Alice Pucher, Bsc.

TITLE

Master's Thesis

to achieve the university degree of

Master of Science

Master's degree programme: Computer Science

submitted to

Graz University of Technology

Supervisor

Dr. Some Body

Institute for Interactive Systems and Data Science
Head: Univ.-Prof. Dipl-Ing. Dr.techn. Some One

Graz, May 2020

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master‘s thesis.

Date	Signature

Abstract

This is a place-holder for the abstract. It summarizes the whole work to give a very short overview. Usually, this the abstract is written when the whole work text is finished. Alternatively, write an initial abstract in the beginning (wish how it should look like in the end), and then rewrite it at the end of the work.

The abstract consists of four parts, plus an optional beginning. 0., optionally start the abstract with 1-2 sentences about the background of the work. 1., describe the topic, or problem of the work. It should be clear what the aim of the work is. 2., describe how you solved the problem. It should be clear what steps were necessary to solve the problem. 3., give a short overview of the results. Should be clear how well the problem is solved. 4., give an outlook of what is now possible, since the problem has been solved.

The abstract is typically written in the past tense. It is uncommon to put references directly into the abstract.

Contents

List of Figures

1 Introduction

StackExchange is a Q&A platform and consists of 174 communities [1]. Each community evolves around a specific topic, for instance, StackOverflow focusing on software engeneering, or AskUbuntu focusing on the Ubuntu operating system. This distincts StackExchange from other Q&A sites such as *Yahoo! Answers* where no such differentiation into topics exists.

In August of 2018, the StackExchange team introduced a small change which may have had a huge impact on the platform. They added a new feature to visibly mark questions from new contributors, as part of their effort to make the site more welcoming for new users [2]. Specifically members who want to answer a question created by a new contributor are shown a notification in the answer box that this question is from a new contributor. The StackExchange team hopes that this little change encourages members to be more friendly and forgiving toward new users.

This thesis evaluates whether this change has a real impact on the community and if so in which direction the community reacts. For this analysis, this thesis utilizes Vader [3], a sentiment analysis tool, to quantify the sentiments of the answers submitted to questions of new contributors. An interrupted time series is then applied to these values to evalutate whether the change achieved its purpose of making the platform more welcoming.

This thesis investigates the ten largest communities of the StackExchange platform measured by number of posts. This includes StackOverflow, MathOverflow, Math, AskUbuntu, SuperUser, and some lesser known communities.

The remaining part of this thesis is structured as follows: Section 2 explains StackExchange, how it works, and shows related work. Section 3 shows the method this thesis uses for analysis. Section 4 contains the investigated datasets. Results are presented in Section 5 and discussed in Section 6. Section 7 conculdes this thesis.

2 Related Work

This section is divided into two parts. The first part explains what StackExchange is, how it developed since its inception, and how it works. The second part shows previous and related work.

2.1 Background

StackExchange[1] is a community question and answering (CQA) platform where users can ask and answer questions, accept answers as an appropriate solution to the question, and up-/downvote questions and answers. StackExchange uses a community-driven knowledge creation process by allowing everyone who registers to participate in the community. Invested users also get access to moderation tools to help maintain the vast community. All posts on the StackExchange platform are publicly visible, allowing non-users to benefit from the community as well. Posts are also accessible for web search engines so users can find questions and anwsers easily with a simple web search. StackExchange keeps an archive of all questions and answers posted, creating a knowledge archive for future visitors to look into. Originally, StackExchange started with StackOverflow[2] in 2008 [4]. Since then StackExchange grew into a platform hosting sites for 174 different topics [1], for instance, programming (StackOverflow), maths (MathOverflow[3] and Math StackExchange[4]), and typesetting (TeX/LaTeX[5]). Questions on StackExchange are stated in natural English language and consist of a title, a body containing a detailed description of the problem or information need, and tags to categorize

[1] https://stackexchange.com
[2] https://stackoverflow.com
[3] https://mathoverflow.net
[4] https://math.stackexchange.com
[5] https://tex.stackexchange.com

the question. After a question is posted the community can submit answers to the question. The author of the question can then accept an appropriate answer which satisfies their question. The accepted answer is then marked as such with a green checkmark and shown on top of all the other answers. Figure 2.1 shows an example of a StackOverflow question. Questions and answers can be up-/downvoted by every user registered on the site. Votes typically reflect the quality and importance of the respective question or answers. Answers with a high voting score raise to the top of the answer list as answers are sorted by the vote score in descending order by default. Voting also influences a user's reputation [1, 5]. When a post (question or answers) is voted upon the reputation of the poster changes accordingly. Furthermore, downvoting of answers also decreases the reputation of the user who voted [6]. Reputation on StackExchange indicates how trustworthy a user is. To gain a high reputation value a user has to invest a lot of time and effort to reach a high reputation value by asking good questions and posting good answers to questions. Reputation also unlocks privileges which may differ slightly from one community to another [7, 8]. With privileges, users can, for instance, create new tags if the need for a new tag arises, cast votes on closing or reopening questions if the question is off-topic or a duplicate of another question, or when a question had been closed for no or a wrong reason, or even get access to moderation tools. StackExchange also employs a badge system to steer the community [9]. Some badges can be obtained by performing one-time actions, for instance, reading the tour page which contains necessary details for newly registered users, or by performing certain actions multiple times, for instance, editing and answering the same question within 12 hours. Furthermore, users can comment on every question and answer. Comments could be used for further clarifying an answer or a short discussion on a question or answer. For each community on StackExchange, a *M*eta page is offered where members of the respective community can discuss the associated community [10, 11]. This place is used by site admins to interact with the community. The *M*eta pages are also used for proposing and voting on new features and reporting bugs. *M*eta pages run the same software as the normal CQA pages so users on vote the ideas and suggestions in the same way they would do on the actual CQA sites.

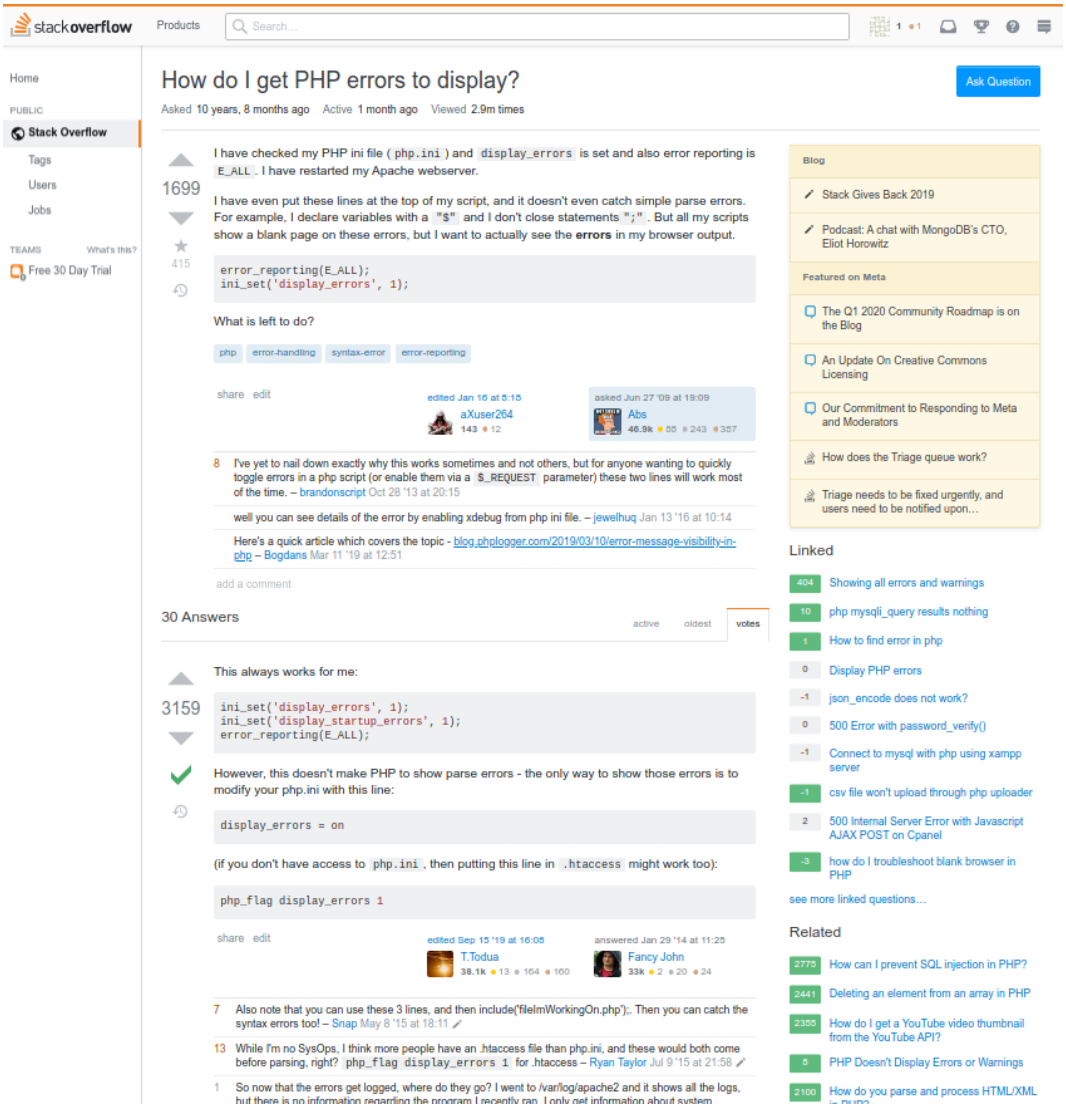


Figure 2.1: A typical question on StackOverflow. In the top middle section of the page, the question is stated. The question has 4 tags and 3 comments attached to it. Beneath the question, all answers are listed by their score in descending order (only one answer is visible in this screenshot). The accepted answer is marked by a green checkmark. To the left of the question and answers, the score (computed via votes) is indicated.

