

# How Modern News Aggregators Help Development Communities Shape and Share Knowledge

Maurício Aniche<sup>1</sup>, Christoph Treude<sup>2</sup>, Igor Steinmacher<sup>3,4</sup>, Igor Wiese<sup>4</sup>  
Gustavo Pinto<sup>5</sup>, Margaret-Anne Storey<sup>6</sup>, Marco Aurélio Gerosa<sup>3</sup>

<sup>1</sup>Delft University of Technology, <sup>2</sup>University of Adelaide, <sup>3</sup>Northern Arizona University

<sup>4</sup>Technological University of Paraná (UTFPR), Campo Mourão, <sup>5</sup>University of Pará (UFPA), <sup>6</sup>University of Victoria

## ABSTRACT

Many developers rely on modern news aggregator sites such as Reddit and Hacker News to stay up to date with the latest technological developments and trends. In order to understand what motivates developers to contribute, what kind of content is shared, and how knowledge is shaped by the community, we interviewed and surveyed developers that participate on the Reddit programming subreddit and we analyzed a sample of posts on both Reddit and Hacker News. We learned what kind of content is shared in these websites and developer motivations for posting, sharing, discussing, evaluating, and aggregating knowledge on these aggregators, while revealing challenges developers face in terms of how content and participant behavior is moderated. Our insights aim to improve the practices developers follow when using news aggregators, as well as guide tool makers on how to improve their tools. Our findings are also relevant to researchers that study developer communities of practice.

## CCS CONCEPTS

• Human-centered computing → Collaborative and social computing;

## KEYWORDS

News aggregators, development communities, knowledge sharing

## ACM Reference Format:

Maurício Aniche<sup>1</sup>, Christoph Treude<sup>2</sup>, Igor Steinmacher<sup>3,4</sup>, Igor Wiese<sup>4</sup>, Gustavo Pinto<sup>5</sup>, Margaret-Anne Storey<sup>6</sup>, Marco Aurélio Gerosa<sup>3</sup>. 2018. How Modern News Aggregators Help Development Communities Shape and Share Knowledge. In *Proceedings of ICSE '18: 40th International Conference on Software Engineering (ICSE '18)*. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3180155.3180180>

## 1 INTRODUCTION

To succeed in today's fast-paced and ever-changing work environment with new technologies and devices being constantly introduced, software developers need awareness of the latest news related to their profession [56]. Successful developers stay abreast

of new trends and technologies and gain important knowledge about a technology they are using from a number of different channels that support knowledge exchange at a community level. In practice, knowing when to pivot to a new technology can help a startup company succeed, while at the developer level, learning about an important security bug can save hours of frustrations.

The collaborative nature of software development continues to evolve, shape, and be shaped by communication channels used by development-related “communities of practice” [63]. Within a community of practice, knowledge about software is the combination of externalized (e.g., code, documentation) and tacit knowledge (e.g., experience of when to use an API, or design constraints that are not written). These channels support developers in collaboratively forming and sharing externalized and tacit knowledge [54].

Several different developer communities and communication channels are focused around rapid knowledge dissemination, Stack Overflow and Twitter being well-known examples. Stack Overflow makes use of gamification features to encourage curation and rapid knowledge dissemination, while Twitter relies more on social networking features. These channels have been extensively studied by researchers, ranging from studies on their importance [50], tweet content [56], tweeting behavior [57], hot topics [48], and automated classification [43, 65] to how developers ask and answer question in Stack Overflow [59], how to design efficient Q&A websites [36], and how to share good code examples [42].

News channels also play an important role in knowledge exchange [32]. News aggregators are regularly featured on compilations such as “Websites Every Developer Should Visit” [1]. These modern *news aggregators*, such as Hacker News [23] and Reddit [44], offer a blend of gamification and social network features that increases developer social capital in a community and enables faster dissemination of curated content. Developers are already relying on modern news aggregator sites; some even place them in their top three communication channels to stay abreast of developments in their community, describing them as “the heartbeat of the current software development industry” [54]. And while the community already has several insights on the role that Stack Overflow and microblogs play, there is little understanding of how news aggregators play a role in community curation and dissemination of knowledge. It is furthermore unclear what benefits developers gain by posting or reading content on these sites, what kind of content is shared, and what is the overall effect of these news aggregators on the technology community.

Up to now, there have only been a small number of studies of news aggregators for software developers. Early work on Slashdot [29, 52] and Digg [18, 33] focused on the moderation of comments and the prediction of popularity, respectively. Research on

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

ICSE '18, May 27–June 3, 2018, Gothenburg, Sweden

© 2018 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery.

ACM ISBN 978-1-4503-5638-1/18/05...\$15.00

<https://doi.org/10.1145/3180155.3180180>

Hacker News and Reddit has investigated the relationship between intrinsic content quality and popularity [53], the importance of a good title [28], the production and curation of news content in specific subreddits [31], and the potential of Hacker News for expanding qualitative research findings [7].

In this paper, we present a study to understand how modern news aggregators, *i.e.*, those that incorporate gamification and social network features, are used by software developers to curate and disseminate knowledge. We consider the popular and widely recommended Reddit and Hacker News aggregators in our study. These aggregators enable users to post links with a short description and support commenting as well as voting on posts and comments.

We first interviewed 22 active developer members from Reddit's r/programming subreddit [46] community to understand their motivations to read, post, comment, and vote on such content, as well as their perceptions of the challenges they face and improvements they would like to see on news aggregators for developers. We then validated the interview findings with an additional 99 participants from Reddit's r/programming subreddit community. The developers we interviewed also, as we expected, use Hacker News. In the final phase of our research, we focused on understanding what kind of development content is posted on Reddit and Hacker News sites.

Our study leads to the following three contributions:

- An understanding of what motivates developers to read, share, and participate in modern news aggregator communities that helps researchers interested in understanding how communities are formed and exchange knowledge (Sec. 4.1).
- A characterization of the existing content in developer news aggregators, bringing understanding to practitioners on what kind of stories they will find on these channels (Sec. 4.2).
- A set of improvement suggestions for news aggregators site owners as they try to improve and evolve these community wide communication channels (Sec. 4.3).

## 2 BACKGROUND

In this section, we describe the news aggregators investigated in this research and how it is useful to view them through the **communities of practice** theoretical lens.

*Reddit.* Reddit is a social news website where registered users submit links or text posts. Users then upvote or downvote each submission, helping to determine its prominence on Reddit's pages. Users also comment on posts and respond in a conversation tree of comments. Content is organized by areas of interest or sub-communities called subreddits [16]. Currently, subreddits such as 'science' and 'worldnews' are among the most popular subreddits. At the time of writing, Reddit had received over 8 billion page views from almost 60 million unique visitors within the last 30 days from the United States alone, and it was the eighth-most visited site in the world [4]. While there are several sub-communities on Reddit for developers (some of which focus on specific technologies, such as r/javascript or r/php), r/programming is the most popular one among those focusing on topics related to software development. According to its front page, "r/programming is a subreddit for discussion and news about computer programming." r/programming's submission rules include the statement "Just because it has a computer in it doesn't make it programming. If there is no code in your

link, it probably doesn't belong here."

*Hacker News.* Hacker News is a social news website focused on technology news [13], with functionality very similar to Reddit. The main difference between the two news aggregators is that Hacker News focuses on technology news and has no sub-communities, while Reddit, in general, does not focus on any topic, but supports sub-communities that do. Hacker News states that "stories on HN don't have to be about hacking, because good hackers aren't only interested in hacking, but they do have to be deeply interesting." [24] Deeply interesting is then defined as "stuff that teaches you about the world." Our content analysis sheds additional light on the content that is shared on Hacker News and contrasts it with the content found on r/programming. At the time of writing, Hacker News had 90 million page views from 3.5 million unique visitors within the last 30 days from the USA [3].

*Communities of Practice.* Communities of practice are groups of people connected by their need to learn together to support their activities in a particular domain [63]. Community members do not need to be collocated nor do they need to have previous social ties, but they share goals within a common domain and may need to solve similar problems. They learn practices and gain knowledge from each other through processes like apprenticeship or mentoring. In these communities, novices learn by watching and taking peripheral activities that add value to the community, in a process called Legitimate Peripheral Participation [30].

An example of a software development community of practice is Open Source [54]. Peripheral developers start by lurking and reading mailing list threads. They then fix bugs and progress to the point where they add their own features. Eventually these developers may be granted commit rights, and as core members may play a role in project decisions [14, 41]. Similarly, we see news aggregators and the people who post to and read their content as a community of practice. Members share learning goals with each other and reify their knowledge through news posts that may be improved by other members. Furthermore, members develop relationships over time (through posting, voting, and commenting) and communities form around important topics, generating a shared repository. The news aggregators we study in this research foster social relationships and support the formation of global, online software developer communities.

