Microblogging in Open Source Software Development: The Case of Drupal Using Twitter

Xiaofeng Wang¹, Ilona Kuzmickaja², Klaas-Jan Stol³, Pekka Abrahamsson¹, and Brian Fitzgerald³

¹Free University of Bozen-Bolzano, Italy ²Facebook Inc., Ireland ³Lero – The Irish Software Engineering Research Centre, University of Limerick, Ireland

Abstract. Microblogging is one of the popular forms of social media that has quickly permeated both enterprise and open source communities. However, how exactly open source communities can leverage microblogging is not yet well understood. We investigate how Drupal’s open source community uses Twitter, a household-name in microblogging. Our analysis of group and individual accounts of Drupal developers reveals that they take on similar but distinct roles. Both serve as communicators of essential links to a vast and growing community knowledge base, such as work artifacts, issues, documentation, and blog posts. Community members often express positive emotions when tweeting about work, which reinforces a sense of community. Finally, Twitter is also used as a crowdsourcing channel to solicit contributions.

Keywords: Twitter, Microblogging, Social Media, Open Source Software Development, Drupal, Communication, Crowdsourcing

Introduction

Social media have become an important means of communication in modern society, evidenced by the widespread use of wikis, blogs, social networking sites, and more recently microblogging, for both personal use and professional use within organizations. Among the different social media a significant one is microblogging [1]. Originally intended to provide brief personal text updates, it has rapidly expanded to play a more informative and interactive role in communication and collaboration across a wide variety of organizations and endeavors, including software development which is inherently a socio-technical activity [2], [3]. Most studies of microblogging have focused on enterprise settings. One such study revealed various types of interactions, including asking questions, sharing information, coordinating and broadcasting updates [4]. The prospect of integrating microblogging tools into software development environments also seems promising (e.g., [5], [6]).

Microblogging is used extensively in open source software (OSS) communities, but so far this hasn’t been investigated in-depth [7]. We studied how microblogging is utilized to benefit OSS development. This will provide a better understanding of the nature of microblogging in an OSS context and allow OSS communities to steer their microblogging efforts toward an optimal utilization. To this end, we investigated
microblogging in one OSS community – Drupal – which has more than 17,000 globally distributed developers and this number is steadily growing. Drupal is a popular and award-winning content management system, with over 10 years of history and seven major releases. Different from enterprise settings that use microblogging tools such as Yammer, StatysNet, Communote, Present.ly [8], the Drupal community has used Twitter since 2007 (see the “What is Twitter” sidebar). Twitter’s popularity in software development in general is increasing as well [9], which makes the study focus on Twitter even more interesting.

Research Approach

In order to understand how the Drupal developers use and benefit from Twitter, we conducted an exploratory case study. We analyzed both Drupal’s group Twitter accounts and a number of individual accounts of Drupal developers. The group accounts helped us understand the community’s collective use of Twitter while the individual accounts show how individual Drupal developers use Twitter in their day-to-day development activities. Given the large Drupal developer community, we focused on the sub-community that works on Drupal 7 (D7) – a recent branch of Drupal which is self-proclaimed as the “best work yet”\(^1\). The process chain in Figure 1 shows D7’s release timeline.

---

1 http://drupal.org/drupal-7.0
We included all 10 Drupal group Twitter accounts listed in the “Social Media Directory” webpage at the Drupal portal. For individual accounts, we obtained a list of 206 D7 contributors in 2008 when D7 was still in its main development phase (this number increased to more than 400 at the time of this study in 2012). We used the Drupal Member Directory to locate the Twitter account names of these contributors. We found that 141 of the 206 developers had a Twitter account. All but one had their accounts before January 2011 (the first official D7 release). To make the study feasible, we selected the Twitter accounts of 12 developers who committed most to the D7 code base, plus the two coordinators’ accounts. They are the core developers and constitute the most active part of D7 development community and therefore the key informants of our study.