2.2 State of the Art

Since the introduction of Web 2.0 and the subsequential spawning of platforms for social interaction, researchers started investigating the emerging online communities. Research strongly focuses on the interactions of users on various platforms. Community knowledge platforms are of special interest, for instance, StackExchange/StackOverflow [12, 13, 14, 5, 15, 16, 17, 18, 19, 20], Quora [21], Reddit [22, 23], Yahoo! Answers [24], and Wikipedia [25]. These platforms allow communication over large distances and facilitate fast and easy knowledge exchange and aquisition by connecting thousands or even millions of users and create valuable repositories of knowledge in the process. Users create, edit, and consume little pieces of information and collectively build a community and knowledge repository. However, not every piece of information is factual [21, 24] and platforms often employ some kind of moderation to keep up the value of the platform and to ensure a certain standard within the community.

All these communities differ in their design. Wikipedia is a community-driven knowledge repository and consists of a collection of articles. Every user can create an article. Articles are edited collaboratively and continually improved an expanded. Reddit is a platform for social interaction where users create posts and comment on other posts or comments. Quora, StackExchange, and Yahoo! Answers are community questions and answer (CQA) platforms. On Quora and Yahoo! Answers users can ask any question regarding any topics whereas on StackExchange users have to post their questions in the appropriate subcommunity, for instance, StackOverflow for programming related questions or MathOverflow for math related questions. CQA sites are very effective at code review [26]. Code may be understood in the traditional sense of source code in programming related fields but this also translates to other fields, for instance, mathematics where formulas represent code. CQA sites are also very effective at solving conceptual questions. This is due to the fact that people have different areas of knowledge and expertise [27] and due to the large user base established CQA sites have, which again increases the variety of users with experise in different fields.

Despite the differences in purpose and manifestation of these communities, they are social communities and they have to follow certain laws. In their book on ”Building successful online communities: Evidence-based social design“ [28] Kraut lie out five equally important criteria online platforms have to fulfill in order to thrive. 1)

When starting a community has to have a critical mass of users who create content. StackOverflow already had a critical mass of users from the beginning due to the StackOverflow team already being experts in the domain [11] and the private beta [4]. Both aspects ensured a strong community core early on. 2) The platform must attract new users to grow as well as to replace leaving users. Depending on the type of community new users should bring certain skills, for example, programming background in open source software developement, or extended knowledge on certain domains; or qualities, for example, a certain illness in medical communities. New users also bring the challenge of onboarding with them. Most newcomers will not be familiar with all the rules and nuances of the community [25, 29]. 3) The platform should encourage users to commit to the community. Online communities are often based on voluntary commitment of their users [30], hence the platform has to ensure users are willing to stay. Most platforms do not have contracts with their users, so users should see benefits for staying with the community. 4) Contribution by users to the community should be encouraged. Content generation and engagement are the backbone of an online community. 5) The community needs regulation to sustain it. Not every user in a community is interested in the wellbeing of the community. Therefore, every community has to deal with trolls and inappropriate or even destructive behavior. Rules need to be established and enforced to limit and mitigate the damage malicious users cause.

All these criteria are heavily intertwined. Attracting new users often depends on the welcomingness and support of users that are already on the platform. Keeping users commited to the platform depends on the engagement with the community and how well the system design supports this. For the purpose of this thesis, the criteria layed out by Kraut and Resnick can be grouped into two main categories: 1) onboarding of new users, 2) keeping users engaged, contributing, and well behaved.

2.2.1 Onboarding of new users

The onboarding process is a permanent challenge for online communities and differs from one platform to another. Slag, Waard and Bacchelli investigated why many users on StackOverflow only post once after their registration [12]. They found that 47% of all users on StackOverflow posted only once and called them one-day-flies. They suggest that code example quality is lower than that of more involved users, which often leads to answers and comments to first improve the

question and code instead of answering the stated question. This likely discourages new users from using the site further. Negative feedback instead of constructive feedback is another cause for discontinuation of usage. The StackOverflow staff also conducted their own research on negative feedback of the community [31]. They investigated the comment sections of questions by recruiting their staff members to rate a set of comments and they found more than 7% of the reviewed comments are unwelcoming.

One-day-flies are not unique to StackOverflow. Steinmacher et al. investigated the social barriers newcomers face when they submit their first contribution to an open-source software project [32]. They based their work on empirical data and interviews and identified several social barriers preventing newcomers to place their first contribution to a project. Furthermore, newcomers are often on their own in open source projects. The lack of support and peers to ask for help hinders them. Yazdanian et al. found that new contributors on Wikipedia face challenges when editing articles. Wikipedia hosts millions of articles [33] and new contributors often do not know which articles they could edit and improve. Recommender systems can solve this problem by suggesting articles to edit but they suffer from the cold start problem because they rely on past user activity which is missing for new contributors. Yazdanian et al. proposed a solution by establishing a framework that automatically creates questionnaires to fill this gap. This also helps matching new contributors with more experienced contributors that could help newcomers when they face a problem. Allen showed that the one-time-contributors phenomenon also translates to workplaces and organizations [34]. They found out that socialization with other members of an organization plays an important role in turnover. The better the socialization within the organization the less likely newcomers are to leave. This socialization process has to be actively pursued by the organization. One-day-flies may partially be a result of lurking. Lurking is consuming content generated by a community but not contributing content to it. Nonnecke, Andrews and Preece investigated lurking behavior on Microsoft Network (MSN) [35] and found that contrary to previous studies lurking is not necessarily a bad behavior. Lurkers show passive behavior and are more introverted and less optimistic than actively posting members of a community. Previous studies suggested lurking is free riding, a taking-rather-than-giving process. However, the authors found that lurking is important in getting to know a community, how a community works and learning the nuances of social interactions on the platform. This allows for better integration into the community when a person decides to join the community.

StackExchange, and especially the StackOverflow community, probably has a large lurking audience. Many programmers do not register on the site and those who do only ask one question and revert to lurking, as suggested by [12].

The StackOverflow team acknowledged the one-time-contributors trend [29, 31] and took efforts to make the site more welcoming to new users [36]. They lied out various reasons: Firstly, they have sent mixed messages whether the site is an expert site or for everyone. Secondly, they gave too little guidance to new users which resulted in poor questions from new users and in the unwelcoming behavior of more integrated users towards the new users. New users do not know all the rules and nuances of communication of the communities. An example is that ”Please“ and ”Thank you“ is not well received on the site as they are deemed unnecessary. Also the quality, clearness and language quality of the questions of new users is lower than more experienced users which leads to unwelcoming or even toxic answers and comments. Moreover, users who gained moderation tool access could close questions with predefined reasons which often are not meaningful enough for the poster of the question [37]. Thirdly, marginalized groups, for instance, women and people of color [29, 38, 39], are more likely to drop out of the community due to unwelcoming behavior from other users [29]. They feel the site is an elitist and hostile place. The team suggested several steps to mitigate these problems. Some of these steps include appealing to the users to be more welcoming and forgiving towards new users [29, 31, 40], other steps are geared towards changes to the platform itself: The *Be nice policy* (code of conduct) was updated with feedback from the community [41]. This includes: new users should not be judged for not knowing all things. Furthermore, the closing reasons were updated to be more meaningful to the poster, and questions that are closed are shown as ”on hold“ instead of ”closed“ for the first 5 days [37]. Furthermore, the team investigates how the comment sections can be improved to lessen the unwelcomeness and hostility and keep the civility up.

The StackOverflow team partnered with Ford et al. and implemented the Mentorship Research Project [13, 42]. The project lasted one month and aimed to help newcomers improve their first questions before they are posted publicly. The program went as follows: When a user is about to post a question the user is asked whether they want their question to be reviewed by a mentor. If they confirmed they are forward to a help room with a mentor who is an experienced user. The question is then reviewed and the mentor suggests some changes if applicable. These changes may include narrowing the question for more precise answers,

adding a code example or adjusting code, or removing of *P*lease and *Thank you* from the question. After the review and editing, the question is posted by publicly the user. The authors found that mentored questions are received significantly better by the community than non-mentored questions. The questions also received higher scores and were less likely to be off-topic and poor in quality. Furthermore, newcomers are more comfortable when their question is reviewed by a mentor. For this project four mentors were hand selected and therefore the project would not scale very well as the number of mentors is very limited but it gave the authors an idea on how to pursue their goal of increasing the welcomingness on StackExchange. The project is followed up by a *Ask a question wizard* to help new users as well as more experienced users improve the structure, quality, and clearness of their questions [36].

Unwelcomeness is a large problem on StackExchange [29, 36, 39]. Although unwelcomeness affects all new users, users from marginalized groups suffer significantly more [29, 43]. Ford et al. investigated barriers users face when contributing to StackOverflow. The authors identified 14 barriers in total hindering newcomers to contribute and five barriers were rated significantly more problematic for women than men. On StackOverflow only 5.8% (2015 [44], 7.9% 2019 [38]) of active users identify as women. David and Shapiro found similar results of 5% women in their work on *Community-based production of open-source software* [45]. These numbers are comparatively small to the number of degrees in Science, Technology, Engineering, and Mathematics (STEM) [46] where 20% are achieved by women [47]. Despite the difference, the percentage of women on StackOverflow has increased in recent years.