## 3 RESEARCH METHOD

This study aims to understand programming-related news aggregators (r/programming and Hacker News), having the following research questions:

- RQ1:** What motivates developers to participate in programming-related news aggregators?
- RQ2:** What are the characteristics of the content posted in programming-related news aggregators?
- RQ3:** What suggestions do the participants have to improve the community and features of programming-related news aggregators?

To answer these questions, we employed interviews, survey, and content analysis.

### 3.1 Interviews

To understand the motivations and perceptions of the participants, we conducted semi-structured interviews. Semi-structured interviews encourage interviewees to freely share their thoughts and enable researchers to follow up and explore interesting topics that might emerge [27]. We recruited active members of the r/programming community for the interviews. As a quantitative proxy for activity, we ordered participants by their number of posts according to our dataset (cf. Section 3.3) and sent a private message to the 229 most active r/programming contributors in terms of number of posts. A potential limitation of recruiting only active r/programming contributors is that their responses might only focus on Reddit. This choice was intentional because Hacker News does not offer a way of contacting its contributors. However, we conjectured that active r/programming users would also use other news aggregators, and thus, would be able to contribute to our research by discussing more than one news aggregator. Indeed, all interviewees do use more than one news aggregator.

We interviewed 22 participants (response rate of 10%), identified below as P1 – P22. All participants work in the software development industry. While most are developers, two are marketing specialists (P8, P20) and two are Computer Science students (P11, P12). Their experience in industry ranges from 3 to 25 years. Interviewees belong to different ethnic groups (living in South America, North America, Europe, Asia, and Oceania). Together, they speak 12 different languages. Table 1 shows participant demographics. 18 interviews were conducted via video calls and four interviews were conducted via e-mail (as requested by the participants). At the start of each interview, we explained our research goal and sought permission to share the data anonymously in our findings. We recorded each audio interview, which took on average 30 minutes.

Our interview guide focused on five main topics: (1) the interviewees' perceptions about what motivates them to use the news aggregator, (2) reasons to contribute (*i.e.*, posting news, commenting, and voting), (3) the content they commonly read and what the learning outcomes are, (4) the differences among news aggregators (*e.g.*, Reddit, Hacker News), and (5) pitfalls and challenges they face, as well as advice for newcomers. The detailed interview guide can be found in our appendix [5].

We conducted interviews until we came to an agreement that theoretical saturation was reached. According to Strauss and Corbin [55], sampling should be discontinued once the already collected data is considered sufficiently dense and data collection no longer generates new information. In addition, in light of findings in psychology that claim human universals are extrapolations from a single social group [26], we made sure to interview people from different ethnic groups before deciding whether saturation was reached. More specifically, the authors who performed the interviews frequently compared their notes and found that they were conducting interviews that were not leading to any new insights.

We analyzed the interview data in two steps. In the first step, we attached codes to any piece of relevant data found in the answers we collected from the interviewees. Simultaneously, we grouped these labels into higher-level categories. The initial set of categories was derived from our interview guides; however, throughout the analysis, other categories emerged. These categories became the

**Table 1: Demographics of our interviewees.**

ID	Location	Years of exp.	Frequency of usage of any news aggregator
P1	Europe	25 years	At least once a day
P2	South America	25 years	Once or twice per day
P3	Europe	11 years	Once or twice per week
P4	Europe	15 years	Twice per day
P5	North America	15 years	At least once a day
P6	Europe	25 years	At least once a day
P7	North America	7 years	At least once a day
P8	North America	12 years	At least once a day
P9	Europe	13 years	At least once a day
P10	North America	7 years	At least once a day
P11	Europe	4th year CS	Multiple times a day
P12	Europe	3rd year CS	Once or twice a day
P13	Europe	6 years	Once a week
P14	Europe	8 years	Few times a week
P15	North America	13 years	At least once a day
P16	Asia	3 years	Three times a week
P17	North America	6 years	At least once a day
P18	North America	4 years	At least once a day
P19	North America	5 years	At least once a day
P20	North America	3 years	Once or twice a week
P21	Europe	5 years	At least once a day
P22	Oceania	7 years	At least once a day

topics we discuss in the Findings section. We share our complete code book on our appendix [5]. Due to privacy reasons, we do not share the original excerpts. Interviews were conducted by three of the authors, and these three authors were involved in the coding of the data and in deriving higher-level categories.

### 3.2 Survey

We designed a survey aimed at triangulating the qualitative findings. Respondents had to indicate their agreement with statements derived from the interview analysis, which represented codes that not all of our interview participants agreed on. Answer options ranged from strongly disagree to strongly agree, in a 5-points Likert scale with a neutral option; we also allowed users to respond "I do not know" or "does not apply." The survey contained 21 statements (see Figures 1 and 2) and demographics questions. The estimated time to answer our survey was four minutes. After conducting a pilot with two participants, we shared the survey on five social networks: Medium, Hacker News, Reddit, Twitter, and Facebook. Due to the nature of these websites, we are not able to estimate how many users we sent it to.

We obtained 99 answers (3 from Medium, 5 from Facebook, 36 from Twitter, 21 from Hacker News, and 34 from Reddit). These 99 participants were from 32 different countries. Respondents vary regarding how long they have been using news aggregators (min=0 years, q1=1.5 years, median=3 years, q3=6 years, max=15 years). 34.3% of the participants visit such websites at least once a day, and another 30.3% between once a day and once a week; only 5.1% indicated that they visit news aggregators less than once a month. All participants are involved with software development activities

(we removed 4 participants that were not). The survey is available in our online appendix [5].

### 3.3 Content Analysis

We collected all posts and comments on Hacker News and r/programming posted between 01 January 2016 and 30 April 2017. We used the data available from Google Big Query because the APIs available in both news aggregators have very restricted rate limits to collect data.<sup>1</sup> We collected “title,” “description,” “author,” “url mentioned,” “user score,” “timestamp,” number of “comments,” and number of “comments per user.” We gathered 492,752 posts from Hacker News and 44,640 posts from r/programming.

To understand the content posted on the selected news aggregators, we randomly selected a sample of 300 posts from Reddit’s r/programming and 300 posts from Hacker News, which gives our conclusions a confidence level of 95% with a margin of error of 6% [47]. To perform the qualitative analysis of the 600 posts, we first defined a classification schema. Four of the authors independently analyzed and tagged 20 posts and met to reach consensus on a common schema. Merging the different perspectives of categorization that emerged from this analysis, we identified the following categories of analysis: type of source, purpose, type of content, theme, related technologies, and related development activity.

The four authors then independently analyzed another sample of 20 posts following this categorization schema and met again to further refine the schema. Once the schema was stable, two of the authors proceeded with the analysis of the remaining posts. They sequentially analyzed rounds of 20 posts independently and met to discuss, until reaching an agreement level greater than 80%. Three rounds were necessary until achieving an agreement of 83.33%. After that, each researcher analyzed and categorized half of the remaining posts, frequently discussing questions and new categories found. After classifying all posts in the sample, the results were presented to a third author. The three authors discussed the results and created meaningful, more generic categories to better organize the technologies and themes identified during the analysis, which were very disperse. We organized the themes in categories inspired by the top tags on Feedly [19], a news aggregator application that compiles news feeds from a variety of online sources. To organize the technologies mentioned in the posts, we used the categorization proposed by Borges *et al.* [9], adding two new categories: Hardware/Gadget and Programming/Markup Languages.

## 4 FINDINGS

In this section, we present and discuss our findings.

### 4.1 Motivations to participate (RQ1)

Participants reported motivations to post, comment, vote, and read.

**4.1.1 Motivations to post.** The analysis of our interviews revealed that there are two main reasons for participants to post links on programming-related news aggregators: self-promoting own work and sharing relevant content. All interviewees except for P6, P17, P19, and P21 reported to self-promote their work, such as blog

posts, open-source systems, and products, to the news aggregators communities. The news aggregators provide participants with access to a large number of readers, which they would not have easily otherwise (P14, P15). P14 said Reddit gave him easy access to thousands of readers.<sup>2</sup> P1 highlighted that without sharing his posts on news aggregators his blog would not get any traffic and P9 reinforced that the news aggregators help to create an audience for his blog. Indeed, as discussed in the next subsection, blog posts are the most common type of content posted on both news aggregators.