In keeping with an existing study of OSS community blogging [10], our analysis focused on two aspects: Twitter usage and Twitter content. To understand Twitter usage, we analyzed quantitative information on the selected Twitter accounts, including number of tweets, frequency of tweeting, number of followers/followees. To gain an understanding of the nature of communication via Twitter, we analyzed a collection of tweets from the selected Twitter accounts. First we retrieved the tweet contents from these accounts using the twitteR package with the statistical data-mining tool R. This package extracts tweets through Twitter’s API. In total, we retrieved 12,167 tweets from both the group accounts and the selected D7 developers’ individual accounts. Since we focused on D7, we decided to restrict the time range of tweets from July 2008 (release of the first development snapshot) to January 2011 (the first official release), which covers the main development phase of D7 (indicated on Figure 1). We filtered the retrieved tweets using four keywords: “drupal 7”, “drupal7”, “#d7” and “(space)d7(space)”, to exclude tweets that are not explicitly related to D7. This resulted in a total of 568 tweets from 16 out of the 24 studied accounts, 363 of them from 10 individual accounts.

We analyzed these D7-related tweets through a coding process. One of us acted as the main coder and another checked the coding. We used the interaction types as found in enterprise blogging mentioned earlier [4], [8] as a set of seed categories. We classified each tweet into one or more categories. As we coded the tweets, new categories emerged for the tweets that didn’t fit any of the seed categories; the tweets already coded were then

---

2 http://drupal.org/social-media
3 We downloaded a spreadsheet from www.knaddison.com/drupal/drupal-7-who-providing-patches-next-release. It contains a list of already compiled D7 contributors based on the commit messages from the Drupal CVS system which allows to obtain all the information needed to decide who has been involved in the code that is ultimately committed.
4 http://drupal.org/profile
5 according to http://cran.r-project.org/web/packages/twitteR/index.html
6 a coordinator in Drupal is a “cat herder” who helps to “get as many new contributors as possible”, see drupal.org/user/24967. D7 had two coordinators during the studied period.
7 www.r-project.org/
8 We admit that filtering tweets using these keywords may inadvertently have excluded any tweets that do not contain them but could be D7 related. Therefore the 568 tweets represent the minimal collection of D7 related tweets.
revisited to check if they would fit any of the new categories. For instance, one new category is “express sentiment”. Based on this finding, we manually conducted a sentiment analysis and further categorized those tweets in terms of “positive sentiment” and “negative sentiment”. We also analyzed the “retweet”, “favorite” and “@” (directedness) properties of all tweets to understand the responses generated by them in the Twitter sphere. As a final step, through a process of synthesizing the categories, we identified a number of higher-level themes.

### Drupal community’s Twitter Usage

Table 1 lists the Twitter usage of the Drupal community represented by the selected accounts (sorted by creation date). As shown, most individual Twitter accounts (86%) were created between 2007 and 2008. Considering that Twitter started in 2006, this suggests that D7 contributors were early Twitter adopters. The group accounts were created much later; most (80%) were created after 2008. Both group and individual accounts are active in terms of the number of accumulated tweets and the tweeting frequencies (tweets per month).