2.2.2 Keeping users engaged, contributing and well behaved

Reputation plays a important role on StackExchange and indicates the credibility of a user as well as a primary source of answers of high quality [5]. Although the largest chunk of all questions is posted by low-reputated users, high-reputated users post more questions on average. To earn a high reputation a user has to invest a lot of effort and time into the community, for instance, asking good questions or providing useful answers to questions of others. Reputation is earned when a question or answer is upvoted by other users, or if an answer is accepted as the solution to a question by the question creator. Mamykina et al. found that the reputation system

of StackOverflow encourages users to compete productively [11]. But not every user participates equally, and participation depends on the personality of the user [14]. Bazelli, Hindle and Stroulia showed that the top-reputated users on StackOverflow are more extroverted compared to users with less reputation. Movshovitz-Attias et al. found that by analyzing the StackOverflow community network, experts can be reliably identified by their contribution within the first few months after their registeration. Graph analysis also allowed the authors to find spamming users or users with other extreme behavior. Although gaining reputation takes time and effort, users can take certain advantages to gain reputation faster by gaming the system [15]. Bosu et al. analyzed the reputation system and found five strategies to increase the reputation in a fast way: Firstly, answering questions with tags that have a small expertise density. This reduces competitiveness against other users and increases the chance of upvotes and answer acceptance. Secondly, questions should be answered promptly. The question asker will most likely accept the first arriving answer that solves the question. This is also supported by [48]. Thirdly, answering first also gives the user an advantage over other answerers. Fourthly, activity during off-peak hours reduces the competition from other users. Finally, contributing to diverse areas will also help in developing a higher reputation.

Complementary to the reputation system StackOverflow also employs a badge system [9] to stimulate contributions by users [49]. The goal of badges is to keep users engaged with the community [50]. Therefore, badges are often used in a gamification setting where users contribute to the community and are rewarded for their behavior if it alignes with the requirements of the badges. Badges are visible in questions and answers as well as the profile page of the user and can be earned by performing certain actions. Badges are often seen as a steering mechanism by researchers [16, 17, 18]. Although users want to achieve badges and are therefore steered to perform certain actions, steering also occurs in the reputation system. However, badges allow a wider variety of goals, for instance, asking and answering questions, voting on questions and answers, or writing higher quality answers. Badges also work as a motivator for users [18]. Users often put in non-trivial amounts of work and effort to achieve badges and so badges become powerful incentives. However, not all users are equal and therefore do not pursue badges in the same way [16]. Contrary to [18], Yanovsky et al. [16] found that users do not necessarily increase their activity prior to achieving a badge followed by an immediate decrease in contribution thereafter but users behave differently based on their type of contribution. The authors found users can be categorized into

three groups: Firstly, some users are not affected at all by the badge system and still contribute a lot to the community. Secondly, users increase their activity too before gaining a badge and keep their level of contribution afterward. Finally, users increase their activity before achieving a badge and return to their previous level of engagement thereafter. Different badges also create status classes [19]. The harder a badge can be earned by users, the more unique it is within the community and therefore the badge symbolizes some sort of status. Often rare badges are hard to achieve and take significant effort. For some users, depending on their type, this can be a huge motivator. Kusmierczyk and Gomez-Rodriguez found first-time badges play an important role in steering users [17]. The steering effect only takes place if the benefit to the user is greater than the effort the user has to put into to obtain the badge. If the effort is greater the user will likely not pursue the badge and therefore the steering effect will not occur.

Quality is often a concern in online communities. Platform moderators and admins want to keep a certain level of quality or even raise it. However, higher-quality posts take more time and effort than lower-quality posts. In the case of CQA platforms, this is an even bigger problem as higher quality posts fight against fast responses. Despite that, StackOverflow also has a problem with low quality and effort questions and subsequent unwelcoming answers and comments [31]. StackOverflow has grown into a large community and larger communities are harder to control. Lin et al. investigated how growth affects a community. They looked at Reddit communities that were added to the default set of subscribed communities of every new user (defaulting) which lead to a huge influx of new users to these communities as a result. The authors found that contrary to expectations, the quality stays largely the same. The vote score dips shortly after defaulting but quickly recovers or even raises to higher levels than before. The complaints of low-quality content did not increase, and the language used in the community stayed the same. However, the community clustered around fewer posts than before defaulting. Tausczik and Pennebaker found reputation is linked to the perceived quality of posts in multiple ways [20]. They suggest reputation could be used as an indicator of quality. Quality also depends on the type of platform. [22] showed that expert sites who charge fees, for instance, library reference services, have higher quality answers compared to free sites. Also, the higher the fee the higher the quality of the answers. However, free community sites outperform expert sites in terms of answer density and responsiveness.

3 Method

StackExchange introduced a *new contributor* indicator to all communities on 21^{st} of August in 2018 at 9 pm UTC [2]. This step is one of many StackExchange took to make the platform and its members more welcoming towards new users. This indicator is shown to potential answerers in the answer text box of a question flagged as from a new contributor as shown in figure 3.1. The indicator is added to a question if the question is the first contribution of a user or if the first contribution (question or answer) of the user was less than 7 days ago [51]. The indicator is then shown for 7 days from the creation date of the question. Note that the user can be registered for a long time and then post their first question and it is counted as a question from a new contributor. Also, if a user decides to delete all their contributions from the site and then creates a new question this question will have the *new contributor* indicator attached. The sole deciding factor for the indicator is the date and time of the first non-deleted contribution and the 7-day window afterward.

To measure the effectiveness of the change this thesis utilizes Vader, a sentiment analysis tool with exceptional performance in analysing and categorizing microblog-like texts as well as good generalization in other domains [3]. The choice is based on the speed and simplicity of Vader. Vader uses a lexicon of words with attached sentiment values and rules related to grammar and syntax to determine a sentiment value between -1 and 1 to a given piece of text. The sentiment range is divided into 3 classes: negative (-1 to -0.05), neutral (-0.05 to 0.05), and positive (0.05 to 1). The outer edges of the value space are rarely reached as the text would have to be extremely negative or positive which is very unlikely. This design allows fast and verifiable analysis.

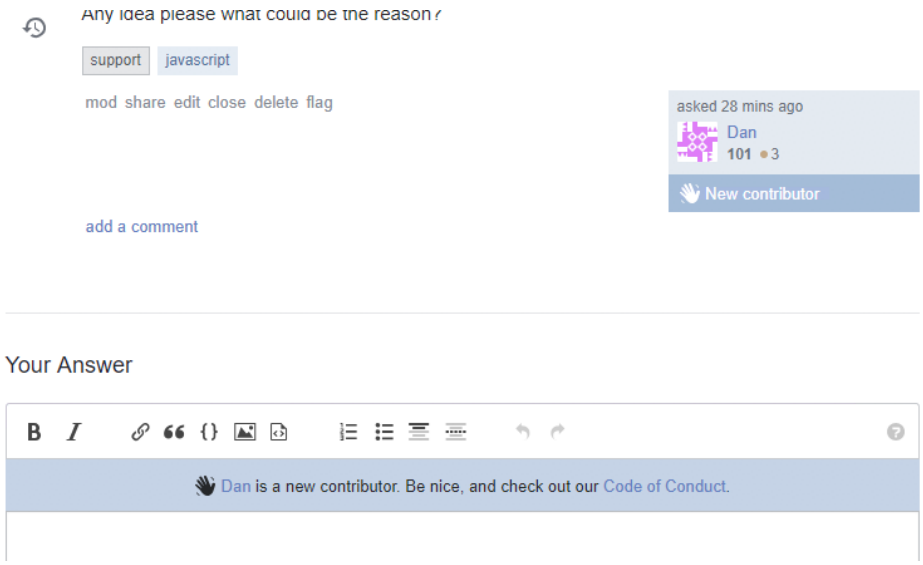


Figure 3.1: The answer box a potential answerers sees when viewing a question from a new contributor. ©Tim Post, 2018, https://meta.stackexchange.com/users/50049/tim-post in [2]

3.1 Data gathering and preprocessing

StackExchange provides anonymized data dumps of all their communities for researchers to investigate at no cost on archive.org [52]. These data dumps contain users, posts (questions and answers), badges, comments, tags, votes, and a post history containing all versions of posts. Each entry contains the necessary information, for instance, id, creation date, title, body, and how the data is linked together (which user posted a question/answer/comment). However, not all data entries are valid and therefore cannot be used in the analysis, for instance, questions or answers of which the user is unknown but this only affects a very small amount entries. So before the actual analysis, the data has to be cleaned. Moreover, the answer texts are in HTML format, containing tags that could skew the sentiment values, and they need to be stripped away beforehand. Additionally, answers may contain code sections which also would skew the results and are therefore omitted.

After preprocessing the raw data, relevant data is filtered and computed. Questions and answers in the data are mixed together and have to be separated and answers have to be linked to their questions. Also, questions in these datasets do not have the *new contributor* indicator attached to them and neither do users. So, the first

contribution date and time of users have to be calculated via the creation dates of the questions and answers the user has posted. Then, questions are filtered per user and by whether they are created within the 7-day window after the first contribution of the user. These questions were created during the period where the *new contributor* indicator would have been displayed, in case the questions had been posted before the change, or has been displayed after the change. From these questions, all answers which arrived within the 7-day window are considered for the analysis. Answers which arrived at a later point are excluded as the answerer most likely has not seen the disclaimer shown in figure 3.1. Included answers are then analyzed with Vader and the resulting sentiments are stored.