Participants also promote their own work to get feedback and possibly improve the content created (P1, P4, P5, P9, P14). P14 stated that *“If you wanna share something and get people’s opinion, it’s a very quick way to do it. I started using Reddit to get feedback about my blog post.”* Participants also reported that number of upvotes and comments is a measurement of how good their work is (P1, P5, P8).

The other major reason for posting is to share third-party content that might be interesting to others (P1, P4, P8, P12, P15-18, P21-22). Active posters also share content because it is part of their jobs as community members (P13, P14, P22). P14, for example, holds a Microsoft Most Valuable Person title. According to Microsoft’s website, an MVP is a technology expert who passionately shares their knowledge with the community [38].

A link may also be shared with the goal of discussing the topic (P1, P10, P12, P16-19, P21). P19 said *“It’s particularly interesting when the consensus differs from the conclusion I came to regarding the article.”*; P5 also exemplified: *“I recently shared a post on functional programming, and it was nice to see how different people reacted to the same story.”* In our survey, 43% of the respondents agreed with the statement that they post to collect opinions on the topic; on the other hand, 32% indicated not to do it (Figure 1, Q5).

Gamification also may play a role in why users post on these websites (P1, P6). P1 said Reddit is a game and it is fun when a post gets upvoted. P6 reported having posted top stories from Hacker News into Reddit to increase his karma (more details about this feature in Sec. 4.3.3). Finally, two participants indicated that one of their goals when sharing content is to improve the software industry (P9, P22). P22 stated: *“I put forward information that can improve the community [...] we can change the culture of the industry.”*

**4.1.2 Motivations to comment.** Most participants comment on a post when they have something that they consider useful to add (P1, P2, P4, P7, P10-12, P14-15, P19-20). P2 considers himself a specialist in a specific technology that has a small community, and he tries to be active, participating in discussions of posts related to that technology. P14, as an MVP, wants to help other participants, providing examples and solutions to problems. P1 and P11 add other references and links that can complement a discussion. P11 exemplified: *“If there’s a discussion on best practices, I sometimes have another blog [post] to share [on the topic].”* On the other hand, P12 said that in most cases he does not have much to comment. According to him, as he mostly browses popular posts, all possible insights are already there, leaving no space for his ideas. P15 and P19 usually do not comment on other users’ posts; according to

<sup>1</sup>The two datasets can be found at <https://bigquery.cloud.google.com/dataset/fh-bigquery:hackernews, fh-bigquery:reddit>.

<sup>2</sup>In fact, we found in our analysis of the posts from Jan 2016 to April 2017 14k distinct participants in Reddit r/programming and 82K in Hacker News (only considering those that posted or commented something — the number of readers might be even greater).

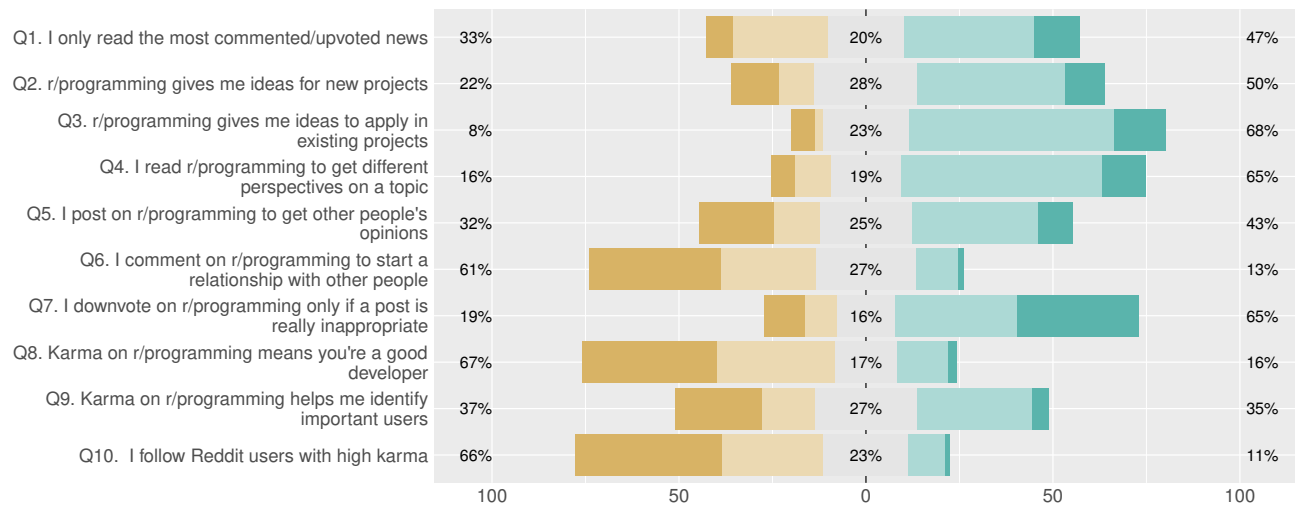


Figure 1: Responses to our survey (part 1).

them, to do so, they have to feel strongly about the particular discussion. In fact, our analysis of the posts from Jan 2016 to Apr 2017 revealed that the median number of comments per user is 2 on r/programming (q1=1; q3=5) and 1 on Hacker News (q1=0; q3=5).

Some participants comment to start a conversation (P3-5, P10, P14). P3 reported having invited participants to write guest blog posts on his website: *"That's the big impact that I have, to take a conversation out of Reddit."* Interestingly, only 13% of our survey respondents indicated to comment to start relationships (Figure 1, Q6). P14 reported starting conversations with participants that gave negative feedback on his posts. P4, P5, and P10 reported to start conversations when they have a different point of view than the one expressed in the post. P5 said *"Most of my comments are questions. The author needs someone to question her assumptions. The reader needs to know why someone has that opinion."*

**4.1.3 Motivations to vote.** Participants upvote posts that they consider interesting (P1, P2, P4, P6, P9, P11-14, P17-20, P22). They see voting as a way to filter good posts so that other people find them (P2-3, P5-6, P11, P17, P20). Other reasons to upvote a post include to see more comments and discussion (P12), to make readers learn something new (P14), to have a more complete profile (P3 said that he does not want to be a user who only posts links), or to bookmark a post for later (P16). Users upvote comments because they agree with a comment (P9, P10, P19), or because comments add to the discussion (P13, P15, P19). Interestingly, participants indicated to upvote posts or comments that go against their beliefs when they make a good point (P4, P6, P9-10, P14): *"I'd upvote it because I'm interested in people's opinions"* (P14).

Conversely, downvoting is not seen as a common activity, as it may be seen as demotivating (P20). Some of our interviewees explicitly mentioned that not liking the post (P1, P2, P12) or not agreeing with it (P11, P12, P14) does not trigger them to downvote. Rather, most participants indicated that they only downvote when a post or a comment is inappropriate, off-topic, offensive, impolite, or spam (P1, P3-4, P7, P9-10, P12-16, P18-22). As examples, P12 said that *"if I downvote a comment, it's because it's terribly off-topic or really offensive and it doesn't add anything to the discussion."* 65% of

our survey respondents also only downvote when the content is inappropriate; for 19%, this is not the only reason (Figure 1, Q7).

**4.1.4 Motivations to read.** Our interviewees consider that programming-related news aggregators provide relevant content (P1, P2, P6-7, P12, P19). The posts make them aware of the trending topics in different areas of software engineering (P5, P10-12, P14, P17-18, P21). P2 stated that he uses Reddit and Hacker News as a kind of newspaper. P12 said: *"[My] main goal is to keep in touch with what happens in the industry. Especially as our industry is really really wide and university is more focused on theory. So, I need to find other sources to see other technologies."* P14 stated: *"I can see what has been upvoted this week, and what has been interesting for most people. It saves you some time, and it's already sorted by the interests of thousands."* Indeed, 47% of the survey respondents affirmed to read only the most upvoted comments and posts; 33% also read non-popular posts (Figure 1, Q1).

Reading about trending topics in other areas can be especially useful to get up-to-date with technologies that developers do not use in their daily job (P4, P5, P12). P4 exemplified: *"I only have time to work with a single stack. So, [in Reddit,] you see perspectives from people that work on different stacks."* P12 stated that he wants to know if something big happens in other areas, even if he does not work on them.

Other participants use news aggregators as a source of ideas for future books, blog posts, and products (P1, P3, P4). P1 writes blog posts based on discussions he reads; P3 said that he commonly goes to the subreddit and observes problems that communities face to build a product later on; P4 claimed that a large part of his published book about a programming language came after reading and participating in discussions about the topic. Indeed, 50% of our survey respondents use news aggregators as a source for new projects, and 68% affirmed the posts give them ideas to apply in their existing projects (Figure 1, Q2 and Q3).