<table>
<thead>
<tr>
<th>Account name</th>
<th>Function/role</th>
<th>Creation date</th>
<th>No. of followers</th>
<th>No. of followees</th>
<th>No. of tweets</th>
<th>Tweeting frequency (tweets/month)</th>
</tr>
</thead>
<tbody>
<tr>
<td>@drupal</td>
<td>Main Drupal account</td>
<td>25-Mar-07</td>
<td>33,889</td>
<td>173</td>
<td>1,382</td>
<td>21.3</td>
</tr>
<tr>
<td>@drupalcon</td>
<td>Drupal Conference account</td>
<td>26-Feb-08</td>
<td>9,499</td>
<td>2,027</td>
<td>4,339</td>
<td>78.9</td>
</tr>
<tr>
<td>@drupalplanet</td>
<td>Drupal related announcements (code, infrastructure, etc.)</td>
<td>19-Feb-09</td>
<td>7,443</td>
<td>3</td>
<td>5,602</td>
<td>130.3</td>
</tr>
<tr>
<td>@DrupalAssoc</td>
<td>The Drupal Association's official twitter account</td>
<td>19-Feb-09</td>
<td>6,248</td>
<td>7</td>
<td>588</td>
<td>13.7</td>
</tr>
<tr>
<td>@drupal_infra</td>
<td>General information about Drupal infrastructure</td>
<td>02-Apr-09</td>
<td>1,264</td>
<td>1</td>
<td>64</td>
<td>1.6</td>
</tr>
<tr>
<td>@drupalsecurity</td>
<td>Republish Drupal Security Advisories &amp; related news</td>
<td>24-Apr-09</td>
<td>2,066</td>
<td>31</td>
<td>394</td>
<td>9.6</td>
</tr>
<tr>
<td>@drupaldocs</td>
<td>Updates from Drupal documentation team</td>
<td>16-Jul-10</td>
<td>1,103</td>
<td>377</td>
<td>297</td>
<td>10.6</td>
</tr>
</tbody>
</table>
Some group accounts provide a constant and steady stream of tweets (such as @drupal or @drupalplanet) while others with specific purposes (such as @drupalcon) have tweeting “peaks” that generally occur during major events, such as a new version release or a Drupal conference.\footnote{The graphics at http://figshare.com/articles/Drupal_Group_Twitter_Accounts_Tweets_Frequency_Diagrams/748 791 show the tweeting frequencies of the 10 Drupal group accounts.} D7 contributors usually have stable tweeting behavior. Generally
speaking, the developers’ tweeting frequencies are higher than those of group accounts. However, developers tweet not only about work, but also about their personal lives. Table 1 also shows that group accounts have large numbers of followers but don’t follow many other accounts, except for @drupalcon and @drupaldocs. The developers’ individual accounts also have larger numbers of followers than followees.

Figure 2 illustrates the “following” linkages within and between the group and individual accounts, a close look at which can reveal the underlying communication structure in the Drupal community and information flow via Twitter.

Figure 2: The “following” links among the selected Twitter accounts

Figure 2(a) shows that the “following” linkages within the group accounts are loose. Few “following” linkages are symmetric. @drupal, the most-followed group account, is
the central node of the network. Isolated group accounts (those that don’t follow and aren’t followed by other group accounts) are not shown in the figure. In comparison, the D7 developers’ accounts are much more closely linked through the “following” relationship, many of which are symmetric. Figure 2(b) focuses on the links between the two D7 coordinators and the other 12 developers. The difference between the two coordinators is evident: @dries, followed by all, follows no other members; @webchick, follows everybody and is followed by everybody except @dries. We can argue that the social network formed by this structure suggests different roles played by the members in the community. In this case, @dries is a leader, and @webchick is a manager.

Figure 2(c) shows that there are also close linkages between individual and group accounts. Most developers follow at least one group account. On the other hand, some group accounts follow D7 developers too. Therefore, some symmetric “following” links are formed. @drupaldocs (the documentation team) is the only group account that none of the D7 developers follow.

D7 Twitter Content

Our analysis of D7-related tweets shows that some group tweets follow a unified format; other group tweets have a more personal touch. But generally tweets from the group accounts are impersonal and informative. In contrast, the D7 developers’ individual tweets are more versatile, interactive and reveal the affective states of developers. Table 2 is an overview of the findings from the analysis of D7 Twitter content.