3.2 Analysis

An interrupted time series (ITS) analysis captures trends before and after a change in a system and fits very well with the question this thesis investigates. ITS can be applied to a large variety of data if the data contains the same kind of data points before and after the change and when the change date and time are known. Bernal, Cummins and Gasparrini published a paper on how ITS works [53]. ITS performes well on medical data, for instance, when a new treatment is introduced ITS can visualize if the treatment improves a condition. For ITS no control group is required and often control groups are not feasible. ITS only works with the before and after data and a point in time where a change was introduced. ITS relies on linear regression and tries to fit a three-segment linear function to the data. The authors also described cases where more than three segments are used but these models quickly raise the complexity of the analysis and for this thesis a three-segment linear regression is sufficient. The three segments are lines to fit the data before and after the change as well as one line to connect the other two lines at the change date. Figure 3.2 shows an example of an ITS. Each segment is captured by a tensor of the following formula $Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 T X_t$, where T represents time as a number, for instance, number of months since the start of data recording, X_t represents 0 or 1 depending on whether the change is in effect, β_0 represents the value at $T = 0$, β_1 represents the slope before the change, β_2 represents the value when the change is introduced, and β_3 represents the slope after the change. Contrary to the method in [53] where the ITS is performed on aggregated values per month, this thesis performs the ITS on single data points, as

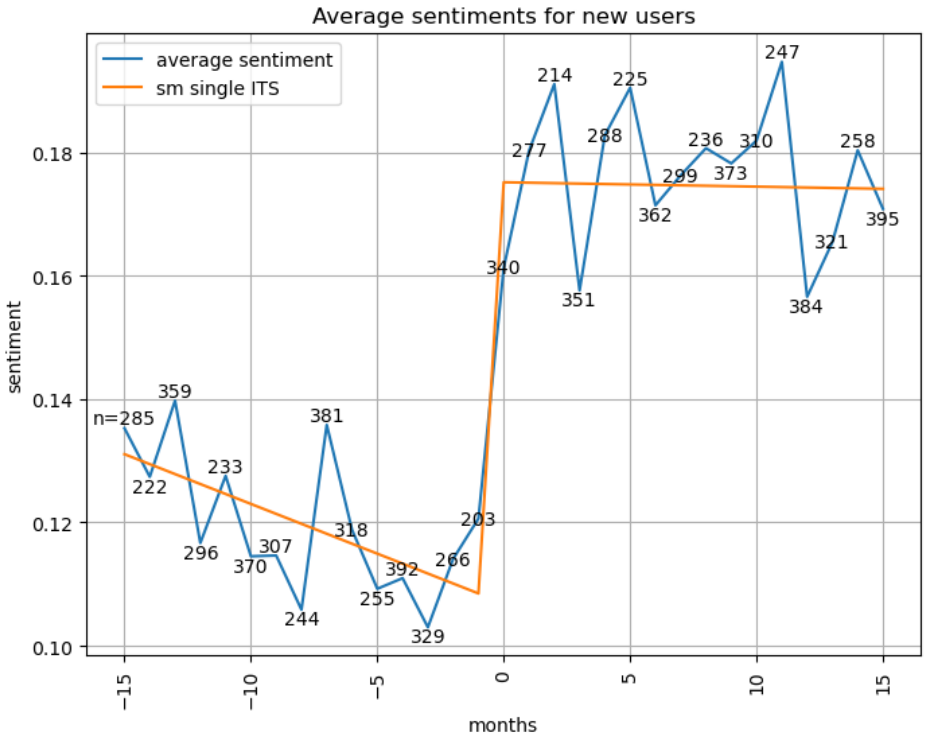


Figure 3.2: An example that visualizes how ITS works. The change of the system occurs at month 0. The blue line shows the average sentiment of fictional answers grouped by month. The numbers attached to the blue line show the number of sentiment values for a given month. The yellow line represents the ITS analysis as a three-segment line. This exmaple shows the expected behavior of the data sets in the following sections.

the premise that the aggregated values all have the same weight within a certain margin is not fulfilled. Performing the ITS with aggregated values would skew the linear regression more towards data points with less weight. Single data point fitting prevents this, as weight is taken into account with more data points.

4 Datasets

StackExchange provides complete datasets of its communities for research purposes on archive.org [52]. StackExchange also provides a short guide on how to interpret the provided data, as some data values are strictly numerical and do not convey any meaning without the knowledge of what these values represent. This thesis investigates the largest datasets available and includes the datasets of the following communities:

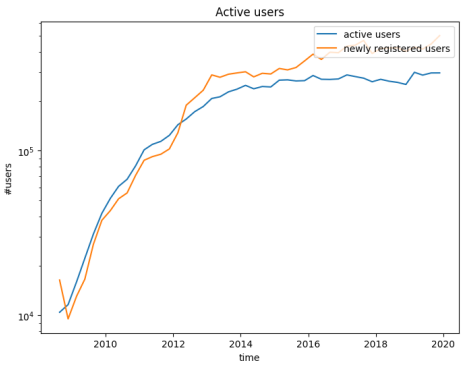
- StackOverflow.com
- math.stackexchange.com
- MathOverflow.net
- AskUbuntu.com
- ServerFault.com
- SuperUser.com
- electronics.stackexchange.com
- stats.stackexchange.com
- tex.stackexchange.com
- unix.stackexchange.com

These datasets are selected due to their size as larger datasets yield more consistent results. Smaller datasets may be too sparse to take any meaningful conclusions. Also, outliers would influence the results more when compared to outliner in bigger datasets. The dataset contain all the necessary data since the creation of the respective community and until the last day of February 2020.

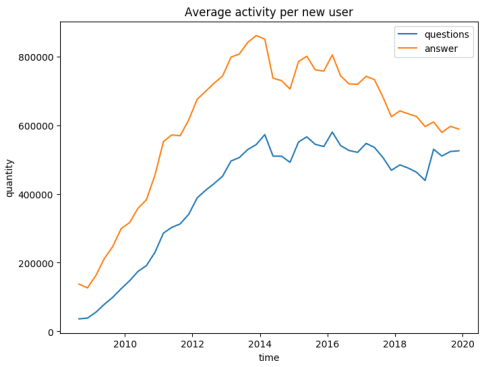
4.1 StackOverflow.com

StackOverflow is the largest and oldest community of the StackExchange platform. The community has 11867244 registered users of which 297192 were active between

December 2019 and February 2020. Members asked 18699974 questions in total and gave 27981749 answers with an average answer density of 1.496 answers per question. New users asked 2880039 questions with an average of 1.240 questions per new user during their first week after their first contribution.



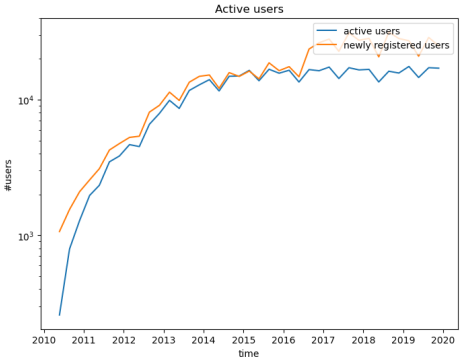
(a) Active users with activity in the last 3 months



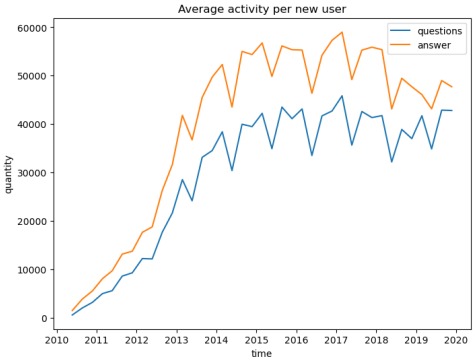
(b) Questions and answers counts over time

4.2 math.stackexchange.com

"Mathematics Stack Exchange is a question and answer site for people studying math at any level and professionals in related fields." [54] The community has 624671 registered users of which 17074 were active between December 2019 and February 2020. Members asked 1170938 questions in total and gave 1565188 answers with an average answer density of 1.336 answers per question. New users asked 265704 questions with an average of 1.336 questions per new user during their first week after first contribution.



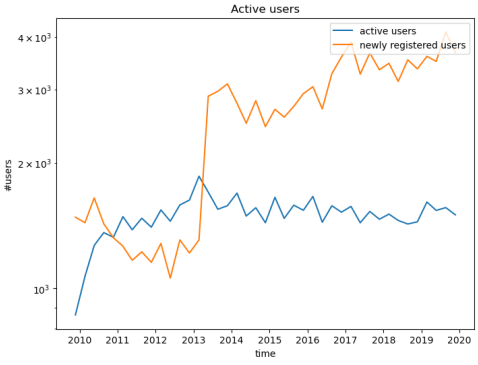
(a) Active users with activity in the last 3 months



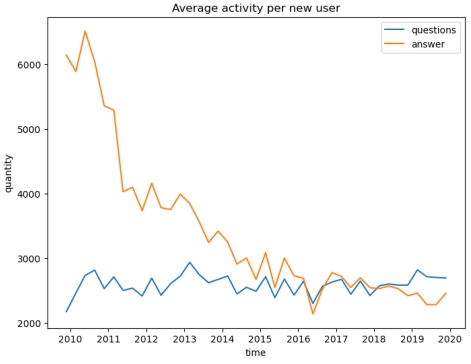
(b) Questions and answers counts over time

4.3 MathOverflow.net

MathOverflow.net is a rather small community for professional mathematicians. The community has 105471 registered users of which 1501 were active between December 2019 and February 2020. Members asked 108083 questions in total and gave 144918 answers with an average answer density of 1.34 answers per question. New users asked 23746 questions with an average of 1.131 questions per new user during their first week after first contribution.



