Investigating Student Experiences with GitHub and Stack Overflow: An Exploratory Study

by

Trishala Bhasin

B.Tech., Guru Gobind Singh Indraprastha University, New Delhi, India, 2018

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

Investigating Student Experiences with GitHub and Stack Overflow:
An Exploratory Study

by

Trishala Bhasin

B.Tech., Guru Gobind Singh Indraprastha University, New Delhi, India, 2018

Supervisory Committee

Dr. Adam Murray, Co-supervisor
(Computer Science Department)

Dr. Margaret-Anne Storey, Co-supervisor
(Computer Science Department)

ABSTRACT

Programmers who want to improve their skills and background in software development rely heavily on developer social platforms such as GitHub and Stack Overflow to enhance their learning. Stack Overflow provides answers to questions they have about languages or library skills they wish to acquire, while contributing to open-source projects hosted on sites like GitHub gives them valuable experience. Students also use these platforms during their education: most will rely heavily on Stack Overflow at some point in their schooling, while many can benefit from contributing to GitHub projects to build their expertise and professional portfolios. We already know from previous research that developers face barriers participating on these platforms, and therefore we may expect that at least some students will experience similar or possibly even bigger barriers. This research describes a semi-structured interview study followed by a survey with university students to explore how they use the GitHub and Stack Overflow platforms. I identified the benefits the students report from using these tools and the barriers they face. I have concluded with some preliminary recommendations on how to reduce the hurdles students may face with these and other developer social platforms, and I have also suggested future work to mitigate these roadblocks.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

Moving to a new country half a world away to pursue a master's degree has been a life changing experience. The three years I spent as an MSc Computer Science student have helped me grow a lot professionally and personally. From deciding to switch to a thesis based degree, teaching seven courses as a TA, to working in a software development company, I have learnt a lot.

I would like to begin with thanking **Dr. Adam Murray** for being co-supervisor and mentor through my research journey. Your detailed feedback at every step of my research, and your giving of so much of your time to this thesis has brought the quality of my work to levels I had not thought possible.

Thank you **Dr. Margaret-Anne Storey** for your mentoring, support, encouragement, and patience. You are my inspiration and I will always be grateful to you for giving me the opportunity to begin my research journey with you. I would also like to extend a big thanks to **Cassie** and all the current and former members of the **CHISEL research group: Leon, Soroush, Jorin, Omar, Andreas, Arman** for always being there to help. It would have been impossible without all of the encouragement and support from you all.

A big thank you to my mom **Mrs. Puja Bhasin**, my dad **Mr. Raman Bhasin**, my uncle **Mr. Nitin Anand** and my grand-parents for always believing in me even in the days when I didn't believe in myself. Special thanks to my world, my sister **Sia Bhasin**. Thank you for always being there for me.

I would like to thank my friends **Deeksha**, **Jaya**, **Ira**, **Ishika**, and **Abhinav** for listening to my rants and busting my stress whenever I got overwhelmed. Lastly, I would like to thank my special friend **Clint Ledding**, this thesis is as much yours as it is mine. My thesis would have been impossibly difficult without your help with edits and your tremendous support.

Trishala Bhasin

DEDICATION

To all the current and future software developers

Chapter 1

Introduction

Open communities, social coding, and collaborative development form the backbone of the modern software development process. The products developed using these processes come under *free license*, wherein anyone can join and work together in order to develop a product or service. The expression **open** in an open source community alludes to the open door for anybody to join and contribute to the collaborative project. The course and objectives of the project being developed are resolved collaboratively by all individuals from the community. The resulting work is made accessible under a free permit, which makes it available to the public to be used and modified freely. As such different communities around projects can get formed and can expand on these projects developed collaboratively[1].

Such communities that are formed around social coding platforms [12] exist because of numerous participants who, after having invested resources in developing new software, freely reveal their proprietary innovations as a public good [47]. In general, the platforms that help in the formation of open communities around them are comprised of open source projects as well as the open forums where people indulge in discussions, problem solving, and code verification.