Table 2: Findings from the D7 Twitter content

<table>
<thead>
<tr>
<th>Key themes</th>
<th>Percentage of group tweets</th>
<th>Percentage of individual tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets containing URLs</td>
<td>87%</td>
<td>53%</td>
</tr>
<tr>
<td>Tweets providing updates</td>
<td>53%</td>
<td>30%</td>
</tr>
<tr>
<td>Crowdsourcing tweets</td>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td>Tweets expressing emotions</td>
<td>6%</td>
<td>40%</td>
</tr>
<tr>
<td>Tweets as thank-you notes</td>
<td>2%</td>
<td>7%</td>
</tr>
</tbody>
</table>

Majority of tweets contain URLs

More than half (53%) of the individual tweets and 87% of the group tweets contain web links to project-related information, issues or discussion. Group tweets containing the links are usually generated automatically by the Twitter service embedded in various Drupal related systems. For example, the tweet below was automatically generated by the Drupal community website and contains a URL pointing to it:
Most of the links can direct followers to different parts of the Drupal.org community website, as well as other Drupal-related websites. Other tweets point to Drupal developer blogs and websites, or some other D7-related websites. This way, project-related knowledge is shared among developers and their followers. These web links, serving as information pointers, constitute an information map of the project that anybody can explore or interact with. For example, some of the information pointers in the analyzed tweets contain references to issues in Drupal’s issue tracker. Registered users can comment on an issue through the tracker’s website, while unregistered users can comment by replying to the tweets directly.

Positive emotions are openly shared

Besides pointing to information sources, many tweets announce upcoming events and provide updates of projects, tasks and developer status. In our set of sampled tweets, about a third (30%) of individual tweets and half (53%) of group tweets provide D7 updates.

What makes these update-oriented tweets from individual accounts interesting is that they also display feelings and emotions. Developers’ personal emotions and feelings are expressed and intertwined with the tasks they work on. Most of these tweets convey positive emotions with respect to the results that have been achieved, and they are often retweeted or marked "favorite" by followers. So, positive emotions are spread to other community members and the outside world. A good example is that, when the first release of the D7 series was out, five of the developers listed in Table 1 tweeted about it, such as @webchick:

“#Drupal 7 is now out! Friendly and powerful #opensource #CMS so awesome it won't fit in 140 chars. Try it! http://drupal.org/7” @webchick, 05-Jan-11.

This was retweeted by 255 of her 8,094 followers (including @dries) and favorited by 18 of them. The retweet from @dries was further retweeted by 80 of his followers. He also tweeted twice about the new release directly, which were retweeted 111 and 103 times, respectively.

We found very few arguments or negative feelings expressed in the tweets. From our sample, only two tweets could be considered to express somehow negative tones. They weren’t retweeted further.

10 http://drupal.org/project/issues
Crowdsourcing appeals are spread with “motivational sparks”

Another important use of Twitter we identified is to broadcast “crowdsourcing” appeals for completing tasks (6% of individual tweets, 8% of group tweets). Tweets don’t directly delegate or assign tasks to specific individuals. The targeted audience is often the community and general public.

Often the crowdsourcing tweets come from the two coordinators, especially @webchick. In contrast to the other general calls for help from group accounts and other individual accounts, @webchick’s crowdsourcing tweets are often combined with short but motivational phrases to encourage participation:

“Attn #drupal people: Core hack sprint tomorrow on IRC! Help D7 be more kick-ass, learn new tricks, and have fun, too! :) #ireviewdrupal” @webchick, 07-Aug-09.

This fits to her role as an active manager, as Figure 2(b) shows.

Some crowdsourcing tweets are retweeted by followers, spreading the calls for help further. For instance:

“Ok, folks! We’re about 24 hours away from #Drupal 7.0. Please git clone git://git.drupal.org/project/drupal.git and test! :D” @webchick, 04-Jan-11.

This was retweeted by 61 of @webchick’s followers, including @dries, whose retweet was in turn retweeted by another 61 of his followers. Crowdsourcing tweets from the group accounts seemed to attract less attention. Very few of them were retweeted, and only by a few followers (usually one to five). This suggests that personal appeals for contributions have more impact than the less personal group tweets. However, within our data set we couldn’t find evidence that these crowdsourcing tweets have actually reached any potential volunteers and elicited responses from them. Therefore, the efficiency of Twitter as a crowdsourcing channel is yet to be demonstrated.