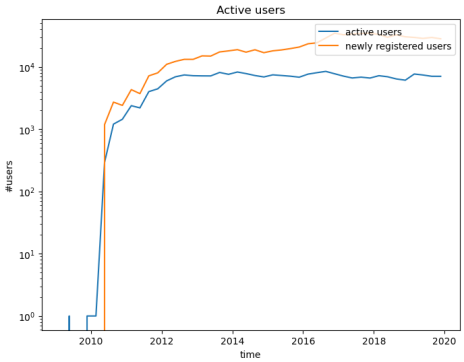
(a) Active users with activity in the last 3 months



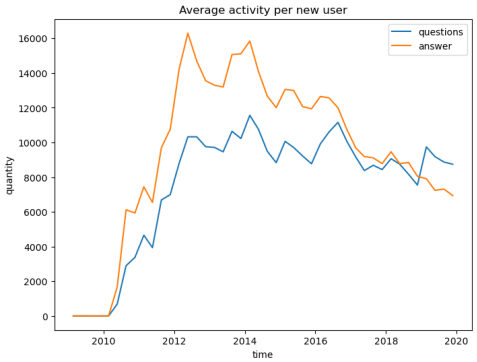
(b) Questions and answers counts over time

4.4 AskUbuntu.com

AskUbuntu.com is a rather small community for Ubuntu users and developers. The community has 783614 registered users of which 7033 were active between December 2019 and February 2020. Members asked 334194 questions in total and gave 418051 answers with an average answer density of 1.25 answers per question. New users asked 157018 questions with an average of 1.101 questions per new user during their first week after first contribution.



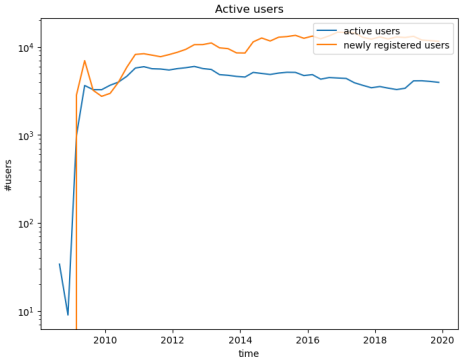
(a) Active users with activity in the last 3 months



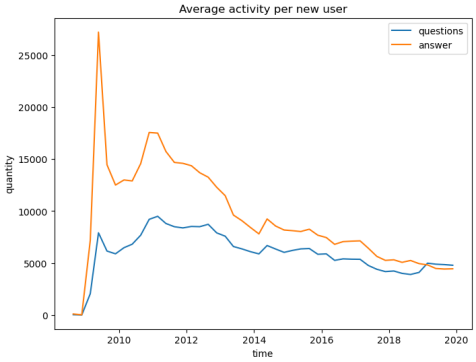
(b) Questions and answers counts over time

4.5 ServerFault.com

ServerFault.com is a rather small community for system and network administrators. The community has 451180 registered users of which 3947 were active between December 2019 and February 2020. Members asked 274564 questions in total and gave 432334 answers with an average answer density of 1.574 answers per question. New users asked 88547 questions with an average of 1.106 questions per new user during their first week after first contribution.



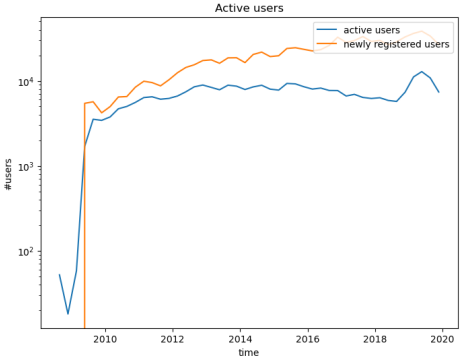
(a) Active users with activity in the last 3 months



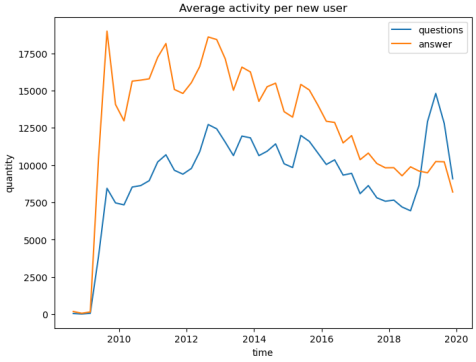
(b) Questions and answers counts over time

4.6 SuperUser.com

SuperUser.com is a rather small community for computer enthusiasts and power users. The community has 861533 registered users of which 7392 were active between December 2019 and February 2020. Members asked 424718 questions in total and gave 587559 answers with an average answer density of 1.383 answers per question. New users asked 161397 questions with an average of 1.085 questions per new user during their first week after first contribution.



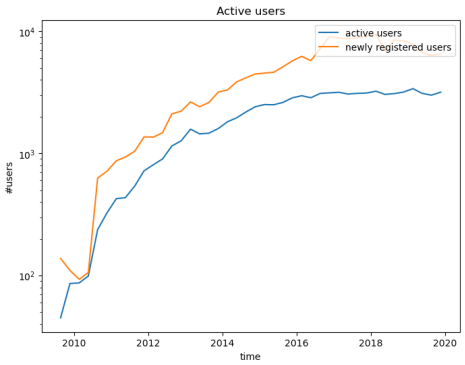
(a) Active users with activity in the last 3 months



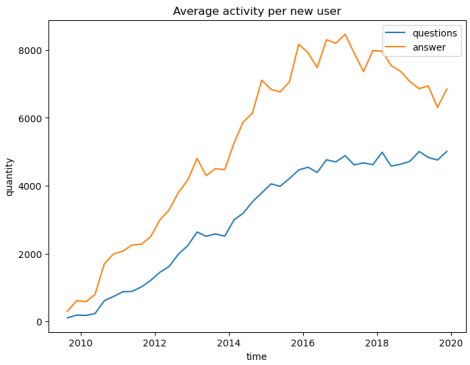
(b) Questions and answers counts over time

4.7 electronics.stackexchange.com

electronics.stackexchange.com is a rather small community for electrical engineering. The community has 184795 registered users of which 3172 were active between December 2019 and February 2020. Members asked 130025 questions in total and gave 221811 answers with an average answer density of 1.705 answers per question. New users asked 47035 questions with an average of 1.126 questions per new user during their first week after first contribution.



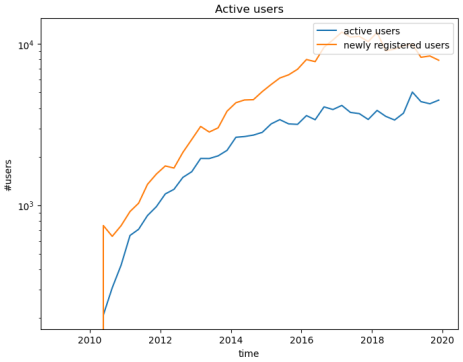
(a) Active users with activity in the last 3 months



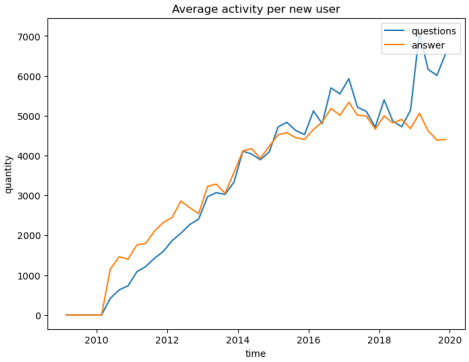
(b) Questions and answers counts over time

4.8 stats.stackexchange.com (Cross Validated)

"Cross Validated is a question and answer site for people interested in statistics, machine learning, data analysis, data mining, and data visualization." [55] The community has 227032 registered users of which 4485 were active between December 2019 and February 2020. Members asked 151777 questions in total and gave 148046 answers with an average answer density of 0.975 answers per question. New users asked 57636 questions with an average of 1.112 questions per new user during their first week after first contribution.



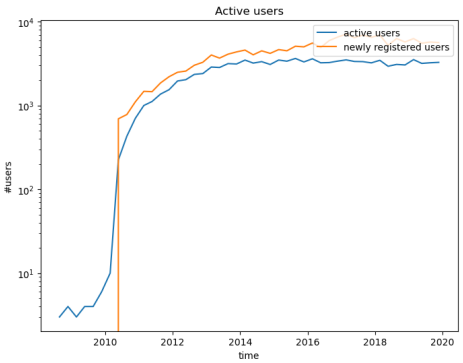
(a) Active users with activity in the last 3 months



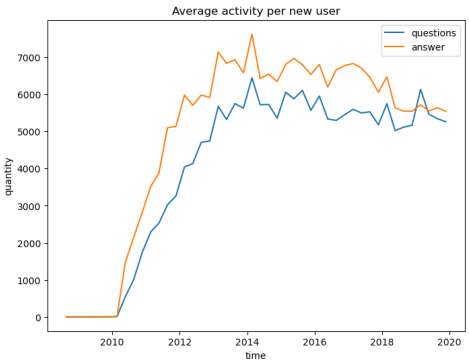
(b) Questions and answers counts over time

4.9 tex.stackexchange.com

tex.stackexchange.com is a rather small community for TEX and related typesetting systems. The community has 171867 registered users of which 3280 were active between December 2019 and February 2020. Members asked 188860 questions in total and gave 227875 answers with an average answer density of 1.206 answers per question. New users asked 59692 questions with an average of 1.191 questions per new user during their first week after first contribution.