In this thesis, I use the term *developer social platforms*, a concept that extends the idea of open communities and collaborative development to platforms such as GitHub, GitLab, Stack Overflow, and Stack Exchange. These are often seen as a valuable resource for today's developers [11]. For example, developers use GitHub

[1]Parts of this thesis that include chapters 1,2, and 4 are translated from a research paper that I co-authored with my supervisors Dr. Adam Murray and Dr. Margaret-Anne Storey. I sincerely thank them for helping me lay the foundation chapters of this research.

and Bitbucket to host software projects and to make meaningful contributions to other projects. They use Stack Overflow and Reddit to enhance their own knowledge and skills, and to solve problems they may face while developing code. Developers benefit in other ways when they explicitly contribute to resources hosted on these open platforms (e.g., contributing by authoring or voting up/down a resource) as they gain visibility in the community due to the social and public nature of the platforms [13][8]. Furthermore, the public contributions that developers make often comprise portfolios they can use when applying for or switching jobs [44].

Due to the importance of developer social platforms in the lives of software developers, this has been an area of interest for researchers: a majority of most recent papers I found involve how developers both contribute to and benefit from these platforms. GitHub and Stack Overflow are the most widely used platforms, and they have been studied the most frequently. A lot of studies have also revealed many barriers (such as lack of confidence or time) that certain developer demographics experience when contributing to resources hosted on GitHub and Stack Overflow [31].

However, most of the studies I came across were focused on developers who have gained software development experience at the industrial level. Relatively little has been done to investigate how **students** use GitHub, Stack Overflow or other popular developer social platforms during their school years. For the students enrolled in degrees related to computer science and engineering, a lot of their development happens while they attend school. Hence, it would be interesting to *investigate if students experience similar or even different benefits of participation than those of senior developers, and do they encounter similar or novel barriers that may interfere with the benefits they could receive?*

We know that students are an important part of these developer social platforms and that they participate in a variety of ways, including hosting their code online, contributing to other people's code, and learning more about programming by finding answers to their questions and by helping others [21]. This was also observed in a study by McLoughlin et al. that focused on the interactive aspects of today's internet such as Reddit and Facebook groups, called the modern Web 2.0 [30]. Web technologies offer a medium for students to learn and grow through personalization, customization, and opportunities for networking and collaboration. Additionally, many software development companies request a student's GitHub profile as a part

of their hiring process[2]. Hence, many students strive to maintain a positive open community presence as it increases their chances of getting hired in the software development industry.

What has been less studied are the barriers students may face while using developer social platforms such as GitHub and Stack Overflow. I believe that this is important to investigate because my investigation may enable educators to be in a position to act on some of these barriers at an early stage of student development. I also want to be able to share insights with the designers of these and other developer social platforms so that they can make accommodations to create a more welcoming environment for students.

In this thesis, I describe the exploratory study I conducted using a mixed-methods approach that included semi-structured interviews and surveys with students at different stages of their studies. The main goal was to investigate how the students use and contribute to content posted on developer social platforms such as GitHub and Stack Overflow. I report in depth on the barriers they encounter and the benefits they perceive from using or contributing to these platforms as these have not been studied much by others.

Many of the barriers I identified in my study that students encounter overlap the barriers reported in earlier research for developers. The existing literature in the field of social interactions on developer social platforms enlist a lack of self-confidence and a disinterest in hosting the content on developer social platforms as barriers to participation experienced by different people on the developer social platforms. However, this barrier was not as strongly reported in other studies as it was in my research. I find that students as a group of users face even more doubt as compared to other users in their abilities to contribute and lack the motivation to help others even though doing so could help them. My thesis also points to recommendations that educators and students may follow to increase participation and meaningful contributions, and my findings may also be relevant to designers of developer social platforms.

[2]https://techbeacon.com/app-dev-testing/what-do-job-seeking-developers-need-their-github

1.1 Motivation

The driving motivation for my work on this research is to understand the dynamics and incentives underlying student participation on social coding platforms such as GitHub, GitLab, and online question/answer forums such as Stack Overflow. The communities that are formed around developer social platforms thrive only when people are willing to participate. There are people who use a lot of content that is available on these platforms such as downloading projects that are created by other people on social coding platforms, or by using the content available on forums such as Stack Overflow to get answers to their questions. Some people also contribute to the content available on these developer social platforms by sharing public-facing projects or by asking or answering questions on platforms like Stack Overflow. These are the kinds of people who become part of the communities that are formed around the platforms in question. One of the main way of keeping these developer social platforms running is for user participation to continue such that more content is created and more content is available for people to use. A lack of active participation threatens the sustainability of online communities [3].

