**Fig. 6.** SPSS output of the Pearson correlation between OpenOffice's number of downloads, number of tweets, number of tweets containing positive signals, number of tweets containing negative signals and the average user rating.

## 4 Discussion

In this research it was investigated whether Twitter can be used as a feedback tool by Open Source developers to determine the success of their projects.

We found no significant correlation between number of downloads of Open Source projects and number of tweets. This indicates when people download an Open Source project this does not necessarily lead to more tweets. The reason for this could be that only people who are already active on Twitter post tweets. When the majority of people who download an Open Source project are not Twitter users it makes sense the number of tweets does not increase.

For the Open Source projects Filezilla and Wampserver a significant correlation was found between the number of downloads and number tweets containing positive signals. This indicates that Twitter users use Twitter to express their opinion when it is positive about an Open Source project. Since there is no significant correlation found between the number of tweets and number of downloads we can conclude that users change their opinion in tweets from non-positive to positive to at least some degree. However, for the Open Source project OpenOffice no significant correlation was found between the number of downloads and number of tweets containing positive signals. This indicates that only for some Open Source projects Twitter can be used for positive feedback. The reason for this may be found in the type of user. Filezilla and Wampserver are typically used by users with a technical background whereas OpenOffice is also used by

non-technical people. The type of user may affect how Twitter is used to express sentiments.

There was no significant correlation found between the number of downloads of Open Source projects and number of tweets containing negative signals. This indicates that Twitter is not frequently used by Open Source project users to express negative opinions through twitter posts. A reason for this could be that users who have a positive view use an Open Source project longer and therefore have a longer period in which they can express their positive view. In contrary, users who have a negative view use an Open Source project for a shorter period and therefore have a shorter period in which they can express their negative view. This results in less negative tweets and may also explain why Open Source projects in general have more tweets containing positive signals than tweets containing negative signals.

One would expect that as the average user rating increases the number of tweets containing positive signals also increases, the same logic can be applied for tweets containing negative signals. However, when looking at the average user ratings of Open Source projects and the number of tweets, number of tweets containing positive signals and number of tweets containing negative tweets no significant correlations were found. This may indicate that the average user rating on Sourceforge has different criteria than tweets containing positive or negative signals. The reason may also be found in the fact that the user rating is a relative variable. The user rating indicates something is either really good, really bad or anything between it. In contrary to user ratings, tweets containing positive or negative signals are absolute. A tweet either indicates the Open Source project is good or bad, there is no middle ground.

## 5 Conclusion

The main goal of this research was to find whether Twitter could be used as a tool to receive reliable feedback of Open Source projects. This has been investigated by performing a statistical analysis on indicators that determine the success of an Open Source project and Twitter signals. The data set of the research contained data about the number of downloads, average user rating and tweets of three of the most popular Open Source projects on Sourceforge.

The results show that two of the three Open Source projects have a significant correlation between the number of downloads and number of tweets containing positive signals. There was no significant correlation found in any of the Open Source projects between the number of downloads and the number of tweets containing negative signals. There was also no significant correlation found between the average user rating and Twitter signals.

Based on the results it can be concluded that Twitter can be used to some degree by Open Source developers to receive positive feedback about their projects. However, more research is necessary to find out for what kind of Open Source projects Twitter can be used as a reliable feedback tool.

In this research there were some limitations that could be addressed in future research. In regards to the data, for a period of two years data was collected from Sourceforge and Twitter of three Open Source projects. In future research, more Open Source projects can be examined over a longer period of time and from other sources such as Github to get more accurate results. In addition to that, in this research only a limited number of indicators were used that are related to the success of an Open Source project. In future research more success indicators can be used to find whether Twitter can be used as a reliable feedback tool.

In this research, to classify Twitter posts into positive and negative signals only English emotional words were used that are presented by Kaminski and Gloor (2014), as a result we excluded non-English. In future research non-English tweets can be included.

In future research one can also look beyond Twitter to find whether other Social Media such as Facebook or Instagram can be used as a reliable feedback source for Open Source projects.

## References

1. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.: Sentiment analysis of twitter data. In Proceedings of the ACL 2011 Workshop on Languages in Social Media, pp. 30–38. (2011)
2. Comino, S., F.M. Manenti, & M.L. Parisi.: From planning to mature: On the success of Open Source projects. *Research Policy*, 36(10), pp. 1575-1586. (2007)
3. Crowston, K., H. Annabi and J. Howison.: Defining Open Source software project success. In Proceedings of Twenty-Fourth International Conference on Information Systems. Seattle, WA, pp. 327-340. (2003)
4. Jansen, S., Finkelstein, A., & Brinkkemper, S.: Business network management as a survival strategy: A tale of two software ecosystems. In Proceedings of the First Workshop on Software Ecosystems. CEUR-WS, vol. 505. (2009)
5. Kaminski, J. and Gloor, P.: Nowcasting the bitcoin market with twitter signals. arXivpreprintarXiv:1406.7577. (2014)
6. Lee, S.T. Kim, H. and Gupta, S.: Measuring Open Source software success. *Omega*, vol. 37, pp. 426 – 438. (2009)
7. Storey, M., Treude, C., van Deursen, A., & Cheng, L.T.: The impact of Social Media on software engineering practices and tools. In Proceedings of the FoSER 2010. ACM Press, pp. 359-363. (2010)
8. Tsay, J. T, Dabbish, L., & Herbsleb, J.: Social Media and success in Open Source projects. In Proceedings of the ACM 2012 conference on computer supported cooperative work companion. ACM, pp. 223–226. (2012)
9. Zhang, X., Fuehres, H., & Gloor, P.: Predicting stock market indicators through Twitter – “I hope it is not as bad as I fear”, In Collaborative Innovations Networks Conference, Savannah, GA, pp. 1-8. (2010)
10. Black, S., Harrison, R., & Baldwin, M.: A survey of Social Media use in software systems development. In Proceedings of the 1st Workshop on Web 2.0 for Software Engineering. ACM Press, pp. 1-5. ACM.(2010)
11. Sourceforge, <http://sourceforge.net>
12. Twitter, <https://twitter.com/search-advanced?lang=en>