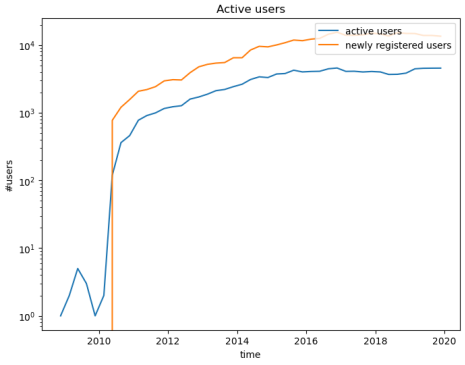
(a) Active users with activity in the last 3 months



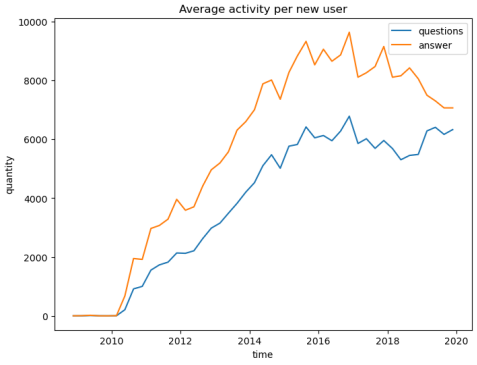
(b) Questions and answers counts over time

4.10 unix.stackexchange.com

unix.stackexchange.com is a rather small community for Linux and Unix-like operating systems. The community has 356498 registered users of which 4565 were active between December 2019 and February 2020. Members asked 174625 questions in total and gave 256007 answers with an average answer density of 1.466 answers per question. New users asked 62437 questions with an average of 1.124 questions per new user during their first week after first contribution.



(a) Active users with activity in the last 3 months



(b) Questions and answers counts over time

5 Results

This section shows the results of the experiments described in section 3 on the data sets described in section 4. In the following diagrams, the blue line states the average sentiment of the answers to questions from new contributors. This line also has numbers attached to it at every datapoint and shows the number of answers that formed the sentiment average. The orange line shows ITS analysis as a 3-segment line.

5.1 StackOverflow.com

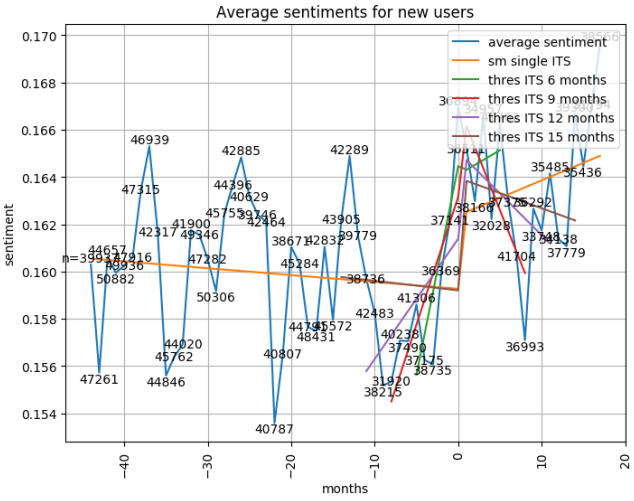


Figure 5.1: An interrupted time series analysis of the sentiments of answer to questions created by new contributors on StackOverflow.com

StackOverflow shows a very slight decrease in average sentiment of time before the change had been introduced. When the change occured the average sentiment jumped up by about 0.003. After the change the sentiments reached higher levels and kept rising.

5.2 math.stackexchange.com

The math.stackexchange.com community shows a decrease in average sentiments prior to the change. The sentiment make a small jump upward when the change is introduced, however, the sentiments decrease faster after the indroduction of the change compared to before the change.

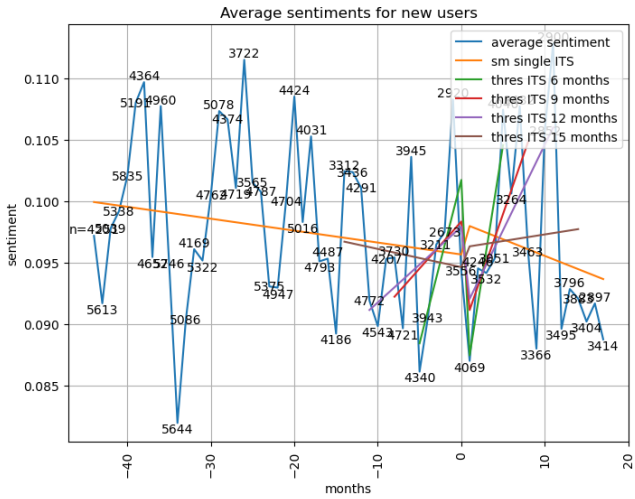


Figure 5.2: An interrupted time series analysis of the sentiments of answer to questions created by new contributors on math.stackexchange.com

5.3 MathOverflow.net

MathOverflow shows a constant regresssion before the change, however, average sentiments are low at about 10 months before the change and spiked high directly

before the change. When the change is introduced regression makes a small jumps up and decreases thereafter. This data set is sparse compared to the other datasets.

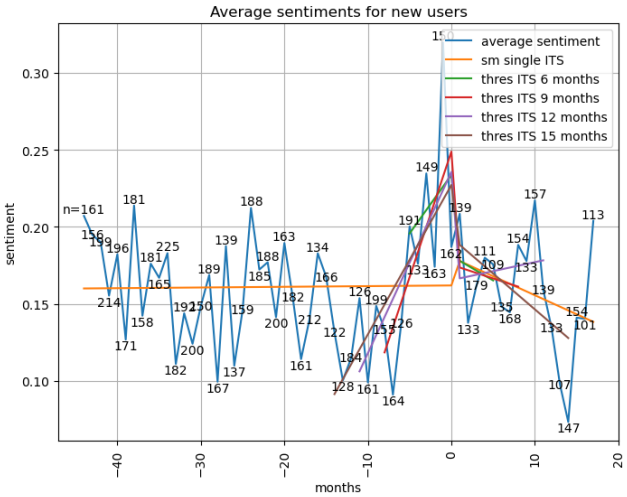


Figure 5.3: An interrupted time series analysis of the sentiments of answer to questions created by new contributors on MathOverflow.com

5.4 AskUbuntu.com

AskUbuntu saw a decrease in average sentiments prior to the change. After the introduction of the change the regression dipped but sentiments keep rising drastically since then.

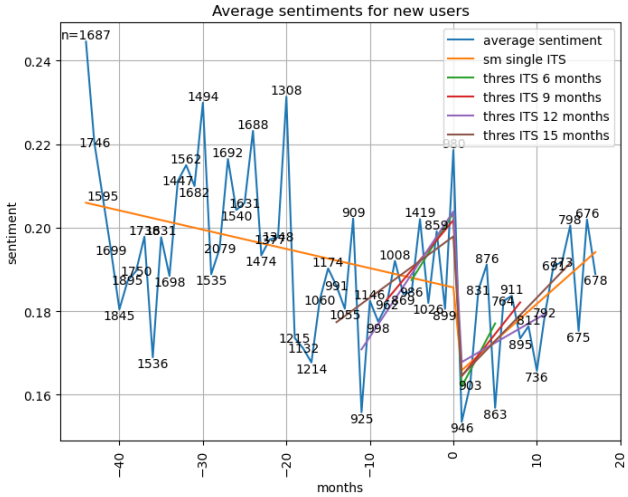


Figure 5.4: An interrupted time series analysis of the sentiments of answer to questions created by new contributors on AskUbuntu.com

5.5 ServerFault.com

ServerFault shows gradually rising average sentiments prior to the change. At the time of the change the regession makes a jump upward and the average sentiment decrease slowly afterward.

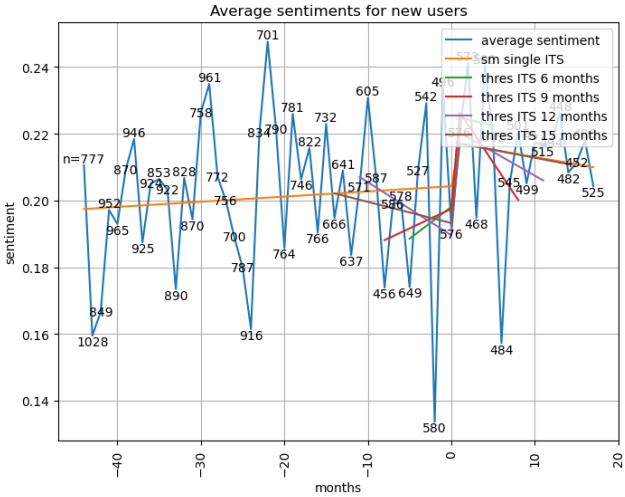


Figure 5.5: An interrupted time series analysis of the sentiments of answer to questions created by new contributors on ServerFault.com

5.6 SuperUser.com

SuperUser shows only sightly decreasing average sentiment up to the change. At the change time the regression takes a dip down and the regression shows a downward trend after the change. Indeed the average sentiments dipped considerably when the change is introducted the average sentiment recovers about 13 months later. Data available in the future will show if the recovery is persistent.