According to online surveys such as the annual Stack Overflow report of 2019, the participation rates of students is low [46], hence I decided to focus my study on student participation on social coding platforms such as GitHub and Stack Overflow, a popular question/answer forum. I chose to investigate students because I was intrigued to understand their interaction with these platforms, understand the barriers they face when they wish to make contribute, and understand the benefits that the students who participate on the developer social platforms experience. By identifying the barriers, I want to make recommendations to different actors involved directly or indirectly in the development and the usage of developer social platforms to take steps to mitigate the barriers in order to increase student participation. I want to explore the benefits that some individuals experience when they participate on the developer social platforms. I believe that by highlighting the benefits to participation, I can motivate the students who do not participate actively to overcome some of the barriers they experience and participate more.

Also, I have used those developer social platforms extensively since my Bachelor degree - including during my time working in software development companies. Also, during my master's degree, I got an opportunity to observe the way students use

developer social platforms closely while working as a Teaching Assistant. I could not find a lot of literature that specifically focused on student interactions, hence I decided to pursue this topic. I followed an exploratory approach where I first conducted semi-structured interviews with different groups of students in order to identify their participation patterns and observe the barriers and benefits they experience while participating in the developer social platforms under investigation. Then I designed a survey which was based on the findings of the semi-structured interviews and distributed it across a larger pool of students to understand how findings from the interviews applied to a larger group of students.

I anticipated that there would be some students who would participate more consistently on these developer social platforms, and that they would help me map practices that could help encourage students, educators, and platform designers to take steps in order to increase participation in these communities.

1.2 Research Questions

Since the overarching goal of this study is to explore the participation trends exhibited by students, I began my study with three research questions.

RQ1: How do students participate and contribute to open content hosted on developer social platforms?

I conducted semi-structured interviews followed by a survey with students at different levels of education in order to understand their interactions with the developer social platforms under investigation, that is, GitHub and Stack Overflow. While numerous participants revealed that they actively use these platforms for their assignments or during their internships, most of their participation is motivated by a **compulsion** to complete school work or job-related work and, as such, they may lack intrinsic motivation to participate in other projects featured in other open communities. There were some participants who did, however, actively participate on these platforms outside of their school or work.

RQ2: What are the barriers that students face in their participation on GitHub and Stack Overflow?

Upon investigating the interactions of students with the developer social platforms they use, I identified some barriers that these students face in using both GitHub and Stack Overflow. These barriers originate from a place of insecurity with regards to

their skill-sets as developers. I discuss the source of these barriers and present specific examples of student experiences in which they face these barriers.

RQ3: For the students who participate, what benefits do they experience from their participation on developer social platforms?

While I identified many barriers that students face, it is notable that the students who actively participate with others on developer social platforms experience certain benefits that help them with their programming skills. By presenting these benefits, I attempt to motivate participation from the students who do not actively contribute on developer social platforms so that they can experience the benefits that their peers enjoy.

1.3 Contributions

This study makes five major contributions that can be used by **students** to understand not just how they themselves interact with developer social platforms, but also how their peers use these platforms. Additionally, this study can help **educators** make efficient study materials which are able to facilitate students' understanding of the culture around project development within the developer social platforms. It can also help the **project owners** who host their work on various developer social platforms understand how students feel when they interact with these platforms. The main contributions of this thesis towards understanding the role that developer social platforms play in the life of students are as follows:

- Identifying which is the preferred social coding platform used by the students to host their programming projects by observing the platform usage trends.
- Understanding how students interact with different developer social platforms when they are creating code related to their school or job versus when they are creating code while developing their personal projects.
- Understanding the role that Stack Overflow plays when students need help with their code, and how they view content on the platform in terms of it being a place to make contributions.
- Exploring the barriers and benefits that students experience when participating on these platforms, and how these barriers differ for students who are enrolled

in an undergraduate degree and the students who are enrolled in a graduate degree.

- Providing recommendations to different stakeholders who could benefit from understanding the findings of this study (as outlined above).

1.4 Thesis Outline

This thesis is structured as follows:

Chapter 2: In this chapter, I discuss previous work that formed the background for my research. it contains the literature around the formation of open communities around different developer social platforms, how different members who are part of these communities use the platforms, and what existing barriers and benefits are experienced by different demographics of participants of various developer social platforms. I also discuss various users of these platforms such as active users and the users who only consume the content without making active contributions (lurkers [51]), and the role the developer social platforms play in the lives of students more generally.

Chapter 3: In this chapter, I describe the methodology I followed to conduct this exploratory study and how I used the semi-structured interview process to derive preliminary results regarding student participation. I also discuss how I incorporated my preliminary results into a survey that helped me strengthen the findings from my interviews and provide more insights.