Only two participants (P4, P21) indicated to visit these news aggregators not only for reading the posts but also for the discussion that may happen. Reading comments helps users filter out sensationalist content (P4) or provides them with a different point

of view (P21). P4 said that “People have strong opinions. ‘This looks sensational,’ let’s see the comments, and the top comments will tell you why it’s sensational... It’s a way to get a balance.” Indeed, 65% of our survey respondents read r/programming to get a different perspective on the content (Figure 1, Q4).

**4.1.5 Impact of the news aggregators.** Programming-related news aggregators impact lives of our participants in several different ways. First, because of the aggregators’ smooth information flow, they become valuable marketplaces with good ideas that developers can take advantage of (P5, P8, P10, P16, P18-19). According to P8, just the fact that news aggregators filter noisy posts and enable him to read only useful technical information already makes a difference in his life as a developer, as also complemented by P19: “I’ve learned a lot from reading articles/blog posts about programming, and that’s (hopefully!) helped make me a better developer.”

These posts also afforded our participants different opportunities, such as ideas and technologies to learn (P3, P9, P15, P22), to evaluate new technologies at work (P7), to try technologies they had never tried before (P18), to see how other developers are building software and learn from their experience (P15), to ask for help (P20), and to be part of a bigger community (P22). To the latter, P22 said “it’s just like your place of work, where you have your coworkers, where you can ask for advice or share stuff and have interesting discussions. The difference is that instead of having a couple of team members, you do have the whole Internet to share things. The scale is bigger, so the benefit is also bigger.” P3 also indicated that programming-related news aggregators enable him to meet experts from many different fields. Interestingly, our survey respondents also notice this benefit: 52% of them agreed that r/programming gives them access to experts; only 21% did not agree (Figure 2, Q11).

## 4.2 Characteristics of the content (RQ2)

Our r/programming dataset contains 44,640 posts, 671,603 comments, and 14,253 participants, and the Hacker News dataset has 492,752 posts, 2,812,034 comments, and 81,532 distinct participants. Table 2 shows the distribution of comments per post. A typical post on Hacker News attracts one comment whereas a typical post on r/programming attracts two comments. Similarly, Table 2 shows the distribution of score (*i.e.*, upvotes minus downvotes) per post. The bottom half of Table 2 shows the number of contributions per contributor on r/programming and Hacker News, in terms of posts and comments. A user on Hacker News contributes, on average, with one post and one comment while a user on r/programming contributes, on average, with two comments, but no posts. We considered all users that had made at least one contribution (post or comment) in our dataset to produce these statistics.

To understand the content of programming-related news aggregators, we analyzed a sample of 300 posts for each aggregator, as described in Sec. 3.3. We used the same sample to understand the extent to which the same post appears on Hacker News and r/programming. We complement the answer with data from the interviews and survey. We also relied on the quantitative analysis of the dataset to have an overview of the contents of the posts.

**4.2.1 Characteristics of the content.** We characterize the content of each news aggregator in five categories: purpose, theme,

**Table 2: Descriptive statistics about our datasets**

	Min	Q1	Median	Q3	Max
comments/post HN	0	1	2	3	5,771
comments/post r/prog	0	0	1	5	2,094
score/post HN	0	0	0	0	2,530
score/post r/prog	0	0	1	7	15,385
posts/user HN	0	0	1	1	3,729
posts/user r/prog	0	0	0	0	866
comments/user HN	0	0	1	5	7,464
comments/user r/prog	0	1	2	5	4,679

technology, development activity, and source.

**Purpose.** Analyzing the primary purpose of the posts, we found a difference between the aggregators. Posts aiming at *learning* are more common on r/programming (r/p: 45%; HN: 18%), while *publicizing news* is more common on Hacker News (r/p: 6%; HN: 29%). Interestingly, on Hacker News, some news are published in scientific venues (r/p: 0%; HN: 10%). Publicizing software, such as libraries, tools, and toy projects, is more common on r/programming (r/p: 19%; HN: 7%), while announcements of new software versions are more common on Hacker News (r/p: 3%; HN: 7%). Asking for feedback, help, and questions (r/p: 2%; HN: 6%) and sharing opinions (r/p: 6%; HN: 12%) are also more common on Hacker News. In addition, we found posts advertising products and services (r/p: 3%; HN: 1%), sharing collections (r/p: 3%; HN: 3%), threat alerts (r/p: 2%; HN: 2%), and humor (r/p: 2%; HN: 1%), among others.

As mentioned before, learning a specific technology or concept plays a significant role on r/programming. Common styles of the *learning* posts are tutorials (r/p: 31%; HN: 17%), specific tips (r/p: 10%; HN: 18%), tech reviews (r/p: 10%; HN: 2%), experience reports (r/p: 7%; HN: 11%), conference talks (r/p: 7%; HN: 2%), screen casts (r/p: 7%; HN: 2%), books (r/p: 2%; HN: 2%), and code explanations (r/p: 2%; HN: 0%). These results confirm some motivations brought by our interviewees, who mentioned that Reddit brings opportunities to experiment and learn new technologies.

**Theme.** We found that on r/programming, as expected, most posts have an IT-related theme (88% of the posts) and of those, 74% focused on development, 15% on technology in general, 6% on security, and 4% on artificial intelligence. On Hacker News, posts are more diverse, with only 55% of posts being IT-related (48% of them focusing on development, 36% on technology in general, 10% on security, and 7% on artificial intelligence). For posts that are not related to IT, job market was the most mentioned theme in r/programming (4% of all posts). On Hacker News, the scope is broader, with some focus on business (12%) and society-related (*e.g.*, religion, living costs, immigration) content (9.6%), including themes such as startups (4%), politics (3.6%), and science (3.3%).

**Technology.** For the posts related to IT, we classified content according to the technologies related to the post. The results are summarized in Table 3, where we can observe that when focusing on programming/technology, r/programming and Hacker News posts

**Table 3: The technologies discussed in IT-related posts (some posts are related to more than one technology)**

Technology	r/p	Hacker News
Programming/ Markup Languages	52%	43%
Libraries/Frameworks	30%	28%
System Software (e.g., operating systems, middleware, DB)	21%	20%
Software Tools (e.g., IDEs package managers, compilers)	11%	18%
Application Software	9%	11%
Hardware/Gadget	1%	6%

are related to similar topics, with Programming/Markup Languages appearing as the most mentioned technologies on both aggregators. By analyzing the technologies that the posts mention, considering r/programming and Hacker News together, we find that JavaScript (7.3%) and Java (6.9%) are the most common in our sample, followed by C++ (4.8%), Python (4.8%), C (4.4%), and React (3.6%).

**Development Activity.** We also classified posts according to the development activity they are related to (for the subset of posts related to software development). We found that most of them are related to programming activities (r/p: 72%; HN: 59%), followed by DevOps (r/p: 13%; HN: 14%). We also found posts that relate to configuration management (r/p: 4%; HN: 3%), testing (r/p: 4%; HN: 5%), and design/architecture (r/p: 2%; HN: 2%).

**Source.** Blog posts (r/p: 47%; HN: 36%) are the most frequent type of source for both aggregators. But only a small proportion of these posts refer to companies' official blog posts (r/p: 12%; HN: 13%) as the majority are from personal blogs (r/p: 88%; HN: 87%). Our interviewees also indicated they were motivated to take advantage of the aggregator's audience to publicize and gather more followers to their own blogs. For the second most common type of post, the aggregators have different characteristics. News websites are common for Hacker News (r/p: 4%; HN: 22%), and software project home pages or repositories are more common for r/programming (r/p: 12%; HN: 6%). Most of these home pages are hosted on GitHub (r/p: 86%; HN: 68%). Other common sources are static web pages (r/p: 14%; HN: 15%), other aggregator posts (r/p: 1%; HN: 8%), YouTube videos (r/p: 6%; HN: 3%), and Q&A sites (r/p: 2%; HN: 0%). We also found (less than 2% each) Wikipedia pages, online courses, crowd-funding platforms, Twitter posts, pictures, online tools, and other subreddits, among others. We observed similar trends in the quantitative analysis of the whole dataset. Medium.com, appears as the most shared domain on Hacker News (22,017 occurrences - 5.83%) and third most shared in r/programming (2,084 - 4.66%), and traditional newspapers are common in Hacker News (NYTimes=7,396 - 1.92%, bloomberg.com=4,293 - 1.11%, washingtonpost.com=2,249 - 0.58%), whereas in r/programming we found only 64 posts from these three newspapers.