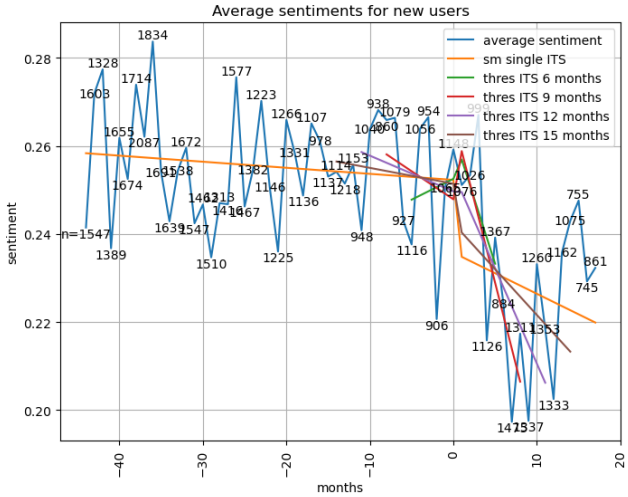


Figure 5.6: An interrupted time series analysis of the sentiments of answer to questions created by new contributors on SuperUser.com

5.7 electronics.stackexchange.com

On electronics.stackexchange.com the average sentiment decreases continuously prior to the change. At the change date the regression makes a little jump upward but the trend from before the change continues afterward. Similarly to SuperUser, the average sentiment recover at about 12 months after the change is introduced and future data will be necessary to determine if the recovery is persistent.s

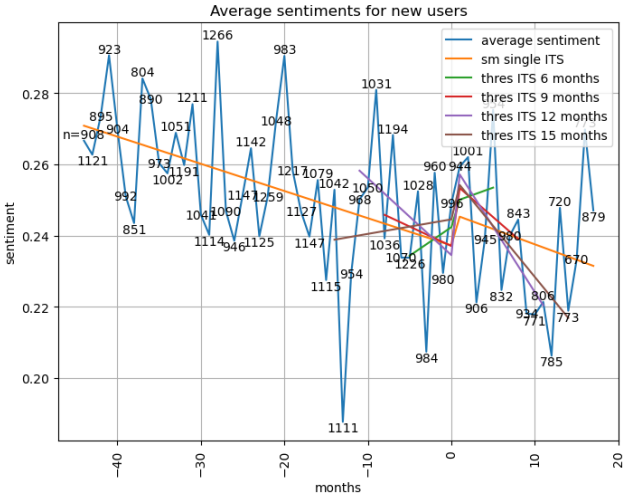


Figure 5.7: An interrupted time series analysis of the sentiments of answer to questions created by new contributors on electronics.stackexchange.com

5.8 stats.stackexchange.com

On stats.stackexchange.com the average sentiment is steadily decreasing prior to the change. The regression dips when the change is introduced. However, the average sentiment after the change indicate a slight upward trend.

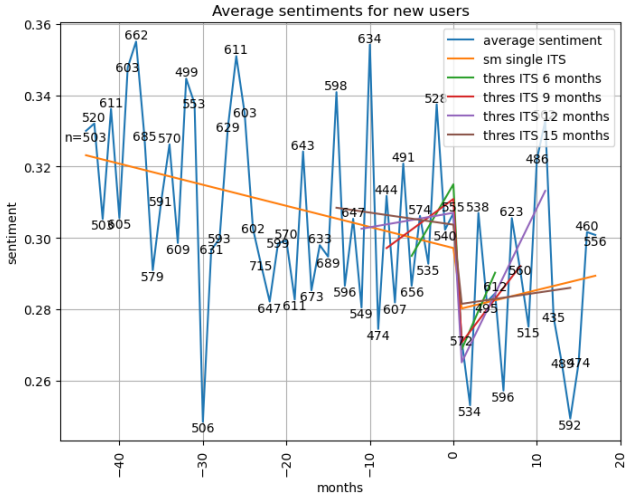


Figure 5.8: An interrupted time series analysis of the sentiments of answer to questions created by new contributors on stats.stackexchange.com

5.9 tex.stackexchange.com

On tex.stackexchange.com the average sentiment is low comapred to the other investigated data sets. Prior to the change the average sentiment only slightly decreases. When the change is introduced the regreesion takes a dip down. After the change the analysis indicates a strong increase in average sentiment. Future data will be required to see if this upward trend continues or evens out.

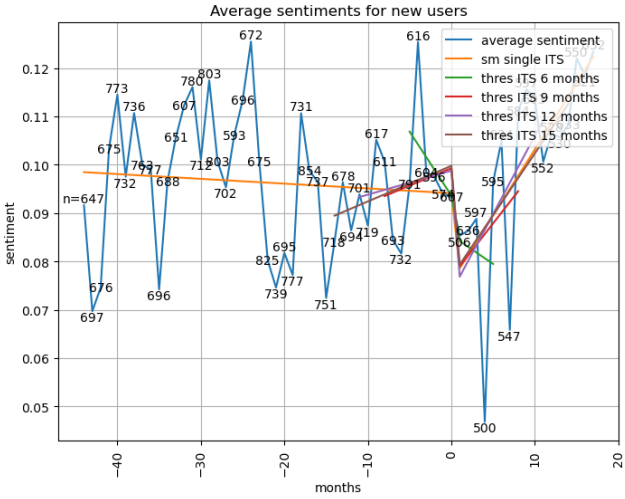


Figure 5.9: An interrupted time series analysis of the sentiments of answer to questions created by new contributors on tex.stackexchange.com

5.10 unix.stackexchange.com

On unix.stackexchange.com the average sentiment is decreasing prior to the change. When the change is introduced the regreesion take a small dip down, however, the average sentiment increases fast after the change.

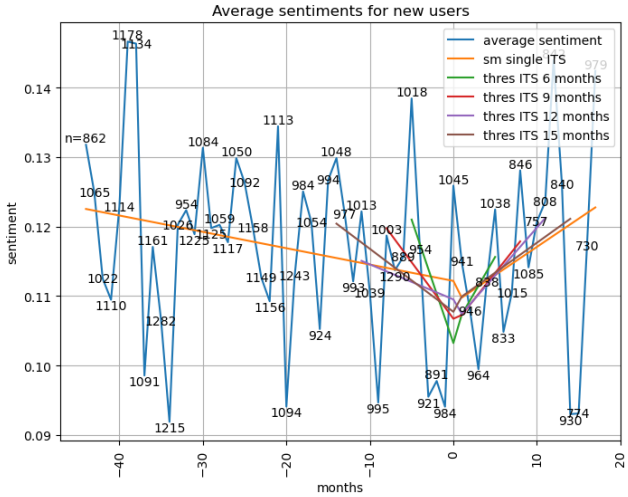


Figure 5.10: An interrupted time series analysis of the sentiments of answer to questions created by new contributors on unix.stackexchange.com

6 Discussion

The ITS analysis of the investigated communities show mixed results. Some communities show an increase in sentiment while others are not affected at all or show a decrease in sentiment. The StackOverflow community has a fairly stable average sentiment prior to the change. The average sentiment jumps into a higher level and keeps rising after the change is introducted. The change has a positive effect on the StackOverflow community. Beside StackOverflow, 4 other communities seem to profit from the change: AskUbuntu, stats.stackexchange.com, tex.stackexchange.com, and unix.stackexchange.com. AskUbuntu shows an interestion zig-zag pattern in the average sentiment graph. Also, the average sentiment is falling prior to the change and raising thereafter, indicating that the change worked for this community. On stats.stackexchange.com the average sentiment is falling prior to the change but since the change the downward trends stopped and the sentiment started to rise slowly, suggesting the change has a positive effect on the community. In the tex.stackexchange.com community sentiments are stable prior to the change and show a stark rising pattern after the change. The change seems to work for this community but future data will be neccessary to see if the rising pattern continues in the shown manner. unix.stackexchange.com also shows a decreasing pattern prior and a rising pattern after the change. So this community also profits from the change.

The other communities do not seem to profit from the change directly. ServerFault is an example where the change does not have a significant impact. The sentiment rises gradually prior to the change, jumps upward by a small value when the change ist introduced and the sentiment is falling slowly thereafter. The data does not inidicate a significant rise or fall in the average sentiment, so this community seem to be largly unaffected by the change. MathOverflow, math.stackexchange.com, and electronics.stackexchange.com show similar results. The average sentiment stay constant on MathOverflow and are falling for math.stackexchange.com and electronics.stackexchange.com. After the change these communities see a decrease

in sentiment. These communities seem to not profit from the change. However math.stackexchange.com has group below average sentiment values at the end which could be a result from another unknown influence. Also the average sentiment on electronics.stackexchange.com seem to recover after about 12 months and future data is required to see if the rise in the end is a long term trend. SuperUser shows a really odd pattern. The average sentiment is stable prior to the change and decreased dramatically shortly afterward. However the sentiment recovers after 12 months. The ITS model chosen in this thesis is not able to capture the apparent pattern. Future data will be necessary to see if the sentiment recovers long term.

By and large, the change introduced by the StackExchange team has a clear positive effect on the average sentiment of half of the investigated communities. Two of the communities have a delayed temporary decrease in sentiment which recovers after about 12 months. The selected ITS model is not designed to capture the sentiment pattern of these communities. For the other three communities the ITS analysis does not show a significant change in the sentiment trend.

Some investigated data sets show intresting patterns. StackOverflow shows the clearest results of all the investigated communities and closely resembles the example ITS shown in section 3. The result matches the expectations and shows that the change introduced by the StackExchange team works well for this community. The AskUbuntu community shows interesting zig-zag pattern where sentiment gradually rises over time and then falls apruptly.

The average sentiment of the StackOverflow community is the most stable in terms of deviation from the regression. This is expected as StackOverflow is the largest community by far and has the most questions created by new comers. On the other hand MathOverflow is the sparsed community and has the least amount questions from new contributors. The level of the average sentiment also varies greatly between communities. stats.stackexchange.com has the highest level of average sentiment compared to the other communities, whereas, tex.stackexchange.com has the lowest level average sentiment. Also, in every community the number of questions from new contributors slowly decreases over time. This may be a result of the filling of gaps in the knowledge repository over time.