Chapter 4: In this chapter, I discuss the semi-structured interview process, providing details about how I collected the data, the steps I followed to analyze the interview data, and the results obtained.

Chapter 5: In this chapter, I discuss the pilot survey I conducted as the next phase of my research, outlining the question formulation process, survey dissemination, data collection, data analysis, and the results obtained.

Chapter 6: In this chapter, I discuss the research findings by combining the

results obtained from the semi-structured interviews and the pilot survey. I also discuss how the barriers I identified regarding student participation resonate with the barriers that are already discussed in related work on the participation on various developer social platforms.

Chapter 7: In this chapter, I discuss threats to the validity of my study, such as the choice of data collection tools, the participant selection process, and the data analysis process.

Chapter 8: In this chapter, I conclude my study and discuss future work.

Chapter 2

Background and Related Work

Developer social platforms such as GitHub, GitLab, and Stack Overflow attract a lot of software developers. These developers perform various activities on the platform such as search projects on the platforms that are hosted by other developers, contributing to the projects, host their own projects, participate in discussions on different coding technologies, or follow repositories of their interest. These people together represent a community that is formed around their content of common interest which is hosted on developer social platforms [57].

Many studies have looked at open source platforms, the structure of the communities formed around these platforms, and how participation occurs among the individuals who are a part of these communities on various developer social platforms. Some of the studies focus on different demographic sectors of the communities, for example, some emphasize understanding the participation of women on different developer social platforms [21] [20], while others work towards understanding the factors that draw participation in the creation and the usage of content hosted on the developer social platforms [11].

The best way to carry out these so far exploratory studies is to identify small demographic groups of participants that exist in the communities formed around various projects hosted on the platforms under investigation, analyze their participation, and provide details on mitigating barriers around their participation on the developer social platforms [15]. Communities on the platforms, such as GitHub and Stack Overflow, are heterogeneous: each group is perceived differently in the communities formed around the content hosted on the platforms, each group uses the platforms differently, and the problems faced as well as the benefits experienced by each of the groups are

different [54]. In this chapter, I discuss the existing research on the formation of the communities around the developer social platforms, user participation and the barriers and the benefits to participation that are reported by the previous studies. I also discuss the existing work on the role of developer social platforms in the life of students, which is the area of interest for my research work.

2.1 Open Platforms, Open Communities and the Emergence of Developer Social Platforms

Open source platforms such as GitHub, GitLab, and Bitbucket; as well as question-answer forums such as Stack Overflow are an important part of collaborative software development [11]. These platforms provide a space where people can actively showcase their programming projects to other people, seek collaboration on their work, ask questions, and resolve issues. Formation of open communities around the content hosted on GitHub, GitLab or similar platforms begins when a developer decides to share the source code publicly for their project and gets attention from people in the form of active collaboration.

One such example of the formation of a community around the development of an open public project is that of the Linux operating system - Linux being one of the most successful software systems available. The entire development process of the Linux kernel happened through the contribution of open source developers who built an ecosystem around it [23]. The sharing of source code on these platforms is associated with the idea that there should be content available online so that the community can use the content in any way they wish.

In the same way, on forums such as Stack Overflow (which are extensively used by software developers for knowledge sharing and acquisition), a community is formed when developers share their knowledge about a specific topic such as R language or JavaScript [56]. These people represent a group of individuals who have common interests and communicate with each other through asking questions, answering other people's questions, commenting on each other's posts or up-voting/down-voting answers.

A big characteristic of the content available on the platforms I discussed above

is *openness*, which means that the content hosted on these public facing platforms is free to view, use, and contribute to [51]. Most of the content that is available on these platforms is added by individuals who decide to invest time in creating content to share publicly, much of the time without any monetary compensation [26]. Such individuals become part of open communities that are formed around different topics of interest on the platforms. As members of the communities they collaborate with each other to create code, resolve each other's issues, and thus develop collaboratively [37].

Finally, all these platforms are collectively called **developer social platforms**, which is a term I use throughout this thesis to describe platforms such as GitHub, GitLab, Stack Overflow, Stack Exchange. These developer social platforms offer a public space for different communities to access content or carry out development activities.

In this study, I further divide developer social platforms into two parts, namely social coding platforms and question-answer forums. Though each of these types of developer social platforms host open content created by users, ultimately the nature of the content is different. Further in this section, I discuss the related work around social coding platforms and the question answer platforms and provide the background decision making process for what specific developer social platforms I chose to investigate as a part of my research.