**4.2.2 Overlap and differences.** For each post in our sample, we manually searched the title in the other aggregator (using double

quotes to avoid false positives). We observed that the overlap between the two news aggregators differs: 34% of the posts posted on r/programming were also posted on Hacker News; 5% of the posts posted on Hacker News were also posted on r/programming. We hypothesize that this difference happens due to breadth and depth. Since Hacker News favors broader content, few users might be tempted to re-post content on r/programming, which is a depth-oriented news aggregator (e.g., although technological relevant, the post might be out of scope for a programming-focused news aggregator). Whereas, content posted on r/programming is also suitable for re-posting on Hacker News.

By analyzing the complete dataset, we found that the five most frequent words in the titles of posts<sup>3</sup> shared on Hacker News are *show* (15,663 occurrences - 1.32%), *new* (13,288 - 1.12%), *app* (9,283 - 0.78%), *google* (8,857 - 0.74%), and *data* (8,841 - 0.74%). Whereas, the five most frequent words in r/programming are *program* (3,987 - 2.37%), *code* (2,977 - 1.77%), *develop* (2,564 - 1.53%), *learn* (1,661 - 0.99%), and *java* (1,624 - 0.96%). This concurs with our previous finding that Hacker News has more general terms while r/programming is more focused on programming.

Our interviewees share this impression, saying that r/programming seems to be more focused on technical aspects of software engineering, such as programming practices and frameworks, whereas Hacker News has a broader focus and touches on more general topics such as entrepreneurship, business, startups, funding, and financing (P1, P5-10, P16-19, P21). P18 provided an example: *"The second story now in Hacker News is about nuclear plants. It's nice, but it's not my main interest."* P21 also said that academic content, i.e., research papers, are better received by the Hacker News community. The difference in content is reflected in posts that go to the main page (P6, P22). This confirms the results of our qualitative analysis on the content *themes*, in which we found that around 50% of Hacker News posts are general topics, while they represent less than 15% of r/programming posts. Additionally, 54% of our survey respondents agreed that Hacker News and Reddit contain different content; however, 78% affirmed that, when the content is good, it will eventually appear in all news aggregators (Figure 2, Q16 and Q17).

Also concerning content, interviewees perceive comments on Hacker News to have better quality than the ones on Reddit (P1, P10, P12, P14, P16). They can be more insightful (P1), no single word comments or off-topic comments as one can find on Reddit (P1, P12), and more interactive (P16). Interestingly, P19 said that, when a post appears on both, he enjoys reading both comments, as they commonly have different perspectives. 35% of our survey respondents prefer comments in Hacker News over comments in Reddit, 27% did not agree with such preference and 38% were neutral (Figure 2, Q18).

**4.2.3 Content Reliability.** It is important to note that a known problem in any social network is the number of fake or wrong news [2, 11, 20]. However, from our interviews, we observed that such a problem is not perceived by our participants, mostly because technical articles are easier to trust, as bad opinions (and bad code) can be easily identified, and HOWTOs are straightforward as they only explain how to do something with a specific technology (P5-6,

<sup>3</sup>For this analysis, stop words were removed and words were stemmed using tm in R.



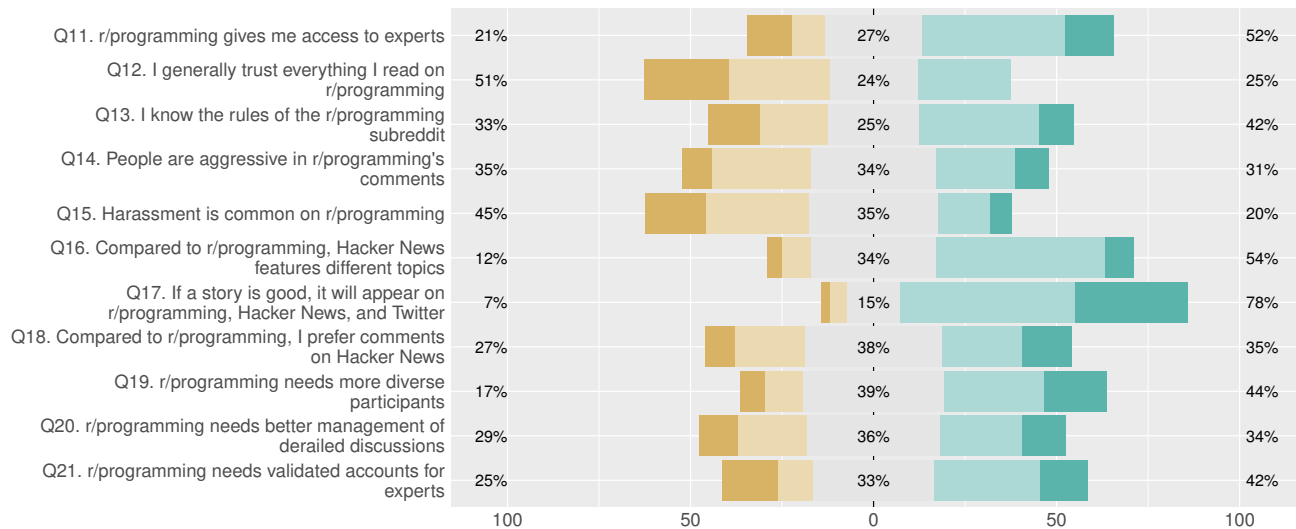


Figure 2: Responses to our survey (part 2).

P12, P14, P17, P21). Participants contrasted technical subreddits and politics-related subreddits: *"If I were on Reddit for some news or politics, I'd be more careful"* (P14). Even though they trust technical articles, they also indicate to read posts and comments with a grain of salt (P1, P4, P10-11, P13, P18). P1 said: *"[Posts and comments] sometimes are true, sometimes are not. You need to make up your mind about them."* P4 affirmed *"Ultimately, I am [the] judge."* As a way to measure the credibility of a post, participants read comments (P4, P7, P18), number of upvotes (P11), and evaluate if the source of the post is a trustworthy website or a well-known developer (P12-14, P19). None of our survey respondents strongly agreed with the idea of trusting all technical content one can find in r/programming: 25% agreed, 24% were neutral, and 51% disagreed (Figure 2, Q12).

### 4.3 Improving the community (RQ3)

Participants discussed how content and behavior is moderated, how the tools may be improved, insights on how karma is relevant, and pitfalls and advice.

**4.3.1 Moderation and Code of Conduct.** Participants suggested different types of moderation, such as to avoid users that are always negative (P7), not to let derailed discussions go to the top of the website (P9), to police (and punish) bad behavior (P13-14, P18), to reduce negative criticism (P20), and, in case of r/programming, to remove posts that are not entirely focused on programming (P19). P14 suggested the introduction of a code of conduct (cf. [58]): *"There's this website, similar to Reddit, and they introduced a code of conduct. If you don't follow, you are expelled from the site... Culture is important, you see in Stack Overflow that people are not aggressive. If you do something bad, you get negative points. When you have -5 points, your question gets removed."* The aggressiveness of users was mentioned by many of our interviewees (P3-7, P14, P17-18, P20).

**4.3.2 New Features.** To frequent contributors, news aggregators could provide detailed analytics on their posts as well as on the community as a whole (P2-3, P5, P15). In addition, according to P4, differentiating posts that require more thinking from others that

are easy to grasp may be helpful; according to him, complex posts attract fewer upvotes and tend to appear less on the front page.

Other suggestions were to provide credentials for experts in different fields (P10), add "subreddits" to Hacker News (P3), improve the user interface when it comes to showing large discussions (P5, P12), promote deeper content with more discussions (P17), and improve user diversity (P7). The survey participants agreed with most of these suggestions: 44% agreed that r/programming needs more participant diversity, 34% agreed on the need for better management of derailed discussions, and 42% would benefit from having validated credentials for experts (Figure 2, Q19, Q20, Q21).

**4.3.3 The role of karma.** For many, karma is not the main reason they post on r/programming (P5, P10, P12-14, P16, P18). P12, in particular, said that *"It's a nice number, right? But other than that, it does not make you better or give you more visibility, so I do not care. They are just internet points."* However, participants also see the indirect effects that karma brings to news aggregators: karma is a motivating factor for users to participate (P12, P15), users respect other users with high karma (P14), the number of karma points in a single post serves as a measurement of how good or interesting that link really is (P5, P8, P17-18), to track relevant users (P8), to track what posts attract more attention (P8), or to be able to create a new subreddit (which requires an undisclosed amount of karma points). 67% of our survey respondents did not believe that users with high karma are necessarily good developers. On the other hand, 35% of them saw karma as a way to identify relevant users. However, our respondents do not tend to follow users with high karma, as only 11% do so (Figure 1, Q8, Q9, and 10).