7 Conclusion

Bibliography

[1] *Tour - StackExchange*. URL: https://stackexchange.com/tour.

[2] Tim Post. *Come Take a Look at our New Contributor Indicator!* 2018. URL: https://meta.stackexchange.com/questions/314287/come-take-a-look-at-our-new-contributor-indicator.

[3] Clayton J Hutto and Eric Gilbert. 'Vader: A parsimonious rule-based model for sentiment analysis of social media text'. In: *Eighth international AAAI conference on weblogs and social media*. 2014.

[4] Jeff Atwood. *Stack Overflow Private Beta Begins*. 1st Aug. 2008. URL: https://stackoverflow.blog/2008/08/01/stack-overflow-private-beta-begins/.

[5] Dana Movshovitz-Attias et al. 'Analysis of the reputation system and user contributions on a question answering website: Stackoverflow'. In: *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*. IEEE. 2013, pp. 886–893.

[6] *Help Center > Privileges > Vote down*. URL: https://stackoverflow.com/help/privileges/vote-down.

[7] *Help Center > Privileges*. URL: https://stackoverflow.com/help/privileges/.

[8] *Help Center > Privileges*. URL: https://mathoverflow.com/help/privileges/.

[9] *Help Center > Badges*. URL: https://stackoverflow.com/help/badges/.

[10] *What is "meta"? How does it work?* URL: https://stackoverflow.com/help/whats-meta/.

[11] Lena Mamykina et al. 'Design lessons from the fastest q&a site in the west'. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2011, pp. 2857–2866.

[12] Rogier Slag, Mike de Waard and Alberto Bacchelli. 'One-day flies on stackoverflow-why the vast majority of stackoverflow users only posts once'. In: *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*. IEEE. 2015, pp. 458–461.

[13] Denae Ford et al. 'We Don't Do That Here: How Collaborative Editing with Mentors Improves Engagement in Social Q&A Communities'. In: *CHI*. 2018.

[14] Blerina Bazelli, Abram Hindle and Eleni Stroulia. 'On the personality traits of stackoverflow users'. In: *2013 IEEE international conference on software maintenance*. IEEE. 2013, pp. 460–463.

[15] Amiangshu Bosu et al. 'Building reputation in stackoverflow: an empirical investigation'. In: *2013 10th Working Conference on Mining Software Repositories (MSR)*. IEEE. 2013, pp. 89–92.

[16] Stav Yanovsky et al. 'One Size Does Not Fit All: Badge Behavior in Q&A Sites'. In: *UMAP*. 2019.

[17] Tomasz Kusmierczyk and Manuel Gomez-Rodriguez. 'On the Causal Effect of Badges'. In: *WWW*. 2018.

[18] Ashton Anderson et al. 'Steering user behavior with badges'. In: *WWW*. 2013.

[19] Nicole Immorlica, Greg Stoddard and Vasilis Syrgkanis. 'Social status and badge design'. In: *Proceedings of the 24th international conference on World Wide Web*. 2015, pp. 473–483.

[20] Yla R Tausczik and James W Pennebaker. 'Predicting the perceived quality of online mathematics contributions from users' reputations'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2011, pp. 1885–1888.

[21] Gang Wang et al. 'Wisdom in the social crowd: an analysis of quora'. In: *Proceedings of the 22nd international conference on World Wide Web*. 2013, pp. 1341–1352.

[22] Zhiyuan Lin et al. 'Better when it was smaller? Community content and behavior after massive growth'. In: *Eleventh International AAAI Conference on Web and Social Media*. 2017.

[23] Eshwar Chandrasekharan et al. 'You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech'. In: *Proceedings of the ACM on Human-Computer Interaction* 1.CSCW (2017), pp. 1–22.

[24] Jiang Bian et al. 'Finding the right facts in the crowd: factoid question answering over social media'. In: *Proceedings of the 17th international conference on World Wide Web*. 2008, pp. 467–476.

[25] Ramtin Yazdanian et al. 'Eliciting New Wikipedia Users' Interests via Automatically Mined Questionnaires: For a Warm Welcome, Not a Cold Start'. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 13. 01. 2019, pp. 537–547.

[26] Christoph Treude, Ohad Barzilay and Margaret-Anne Storey. 'How do programmers ask and answer questions on the web?(NIER track)'. In: *Proceedings of the 33rd international conference on software engineering*. 2011, pp. 804–807.

[27] Pierre N Robillard. 'The role of knowledge in software development'. In: *Communications of the ACM* 42.1 (1999), pp. 87–92.

[28] Robert E Kraut and Paul Resnick. *Building successful online communities: Evidence-based social design*. Mit Press, 2012.

[29] Jay Hanlon. *Stack Overflow Isn't Very Welcoming: It's Time for That to Change*. 2018. URL: https://stackoverflow.blog/2018/04/26/stack-overflow-isnt-very-welcoming-its-time-for-that-to-change/.

[30] Panagiotis G Ipeirotis and Evgeniy Gabrilovich. 'Quizz: targeted crowdsourcing with a billion (potential) users'. In: *Proceedings of the 23rd international conference on World wide web*. 2014, pp. 143–154.

[31] Julia Silge and Jason Punyon. *Welcome Wagon: Classifying Comments on Stack Overflow*. 2019. URL: https://stackoverflow.blog/2018/07/10/welcome-wagon-classifying-comments-on-stack-overflow/.

[32] Igor Steinmacher et al. 'Social barriers faced by newcomers placing their first contribution in open source software projects'. In: *Proceedings of the 18th ACM conference on Computer supported cooperative work & social computing*. 2015, pp. 1379–1392.

[33] *Wikipedia:Size of Wikipedia*. 2020. URL: https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia.

[34] David Allen. 'Do organizational socialization tactics influence newcomer embeddedness and turnover?' In: *Journal of Management* (2006).

[35] Blair Nonnecke, Dorine Andrews and Jenny Preece. 'Non-public and public online community participation: Needs, attitudes and behavior'. In: *Electronic Commerce Research* 6.1 (2006), pp. 7–20.

[36] Joe Friend. *Rolling out the Welcome Wagon: June Update*. 2018. URL: https://stackoverflow.blog/2018/06/21/rolling-out-the-welcome-wagon-june-update/.

[37] Jay Hanlon. *The War of the Closes*. 2013. URL: https://stackoverflow.blog/2013/06/25/the-war-of-the-closes/.

[38] *Developer Survey Results 2019*. 2019. URL: https://insights.stackoverflow.com/survey/2019.

[39] Denae Ford et al. 'Paradise unplugged: Identifying barriers for female participation on stack overflow'. In: *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering*. 2016, pp. 846–857.

[40] Joel Spolsky. *Kicking off the Summer of Love*. 2012. URL: https://stackoverflow.blog/2012/07/20/kicking-off-the-summer-of-love/.

[41] Jaydles. *The NEW new "Be Nice" Policy ("Code of Conduct") — Updated with your feedback [duplicate]*. 2014. URL: https://meta.stackexchange.com/questions/240839/the-new-new-be-nice-policy-code-of-conduct-updated-with-your-feedback.

[42] Kristina Lustig. *Mentorship Research Project - Results + Wrap-Up*. 2017. URL: https://meta.stackoverflow.com/questions/357198/mentorship-research-project-results-wrap-up.

[43] Bogdan Vasilescu, Andrea Capiluppi and Alexander Serebrenik. 'Gender, representation and online participation: A quantitative study'. In: *Interacting with Computers* 26.5 (2014), pp. 488–511.

[44] *Developer Survey Results 2015*. 2015. URL: https://insights.stackoverflow.com/survey/2015.

[45] Paul A David and Joseph S Shapiro. 'Community-based production of open-source software: What do we know about the developers who participate?' In: *Information Economics and Policy* 20.4 (2008), pp. 364–398.

[46] Jacob Clark Blickenstaff*. 'Women and science careers: leaky pipeline or gender filter?' In: *Gender and education* 17.4 (2005), pp. 369–386.

[47] Catherine Hill, Christianne Corbett and Andresse St Rose. *Why so few? Women in science, technology, engineering, and mathematics*. ERIC, 2010.

[48] Ashton Anderson et al. 'Discovering value from community activity on focused question answering sites: a case study of stack overflow'. In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2012, pp. 850–858.

[49] Huseyin Cavusoglu, Zhuolun Li and Ke-Wei Huang. 'Can gamification motivate voluntary contributions? The case of StackOverflow Q&A community'. In: *Proceedings of the 18th ACM conference companion on computer supported cooperative work & social computing*. 2015, pp. 171–174.

[50] Zhuolun Li, Ke-Wei Huang and Huseyin Cavusoglu. 'Quantifying the impact of badges on user engagement in online Q&A communities'. In: (2012).

[51] Sonic the Anonymous Hedgehog and Brian Nickel. *What are the exact criteria for the "new contributor" indicator to be shown?* 2018. URL: https://meta.stackexchange.com/questions/314472/what-are-the-exact-criteria-for-the-new-contributor-indicator-to-be-shown.

[52] *Files for stackexchange*. URL: https://archive.org/download/stackexchange.

[53] James Lopez Bernal, Steven Cummins and Antonio Gasparrini. 'Interrupted time series regression for the evaluation of public health interventions: a tutorial'. In: *International journal of epidemiology* 46.1 (2017), pp. 348–355.

[54] *Mathematics Stack Exchange Main page*. URL: https://math.stackexchange.com/.

[55] *Cross Validated Main page*. URL: https://stats.stackexchange.com/.

Appendix