2.1.1 Social Coding Platforms

Platforms such as GitHub, GitLab, and Google code offer a lot of facilities to the developers such as: providing developers a space to store their projects online, offering version control capabilities to help the developers keep track of the code they modify, and offering the developers an ability to facilitate collaborative development [52].

The users of these platforms share their projects either publicly: which makes their code available for anyone who visits the platform, or privately: where users can share their code with a limited audience [37].

These platforms are popularly known as social coding platforms as they offer a space for developers to share their code and seek collaboration from people that they may not otherwise come across.

The social aspect of these platforms open doors for developers to expand their pro-

gramming circle beyond known individuals and allows collaboration with new people who share common programming interest. Open communities are usually formed on these platforms when a group of programmers contributes to common projects and develops software under a free licence. The members of such communities offer support to each other, share their knowledge with each other, review each other's code and resolve each other's issues. For example, R community is a group of programmers who contribute to different modules of R project hosted on GitHub.

Out of all the social coding platforms that offer facilities hosting public facing projects, the existing studies report GitHub as the most popular social coding platform [7]. Due to its popularity, many developers try to maintain a presence on this platform in order to gain recognition in the community and to increase the reach of the projects they develop [57]. Also, due to the popularity of GitHub as the most widely used platform with a lot of users, it is a popular choice for researchers studying different aspects of these kinds of platforms - their research ranging from analyzing the content hosted on the platform to understanding the dynamics of different members of the open communities formed around different projects hosted on the platform.

Among all social coding platforms, GitHub therefore has the greatest popularity and among all social coding platforms, I chose to study student participation trends on GitHub for this reason.

2.1.2 Question-Answer Forums

Writing a piece of code requires multiple steps such as identifying the code logic, using the syntax of the language correctly, and interfacing various code modules such that different code blocks run together smoothly. With all of these steps involved in programming, it is inevitable that the individuals who are writing the code will run into programming problems. In such cases, the most natural instinct of the programmer is to type their problems into general purpose search engines and then canvass the results available [55]. There are two ways in which the participants seek help:

1. Through blog posts that are shared by an individual. While it is a good source of acquiring knowledge, the information that the developers access there is coming from a single person's perspective.

2. Another way is to access the content on platforms such as Stack Overflow, where users host their programming problems and the discussions around the solutions provided by the users who know the solution. This can help the developers gain the perspective of multiple developers who encountered similar problems and seek the solutions faster.

Question answering websites are large repositories of valuable knowledge. In addition to the question-answer forums being aimed at providing useful answers to the question asked, there has also been a marked shift towards question answering as a community-driven knowledge creation process whose end product can be of enduring value to a broad audience [5]. Stack Overflow is one of the most popular question answering forums that hosts questions about several programming languages [52]. Due to its popularity among software developers, Stack Overflow is a platform studied by a plethora of researchers. Studies have been done to understand: the community formation and value [5], the participation of different members of Stack Overflow [21], and how knowledge is shared over this platform [55].

I speculate that the students who are enrolled in degrees that involved the creation of programs extensively use the content on Stack Overflow. I did not come across many studies that focus on student participation on Stack Overflow, hence I decided to investigate student participation on Stack Overflow as a part of my thesis.

2.2 Students and Their Interactions with Developer Social Platforms

Students can participate in communities that are formed around a given developer social platform [19]. In their study on the impact of Web 2.0 pedagogy in participation on online communities, McLoughlin [30] ostensibly believes that participatory learning lends itself well to education as students benefit from learning opportunities where they can connect with, and learn from each other. It has been observed that students interact with developer social platforms during their degrees to seek help with their assignments or to collaboratively develop projects and learn new languages[1]. Sometimes students get involved in internships where they have the opportunity to work

[1]https://stackoverflow.blog/2017/02/15/how-do-students-use-stack-overflow/

on open source projects[2]. Also, educators are discovering the potential for GitHub as an emerging collaborative platform with pedagogical utility as a tool for managing, improving, and transforming the learning experience [58].

By being active on developer social platforms, students get an opportunity not just to create, but also to showcase their work to other people and receive collaborative contributions to that work. The projects that students are able to collaborate on are beyond those created by the students they take courses with or the developers in the companies they work with. However, according to online surveys such as the annual Stack Overflow report of 2019, the participation rates of students is low [46]. Not much is known about how students participate, or if they experience any barriers to participation and whether the barriers that the students face are any different from the barriers that are experienced by developers not enrolled in school.