Interestingly, two participants indicated that, although karma points in Reddit are not so important, they care about their points in Stack Overflow (P13-14). P14 said that *"In Reddit, if you are rude and people find it funny, you get karma for basically being a troll. On the other hand, points in Stack Overflow show how competent you are, as one only gets points there for asking interesting questions or for providing good answers."*



**4.3.4 Pitfalls and Advice.** Our study participants shared several pieces of advice, ranging from how to behave on news aggregators to how to make sure a post has enough quality before publishing. To cope with the aggressiveness problem, readers and authors should not engage with trolls<sup>4</sup> (P12-13, P17). Also, they suggested to avoid being part of flame wars (P7, P10, P18). P18 advised: *"If you are in a heated conversation, take a step back and ask why you are in a heated conversation, and if you can possibly defuse it somehow. This is a skill that I think a lot of people don't practice."*

Other participants (P11-13, P16, P18, P21) indicated that they do not face many aggressive users, and when they do, they mostly ignore them. P13, in particular, said that programming-related news aggregators face fewer problems when compared to others (e.g., politics). Our survey respondents seem to have more mixed feelings: 31% of them believed that people are aggressive in r/programming, 34% were neutral, and 35% did not perceive people as aggressive. In addition, some survey respondents also see harassment as a problem: 20% of them believed this is a common behavior on r/programming, while 35% were neutral, and 45% did not think harassment is common (Figure 2, Q14-15). When participating in discussions, participants should be positive and always pose opinions constructively (P5-7, P11). P14 told us that a famous JavaScript developer asks his readers not to share his posts on Hacker News, mostly because of negative comments.

Participants recommended that readers watch for new trends (P5), avoid *clickbait*s<sup>5</sup> (P10), read the comments of a good post (P10, P12), post the primary source of the content rather than someone's version of it (P19), and be critical when it comes to applying something from the news aggregators elsewhere (P4, P18). To the latter, P4 said: *"If you are not critical, you can jump on this technology that people are talking about, and maybe [it is] not the best one."*

As a user who wants to share content, one should look for the right sub-community to follow and post into (P3, P16). In addition, users should know the rules of their news aggregator (P11, P13): some of them do not allow users to post specific content, or post more than 10% self-promotion. Curiously, several of our interviewees have been punished for that (P2, P5, P9, P22). Interestingly, only 42% of our survey respondents indicated to know the rules of the news aggregators (and subreddits) they visit (Figure 2, Q13).

## 5 RECOMMENDATIONS

Our research leads to recommendations for developers, technology stewards and researchers.

Some recommendations for software developers and users of news aggregators are:

- *Incorporate r/programming and Hacker News in the reading activities*, as these websites can be an efficient way to stay up-to-date, learn about technologies, and discuss different points of view on the practice of software development.
- *Be part of the community*, upvotes and comments enrich the information and ensure the sites succeed.

<sup>4</sup>An internet troll is a person who posts messages whose content solely intends to irritate other participants. Interestingly, P17 was the only participant who admitted acting like an internet troll sometimes.

<sup>5</sup>Clickbaits are links whose primary goal is to make users click and visit a web page. They usually have titles that instigate curiosity, but do not contain enough information.

- *Share findings and new ideas*, as the community will help to improve the content quality. However, before sharing, be sure that the content is coherent and be prepared to deal with feedback.

In communities of practice, technology stewardship plays an important role in ensuring community technologies do not impede the success of the community overall. These stewards need to consider how the technology or tools should be improved over time. Here we suggest several ways to improve the quality of the services provided by news aggregators:

- *A better moderation system* to deal with derailed discussions and avoid single worded comments. Moderators could also help to detect and flag aggressive posts and harassment, and foster constructive over negative feedback.
- *An analytics system* for frequent posters. Such users would benefit from knowing the characteristics of popular posts and the demographics of their readers. By providing them with more information, we expect them to further improve their content.
- *Provide different levels of upvoting* so that users can separately classify regular, good, and amazing posts. A future experiment could also be to see how users would behave without having the possibility of downvoting comments; Van Mieghem [60] has shown that the number of downvotes depends in power law fashion on the total number of upvotes in Reddit, and that, surprisingly, those downvotes increase faster than upvotes.

We also suggest future work that can be tackled by software engineering researchers:

- *Further understand the behavior of developers through news aggregators.* Our study provides a first step towards understanding the role of news aggregators in the software engineering community. We found that r/programming and Hacker News have become communities of practice for developers, and, as researchers who aim to improve the practice of software development, it is important to be aware of the discussions on news aggregators.
- Another opportunity is to study peripheral participation in news aggregators—how newcomers (or readers) may move from peripheral to central participation (posters or moderators) to maximize their own and others' learning (cf. Lave-Wenger [30]).
- *Look beyond formal documentation and Stack Overflow as sources for developer knowledge.* Although Reddit has been explored by researchers from other communities (e.g., [8, 21, 51, 60]), it is still not very popular in the software engineering research community. We suggest researchers to include Reddit and Hacker News in their datasets when studying developer knowledge sources. At a technology level, we also need to understand how these news aggregators impact the diffusion of innovation [45] in the software development domain. From a tool builder's perspective, we can attempt to improve, summarize, repackage, or filter the rich information available on these news aggregators.

## 6 LIMITATIONS

To be able to triangulate findings, we used a mix of research methods to answer our research questions, including semi-structured interviews, a validation survey, and content analysis. While we achieved saturation regarding the answers to our research questions, interviewing and surveying the individuals we selected may have biased our results. All interview and survey participants were

ultimately self-selected individuals within the respective target populations. Individuals who did not respond to our invitations may have different views on some of the questions that we asked. However, we did ensure to talk to individuals from different ethnic backgrounds before claiming saturation.

We identified candidates to interview by messaging active contributors in a single news aggregator (r/programming), which possibly reduces the generalizability of our results to other news aggregators. We did this because Hacker News does not offer a way of contacting its contributors. During the design of the study, we conjectured that active r/programming users would also use other news aggregators, and thus, would be able to contribute to our research by discussing more than one news aggregator. Indeed, all interviewees use more than one news aggregator.

We targeted active contributors in terms of number of posts for our interviews. We did this to ensure that we talked to core members of the r/programming community, but it is possible that less active members would have different perceptions. Our validation survey, however, mitigated this threat to validity.

The use of news aggregators may vary over time. We collected data from 01 January 2016 to 30 April 2017, and the interviews were conducted in June and July of 2017. Thus, our results reflect the use in this time frame. Our findings cannot be generalized to news aggregators other than r/programming and Hacker News. However, to the best of our knowledge, these two are the most active news aggregators used by software developers (cf. page views reported in Section 2). We also asked our interviewees about which news aggregators they used, and the majority reported using Reddit and Hacker News, with a few mentions of Lobsters [35].

## 7 RELATED WORK

There has only been a small number of studies investigating the role of programming-related news aggregators.

*News aggregators.* In early work on Slashdot, Lampe and Resnick found that Slashdot's mostly tech savvy, younger users, may be especially good at using moderation tools such as upvoting or downvoting [29]. This is confirmed by our findings: All of our interviewees felt comfortable using the moderation tools provided by the aggregators. More recent work has focused on Reddit: Mills [39] reports that "*as the number of active subreddits continues to grow, each with its own subject and driven by its own independent community, Reddit's offer to users shifts away from the once unifying, once unitary, 'Front page of the Internet', and towards a highly customizable 'Daily Me'.*" In our work, we have studied one such community, i.e., r/programming, and we found that it has become a daily source of information for many of our participants.

Other aspects of news aggregators that have been investigated include the role of re-submissions [21], the interplay between titles/content and popularity [28, 53], the interplay between user expertise and content quality [34], motivators [40], conversation patterns [12], design parameters [17], predictors of comment scores [64], and Reddit's evolution [51].

*News aggregators in software engineering research.* In more closely related research, news aggregators targeted at developers such as r/programming and Hacker News have been discussed in

the context of social coding where they provide a common source of information [15, 37]. Wu *et al.* [66] found that their survey respondents use Twitter to learn about ideas and converse with others while they use Hacker News to be aware of trending projects. We confirm these preliminary findings and shed more light on the role of news aggregators in software development.