Contributions are often appreciated with explicit thank-you notes

Tweets are also used as Thank-You notes (7% of individual tweets, 2% of group tweets), either directed to specific persons, to groups of people (using hashtags such as #drupalchix for Drupal’s female developers), or to a general audience:

“Much to be done yet, but thanks so much to everyone who's helped on the D7 docs up to the launch, you rock! #rockthedocs #drupal” @drupaldocs, 05-Jan-11.

One interesting aspect of these tweets is that these Thank-You notes often include a URL to the contribution:
“Drupal 7.0 article in @LinuxJournal: http://bit.ly/id42qQ Thanks @katherined!” @dries, 05-Jan-11.

This way, the names of contributors as well as their concrete contributions are shared with the community. Such tweets are not often retweeted as they are usually directed to a specific community member.

Twitter in Open Source, So What?

Our study shows that Twitter can support developers in open source communities in a number of ways. We discuss the key functions we identified below.

Information Radiator

Twitter can serve as a virtual information radiator for an open source community. When a community grows more connected through Twitter use, the distribution of information within the community should become quicker, minimizing misunderstandings between colleagues who do not meet face-to-face very often if at all [11]. Both the group tweets and individual tweets contain the latest updates as well as links to relevant information sources. Furthermore, information is spreading efficiently through retweeting, which generally happens within the first hour of the original tweet being published [12]. A distinctive characteristic of the OSS development model is that developers are dispersed and work in an asynchronous manner. Therefore, work must be visible to distributed members to facilitate development. Our study shows that Twitter makes the Drupal development process more transparent through constant streams of updates of the project’s progress.

The majority of analyzed tweets contain URLs that point to artifacts that developers have worked on. These include several aspects of the development process, such as submitted requests and changes, crucial decisions, important milestones, or even team retrospectives. This pervasiveness of URLs is also reported to be important by Github, another open source project [13]. This facilitates easy traceability of an artifact’s development history, which in turn helps to understand its evolution. The Drupal community’s use of Twitter demonstrates an effective way of making these URLs more accessible to a wide audience.

Knowledge Preservation

Knowledge preservation is another significant role of Twitter. While spoken communication and chat tools such as IRC (Internet Relay Chat) are “mostly fleeting and not useful as a durable process of knowledge sharing and further development” [6], in
Drupal’s case Twitter externalizes tacit knowledge during informal communication, saves it persistently and makes it publicly available.

This is particularly significant for potential contributors. Both the group accounts and individual developers’ individual accounts attract large numbers of followers, many of whom can be potential contributors, if properly motivated and given the right access to the community knowledge base. Twitter provides another channel to access to the community knowledge base in addition to forums, mailing lists, etc. Meanwhile, crowdsourcing tweets can suggest areas where contributions are most needed at any one moment. The motivational phrases in these tweets encourage them to join and become part of the community. Microblogging as a crowdsourcing mechanism has not received much attention in the past studies [14] but our study shows a clear potential for this.

**Self-organization Facilitator**

Twitter facilitates an effective self-organization of developers. In enterprise microblogging, “coordinating others” is the second largest interaction type after “providing updates” [4], which implies that communication via enterprise microblogging influences what other team members do. In our study, however, we didn’t find directed tweets to assign/delegate tasks to a developer directly. While some community members play coordinating roles (such as @dries and @webchick), they are not there to grant “commit bits to repositories before you can do work” [13], or to approve releases.

Our analysis supports the “cooperation without coordination” model common to many OSS projects [13]. Coordination in OSS projects is often through a self-organizing process of task selection [15]. Our study shows that the coordination value of Twitter, if any, is minimal, perhaps due to its broadcast nature and limited message length [14]. Therefore we argue that Twitter shouldn’t be stretched to support interactions such as coordination that requires more media-rich communication channels.