In the upcoming sections I discuss the types of participation seen on various developer social platforms as well as the the barriers and the benefits to participation on developer social platforms experienced by users of said platforms.

2.3 Types of Participation

The existence of developer social platforms is based on the idea of participation and collaborative development. The term **participation** is defined as the act of taking part in something with or without leaving a trace behind. In terms of open communities formed around developer social platforms, participation can be of two types [37]: Active Participation (which means using content posted on developer social platforms and adding content back for other users to see) and Passive Observation (Which means to use the content being hosted on the developer social platforms without leaving any trace behind for other users to see)[3]. **Contribution** is a subset of participation, which means that everyone who actively contributes definitely participates, but not everyone who does participate in online communities actively contributes. For instance, using an answer from Stack Overflow is participation but

[2]https://thenextweb.com/podium/2019/10/02/why-github-is-the-gold-standard-for-developer-focused-companies/

[3]https://blog.studentlifenetwork.com/2019/02/28/type-of-student-active-participant-passive-observer/

using an answer and up voting it upon finding it useful is contribution.

In this section, I discuss the different forms of participation so that I can classify the participants of my study. Defining the different types of participation helps in mitigating the potential confusion for the readers as many people believe that participation also refers to just accessing the content and not adding anything back to the content from their side.

2.3.1 Active Participants (Contributors) on Developer Social Platforms

Active participation in open communities refers to performing actions on the developer social platforms that leave a trace of the presence of a user. For instance, on platforms such as GitHub, active participation can refer to making pull requests on other people's repositories, setting their repositories to public in order to seek collaboration from other users, marking the repositories of different users, raising issues by commenting in forums or on projects, contributing to documentation of projects, to name a few [38]. In the same way, on question-answer forums like Stack Overflow, active participation corresponds to activities such as asking questions, answering questions being asked by others, up-voting, down-voting the responses, or moderating the website [1].

2.3.2 Passive Observers on Developer Social Platforms

Passive observation is a term used to describe the behaviour of individuals who consume the content on a given platform without making contributions. These are also commonly known as **lurkers** [17]. It is noteworthy that lurkers are also a part of the developer social platforms as they act as an *audience*[4] for the content being generated by active participants. Their participation is not limited to viewing the content, often they use the content as well. Lurking is seen as a derogatory term in the literature, as the lurkers are seen as users of content who do not give back to the community [51]. Passive observers or lurkers are not the same as non-users, as non-users are individuals who do not even access the content for a given developer social platforms [17].

[4]https://blogs.ch.cam.ac.uk/pmr/2006/10/01/open-source-and-the-tragedy-of-the-lurkers/

2.4 Participation on Developer Social Platforms

In this section, I discuss the literature around the barriers and benefits that people who participate in the developer social platforms experience.

2.4.1 Barriers

There have been studies to identify barriers to participation experienced by people who are using developer social platforms [48]. 90 percent of people who use content posted on developer social platforms don't effectively partake in online conversations, while 9 percent of people add somewhat, and just 1 percent of people represent practically all of the activities on developer social platforms [3].

The barriers that people in open communities face can be social as well as technical in nature. Some of the social barriers include the inability to participate in discussions as a result of encountering an unwelcoming environment, the existence of stereotypes around participation, and discrimination in the participation process due to a lack of experience [3]. Some of the common technical barriers most prevalent on code sharing platforms like GitHub are a lack of proper guidelines for making contributions, and a lack of proper documentation available to explain the code of the project being hosted on the platform [31]. For those who wish to participate, the barriers can be intrinsic as well as extrinsic [23].

In Figure 2.1, I synthesize the barriers that are discussed in the existing literature and that I anticipated I might find in my study with students.