Two recent efforts have explored potential links between Reddit and Stack Overflow. Bagheri and Ensan [6] annotated content from Stack Overflow and Reddit with semantic tags, which enabled them to recommend Reddit content for any Stack Overflow question. Hardin and Berland [25] compared links posted on Reddit and Stack Overflow to investigate how these communities use and talk about computer science learning resources, although their results are not available for inspection. We offer a similar comparison between content posted on Hacker News and r/programming.

*Twitter use by software developers.* Several researchers have investigated the role that Twitter can play in software engineering, ranging from explorations of its use by software developers [10, 61] to studies on tweet content [56], tweeting behavior [57], hot topics [48], automated classification [43, 65], prioritizing development-related tweets [22], and extracting URLs from tweets [49]. In a study on how developers stay current using Twitter, Singer *et al.* [50] surveyed and interviewed developers and found that Twitter helps them keep up with the fast-paced development landscape, stay aware of industry changes, learning, and building relationships. Our study complements their work by using a similar methodology to study programming-related news aggregators.

## 8 CONCLUSION

Although we knew before our research that news aggregators are popular and many see them as essential knowledge resources, we lacked insights on the role they play in software development and how they could be improved. To understand how these communities of practice shape and share knowledge using modern news aggregators, we interviewed and surveyed active contributors to r/programming and we analyzed the content of posts on Hacker News and r/programming using quantitative and qualitative methods. We characterized developers' motivations for participating, the challenges they face, and the content they read and share.

Programming-related news aggregators form a community of practice [63] for software developers. As P22 mentioned: "*Reddit is a community of practice inside the world.*" Developers now have a global place to discuss software engineering rather than with just a small group of colleagues. Such discussions are not new and happened before in other sites such as Usenet [62], but the size of the community, the scale of the interactions, and the new social features that emerge are disrupting and are likely to continue to change how development communities share and shape knowledge.

## ACKNOWLEDGMENTS

We thank the 99 survey respondents and the 22 interviewees for their availability. We also thank the reviewers for their valuable comments. This work is supported by the CNPq (406308/2016-0 and 430642/2016-4); PROPESP/UFGA; and FAPESP (2015/24527-3).

## REFERENCES

- [1] 18 Websites that every developer should visit right now [n. d.]. Best Developer Websites: Programming News, Tutorials & More. <https://stackify.com/18-websites-every-developer-should-visit-right-now/>. ([n. d.]).
- [2] Kayode Sakariyah Adewole, Nor Badrul Anuar, Amirrudin Kamsin, Kasturi Dewi Varathan, and Syed Abdul Razak. 2017. Malicious accounts: Dark of the social networks. *Journal of Network and Computer Applications* 79 (2017), 41–67. <https://doi.org/10.1016/j.jnca.2016.11.030>
- [3] Alexa Hacker News [n. d.]. Ycombinator.com Traffic, Demographics and Competitors - Alexa. <https://www.alex.com/siteinfo/ycombinator.com>. ([n. d.]). Verified 25 August 2017.
- [4] Alexa Reddit [n. d.]. Reddit.com Traffic, Demographics and Competitors - Alexa. <https://www.alex.com/siteinfo/reddit.com>. ([n. d.]). Verified 25 August 2017.
- [5] Mauricio Aniche, Christoph Treude, Igor Steinmacher, Igor Wiese, Gustavo Pinto, Margaret-Anne Storey, and Marco A. Gerosa. 2018. How Modern News Aggregators Help Development Communities Shape and Share Knowledge: Appendix. [online]. (2018). <https://www.zenodo.org/record/1167838#.WnnlJYG1vp>
- [6] Ebrahim Bagheri and Faezeh Ensani. 2016. Semantic Tagging and Linking of Software Engineering Social Content. *Automated Software Engineering* 23, 2 (2016), 147–190.
- [7] Titus Barik, Brittany Johnson, and Emerson Murphy-Hill. 2015. I Heart Hacker News: Expanding Qualitative Research Findings by Analyzing Social News Websites. In *Proceedings of the 10th Joint Meeting on Foundations of Software Engineering*. ACM, New York, NY, USA, 882–885.
- [8] Kelly Bergstrom. 2011. “Don’t feed the troll”: Shutting down debate about community expectations on Reddit.com. *First Monday* 16, 8 (2011).
- [9] Hudson Borges, Andre Hora, and Marco Tulio Valente. 2016. Understanding the factors that impact the popularity of GitHub repositories. In *Proceedings of the International Conference on Software Maintenance and Evolution*. IEEE Computer Society, Washington, DC, USA, 334–344.
- [10] Gargi Bougie, Jamie Starke, Margaret-Anne Storey, and Daniel M. German. 2011. Towards Understanding Twitter Use in Software Engineering: Preliminary Findings, Ongoing Challenges and Future Questions. In *Proceedings of the 2nd International Workshop on Web 2.0 for Software Engineering*. ACM, New York, NY, USA, 31–36.
- [11] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information Credibility on Twitter. In *Proceedings of the 20th International Conference on World Wide Web*. ACM, New York, NY, USA, 675–684.
- [12] Daejin Choi, Jinyoung Han, Taejoong Chung, Yong-Yeol Ahn, Byung-Gon Chun, and Ted Taekyoung Kwon. 2015. Characterizing Conversation Patterns in Reddit: From the Perspectives of Content Properties and User Participation Behaviors. In *Proceedings of the on Conference on Online Social Networks*. ACM, New York, NY, USA, 233–243.
- [13] Alceu Ferraz Costa, Yuto Yamaguchi, Agma Juci Machado Traina, Caetano Traina Jr., and Christos Faloutsos. 2017. Modeling Temporal Activity to Detect Anomalous Behavior in Social Media. *ACM Transactions on Knowledge Discovery from Data* 11, 4 (2017), 49:1–49:23.
- [14] Kevin Crowston and Ivan Shamshurin. 2017. Core-periphery communication and the success of free/libre open source software projects. *Journal of Internet Services and Applications* 8, 1 (jul 2017), 10. <https://doi.org/10.1186/s13174-017-0061-4>
- [15] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. 2012. Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository. In *Proceedings of the Conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 1277–1286.
- [16] Munmun De Choudhury and Sushovan De. 2014. Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*. AAAI Press, Menlo Park, CA, USA, 71–80.
- [17] Chrysanthos Dellarocas, Juliana Sutanto, Mihai Calin, and Elia Palme. 2015. Attention allocation in information-rich environments: the case of news aggregators. *Management Science* 62, 9 (2015), 2543–2562.
- [18] Digg [n. d.]. Digg - What the Internet is talking about right now. <http://digg.com/>. ([n. d.]). Verified 25 August 2017.
- [19] Feedly [n. d.]. <https://feedly.com/i/discover>. ([n. d.]). Verified 25 August 2017.
- [20] M. Fire, R. Goldschmidt, and Y. Elovici. 2014. Online Social Networks: Threats and Solutions. *IEEE Communications Surveys Tutorials* 16, 4 (Fourthquarter 2014), 2019–2036. <https://doi.org/10.1109/COMST.2014.2321628>
- [21] Eric Gilbert. 2013. Widespread Underprovision on Reddit. In *Proceedings of the Conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 803–808.
- [22] Emitza Guzman, Mohamed Ibrahim, and Martin Glinz. 2017. Prioritizing User Feedback from Twitter: A Survey Report. In *Proceedings of the 4th International Workshop on CrowdSourcing in Software Engineering*. IEEE Press, Piscataway, NJ, USA, 21–24.
- [23] Hacker News [n. d.]. Hacker News. <https://news.ycombinator.com/>. ([n. d.]). Verified 25 August 2017.
- [24] Hacker News Welcome [n. d.]. Hacker News: Welcome. <https://news.ycombinator.com/news/welcome.html>. ([n. d.]). Verified 25 August 2017.
- [25] Carolin D. Hardin and Matthew Berland. 2016. Learning to Program Using Online Forums: A Comparison of Links Posted on Reddit and Stack Overflow (Abstract Only). In *Proceedings of the 47th Technical Symposium on Computing Science Education*. ACM, New York, NY, USA, 723–723.
- [26] Joseph Henrich, Steven J. Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33, 2-3 (2010), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- [27] Siw Elisabeth Hove and Bente Anda. 2005. Experiences from Conducting Semi-structured Interviews in Empirical Software Engineering Research. In *Proceedings of the 11th International Software Metrics Symposium*. IEEE Computer Society, Washington, DC, USA, 23–.
- [28] Himabindu Lakkaraju, Julian McAuley, and Jure Leskovec. 2013. What’s in a Name? Understanding the Interplay between Titles, Content, and Communities in Social Media. In *Proceedings of the International AAAI Conference on Web and Social Media*. AAAI, Palo Alto, CA, USA, 311–320.
- [29] Cliff Lampe and Paul Resnick. 2004. Slash(Dot) and Burn: Distributed Moderation in a Large Online Conversation Space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 543–550.
- [30] Jean Lave and Etienne Wenger. 1999. Legitimate peripheral participation. *Learners, learning and assessment*, London: The Open University (1999), 83–89.
- [31] Alex Leavitt and Joshua A. Clark. 2014. Upvoting Hurricane Sandy: Event-based News Production Processes on a Social News Site. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1495–1504.
- [32] Angela M Lee and Hsiang Iris Chyi. 2015. The rise of online news aggregators: Consumption and competition. *International Journal on Media Management* 17, 1 (2015), 3–24.
- [33] Kristina Lerman and Tad Hogg. 2010. Using a Model of Social Dynamics to Predict Popularity of News. In *Proceedings of the 19th International Conference on World Wide Web*. ACM, New York, NY, USA, 621–630.
- [34] Wern Han Lim, Mark James Carman, and Sze-Meng Jojo Wong. 2017. Estimating Relative User Expertise for Content Quality Prediction on Reddit. In *Proceedings of the 28th Conference on Hypertext and Social Media*. ACM, New York, NY, USA, 55–64.
- [35] Lobsters [n. d.]. Lobsters. <https://lobste.rs/>. ([n. d.]). Verified 25 August 2017.
- [36] Lena Mamkina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. 2011. Design Lessons from the Fastest Q&A Site in the West. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2857–2866.
- [37] Jennifer Marlow, Laura Dabbish, and Jim Herbsleb. 2013. Impression Formation in Online Peer Production: Activity Traces and Personal Profiles in Github. In *Proceedings of the Conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 117–128.
- [38] Microsoft MVP [n. d.]. Microsoft MVP Award. <http://mvp.microsoft.com/>. ([n. d.]). Verified 25 August 2017.
- [39] Richard A. Mills. 2015. Reddit.Com: A Census of Subreddits. In *Proceedings of the Web Science Conference*. ACM, New York, NY, USA, 49:1–49:2.
- [40] Carrie Moore and Lisa Chuang. 2017. Redditors Revealed: Motivational Factors of the Reddit Community. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.
- [41] Kumiyo Nakakoji, Yasuhiro Yamamoto, Yoshiyuki Nishinaka, Kouichi Kishida, and Yunwen Ye. 2002. Evolution Patterns of Open-source Software Systems and Communities. In *International Workshop on Principles of Software Evolution (IWPSE ’02)*. ACM, New York, NY, USA, 76–85. <https://doi.org/10.1145/512035.512055>
- [42] Seyed Mehdi Nasehi, Jonathan Sillito, Frank Maurer, and Chris Burns. 2012. What Makes a Good Code Example?: A Study of Programming Q&A in StackOverflow. In *Proceedings of the International Conference on Software Maintenance*. IEEE Computer Society, Washington, DC, USA, 25–34.
- [43] Philips K. Prasetyo, David Lo, Palakorn Achananunparp, Yuan Tian, and Ee-Peng Lim. 2012. Automatic Classification of Software Related Microblogs. In *Proceedings of the International Conference on Software Maintenance*. IEEE Computer Society, Washington, DC, USA, 596–599.
- [44] Reddit [n. d.]. reddit: the front page of the internet. <https://www.reddit.com>. ([n. d.]). Verified 25 August 2017.
- [45] Everett M. Rogers. 2003. *Diffusion of innovations* (5th ed.). Free Press, New York, NY, USA.
- [46] r/programming [n. d.]. programming. <https://www.reddit.com/r/programming/>. ([n. d.]). Verified 25 August 2017.
- [47] Sample Size Calculator [n. d.]. Sample Size Calculator. <https://www.surveysystem.com/sscalc.htm>. ([n. d.]). Verified 25 August 2017.
- [48] Abhishek Sharma, Yuan Tian, and David Lo. 2015. What’s Hot in Software Engineering Twitter Space?. In *Proceedings of the International Conference on Software Maintenance and Evolution*. IEEE Computer Society, Washington, DC, USA, 541–545.
- [49] Abhishek Sharma, Yuan Tian, Agus Sulistya, David Lo, and Aiko Fallas Yamashita.