**Positivity Spreader**

Distributed software developers feel more connected to each other when they are able to share not only activities but also mood [16]. Drupal team members use Twitter to share it. Tweets from the Drupal community are much more emotional. This is distinctively different from enterprise microblogging, which is mainly dedicated to work and is kept impersonal [4]. Being a “public board for people,” Twitter enables more socially open communication as developers often display emotions while they tweet about their work. However, community developers seem to refrain from both tweeting and spreading tweets with negative tones, but rather focus on spreading positive emotions. The frequent Thank-You tweets also contain positive messages and help motivate developer involvement in the community, and offer its members recognition, a sense of achievement and belonging.
to a community. This is an important and relatively easy way for community members to maintain their social bonds [14].

Conclusions

Similar to its role in society at large, microblogging also plays an increasingly important role in open source communities. Twitter enriches Drupal’s ecosystem by providing a key channel for information sharing, project updating, motivating developers and crowdsourcing.

Due to our limited sample, we provide only a partial view of the tweeting phenomenon in the Drupal community. It remains to be seen whether other OSS communities demonstrate similar microblogging behaviors. It is not clear how effective Twitter is compared to other communication media commonly used in OSS communities, such as mailing lists, forums, and IRC. It is also intriguing to see how often and voluntarily developers use Twitter as a sideline, and what factors influence their microblogging behaviors. While more research is needed to better understand these and other questions, our study of Drupal clearly demonstrates how microblogging can support a large distributed software development endeavor. Microblogging may reshape the way people collaborate and engage in software development.

What is Twitter?

Twitter is a well-known implementation of microblogging, which started in April 2006. Twitter messages are called “Tweets,” which have a maximum length of 140 characters. Relationships between people with a Twitter account are unidirectional, which means that one user can “follow” another, but the followee doesn’t need to “follow” back the follower.

All tweets are public by default, though it’s possible to “protect” tweets by making them visible to selected users only. Messages can be directed to a particular person, by prefixing the recipient’s username with a “@” sign. Interesting tweets can be “retweeted” (RT), so the original tweet can reach a wider audience. People can “like” a tweet by clicking the “favorite” link associated with each tweet.

Tweets can be “tagged” using hashtags to highlight a topic. For instance, people tweeting about “microblogging” could put the hashtag “#microblogging” in their tweet.

A Twitter account can belong to either an individual or a group of people, even though it is always represented by a single Twitter name. By group account we mean a Twitter account that is maintained by a group of people, and by individual account an individual user.
Acknowledgments. This work was supported, in part, by Science Foundation Ireland grant 10/CE/I1855 to Lero—the Irish Software Engineering Research Centre (www.lero.ie).

References


About the Authors

Dr. Xiaofeng Wang is a researcher at the Free University of Bozen-Bolzano. Her research areas include software development process, methods, agile software development, and complex adaptive systems theory. Contact her at xiaofeng.wang@unibz.it.

Ilona Kuzmickaja is an analyst at Facebook. She received a master’s degree in software engineering from the University of Bozen-Bolzano. Her research interests include social media in distributed software development. Contact her at i.kuzmickaja@gmail.com.

Dr. Klaas-Jan Stol is a researcher at Lero, the Irish Software Engineering Research Centre at the University of Limerick. His research interests include open source software, Inner Source and agile and lean software development. Contact him at klaas-jan.stol@lero.ie.

Prof. Pekka Abrahamsson is a full professor of computer science at Free University of Bozen-Bolzano. His research interests are centered on empirical software engineering, agile & lean software development and cloud computing. Contact him at pekka.abrahamsson@unibz.it.

Prof. Brian Fitzgerald holds an endowed chair, the Frederick Krehbiel Chair in Innovation in Business and Technology at the University of Limerick, where he is also Principal Investigator in Lero – the Irish Centre for Software Engineering. His research interests include open source software, lean and agile methods. Contact him at bf@ul.ie.