2017. Harnessing Twitter to support serendipitous learning of developers. In *Proceedings of the 24th International Conference on Software Analysis, Evolution and Reengineering*. IEEE Computer Society, Washington, DC, USA, 387–391.
- [50] Leif Singer, Fernando Figueira Filho, and Margaret-Anne Storey. 2014. Software Engineering at the Speed of Light: How Developers Stay Current Using Twitter. In *Proceedings of the 36th International Conference on Software Engineering*. ACM, New York, NY, USA, 211–221.
- [51] Philipp Singer, Fabian Flöck, Clemens Meinhardt, Elias Zeitfogel, and Markus Strohmaier. 2014. Evolution of Reddit: From the Front Page of the Internet to a Self-referential Community?. In *Proceedings of the 23rd International Conference on World Wide Web*. ACM, New York, NY, USA, 517–522.
- [52] Slashdot [n. d.]. Slashdot: News for nerds, stuff that matters. <https://slashdot.org/>. ([n. d.]). Verified 25 August 2017.
- [53] Greg Stoddard. 2015. Popularity and Quality in Social News Aggregators: A Study of Reddit and Hacker News. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, New York, NY, USA, 815–818.
- [54] Margaret-Anne Storey, Alexey Zagalsky, Fernando Figueira Filho, Leif Singer, and Daniel M. German. 2017. How Social and Communication Channels Shape and Challenge a Participatory Culture in Software Development. *IEEE Transactions on Software Engineering* 43, 2 (2017), 185–204.
- [55] Anselm Strauss and Juliet M Corbin. 1997. *Grounded theory in practice*. Sage.
- [56] Yuan Tian, Palakorn Achananuparp, Ibrahim Nelman Lubis, David Lo, and Ee-Peng Lim. 2012. What Does Software Engineering Community Microblog About?. In *Proceedings of the 9th Working Conference on Mining Software Repositories*. IEEE Press, Piscataway, NJ, USA, 247–250.
- [57] Yuan Tian and David Lo. 2014. An Exploratory Study on Software Microblogger Behaviors. In *Proceedings of the 4th Workshop on Mining Unstructured Data*. IEEE Computer Society, Washington, DC, USA, 1–5.
- [58] Parastou Tourani, Bram Adams, and Alexander Serebrenik. 2017. Code of conduct in open source projects. In *Proceedings of the 24th International Conference on Software Analysis, Evolution and Reengineering*. IEEE Computer Society, Washington, DC, USA, 24–33.
- [59] Christoph Treude, Ohad Barzilay, and Margaret-Anne Storey. 2011. How Do Programmers Ask and Answer Questions on the Web? (NIER Track). In *Proceedings of the 33rd International Conference on Software Engineering*. ACM, New York, NY, USA, 804–807.
- [60] Piet Van Mieghem. 2011. Human psychology of common appraisal: The Reddit score. *IEEE Transactions on Multimedia* 13, 6 (2011), 1404–1406.
- [61] Xiaofeng Wang, Ilona Kuzmickaja, Klaas-Jan Stol, Pekka Abrahamsson, and Brian Fitzgerald. 2014. Microblogging in Open Source Software Development: The Case of Drupal and Twitter. *IEEE Software* 31, 4 (2014), 72–80.
- [62] M McLure Wasko and Samer Faraj. 2000. "It is what one does": why people participate and help others in electronic communities of practice. *The Journal of Strategic Information Systems* 9, 2 (2000), 155–173.
- [63] Etienne Wenger. 1998. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, Cambridge, United Kingdom.
- [64] Tim Weninger, Xihao Avi Zhu, and Jiawei Han. 2013. An Exploration of Discussion Threads in Social News Sites: A Case Study of the Reddit Community. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*. ACM, New York, NY, USA, 579–583.
- [65] Grant Williams and Anas Mahmoud. 2017. Analyzing, Classifying, and Interpreting Emotions in Software Users' Tweets. In *Proceedings of the 2nd International Workshop on Emotion Awareness in Software Engineering*. IEEE Press, Piscataway, NJ, USA, 2–7.
- [66] Yu Wu, Jessica Kropczynski, Patrick C. Shih, and John M. Carroll. 2014. Exploring the Ecosystem of Software Developers on GitHub and Other Platforms. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, New York, NY, USA, 265–268.