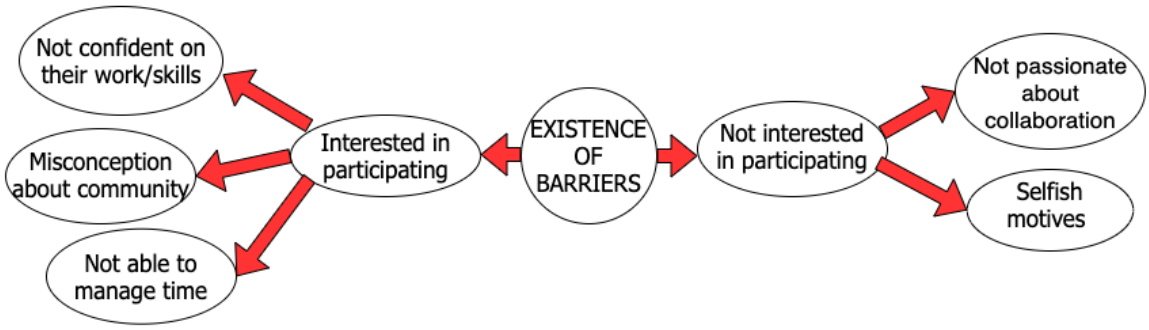


Figure 2.1: Barriers seen in literature around general participation in open communities

In particular, this figure divides the barriers that are experienced by the people who interact with developer social platforms into two sets. Firstly, barriers that relate to a lack of interest [31] that include: a lack of passion and lack of altruism [23]. Secondly, barriers that are faced when developers are interested in participating [31] including: a lack of confidence [48], misconceptions about the community culture and support [3] along with poor management of time.

2.4.2 Benefits

Developer social platforms are used often by of people working in software development. With so many people accessing these platforms, large communities tend to form around these platforms. Much of the existing literature discusses the benefits of large open communities associated with developer social platforms which are enjoyed not only by developers but also the businesses involved in software development[34]. Some of the technical benefits include:

- Being able to *rely* on a community with different levels of expertise.
- *Security* in terms of good integration with code testing and access authentication tools that provides a safe space for developers to develop and share their code.
- The opportunity to collaborate with different people and enhance code *quality* and *performance*.
- Strong *developer and tester base* for the code being generated.

These benefits also extend to businesses:

- *Low cost* associated with development due to the content being open source and the community members that often contribute without charging salaries.
- *Freedom from licensing* due to the open source nature of the code.
- *Increase in collaboration* that encourages innovation.
- *Extra business functionality* due to greater code accessibility by groups of people.

In addition to the benefits listed above, many existing studies in the field of participation on the developer social platforms discuss **future rewards** that are gained by the individuals who participate actively [23]. These future rewards come into play in the form of increasing one's *human capital* [6] by means of education, training, learning, and practicing that leads to better job opportunities, higher salaries, and more fulfilling jobs.

In 2007, Sauer et.al interviewed 2718 developers asking them for the reasons why they joined open source communities hosted on developer social platforms[42]. Their responses included to: learn and develop new skills, share knowledge and skills, participate in new form of cooperation, improve OS products for others, participate in OS scene, development and use of software should be free, solve problems that proprietary software could not solve, get help in understanding the idea of software products, limit the power of large software companies, and build reputation in the community.

As discussed earlier, developer social platforms can also help students build a portfolio that they can use to show their skills to prospective employers in the software development field.

2.5 Chapter Summary

In this chapter, I explained the emergence of the concept of developer social platforms, followed by explaining the role these platforms play in the life of students, and the existing literature around understanding different cohorts of participants in the developer social platforms. I also explored the work that has been done towards understanding barriers and benefits of participating on the developer social platforms.

In the next chapter, I explain the methodology I followed to conduct my study which include the design of the data gathering tool, role of the research team in different phases of the study and the research ethics I followed.

Chapter 3

Methodology

Empirical studies involve gaining knowledge on a particular topic by means of direct and indirect observation. The evidence and data gained from them can be analyzed both qualitatively and quantitatively. The nature of my exploratory study is empirical, and there are many ways to conduct empirical research especially in the field of software engineering [16]. The technique that is to be used to conduct a study heavily depends on the type of information that needs to be gathered and the expected outcome [28]. Hence, before I elucidate upon my research, it is important to discuss the research techniques I selected in order to conduct my study [2].

In this chapter, I discuss the data gathering instruments I selected in order to explore the thought processes of the students regarding their interactions with the developer social platforms under study. I began by revisiting the studies I discussed in Chapter 2 paying special attention to the research techniques they used. This helped me understand the current state of research techniques used in empirical studies and choose the best techniques that would result in usable findings. In the next chapters (chapter 4, 5) I discuss the steps I undertook to gather data, the analysis process, and how I obtained the results for each of the data gathering instrument selected.

3.1 Mixed Methods Research

Mixed Methods research is defined as research that involves a combination of different qualitative and quantitative research techniques[39]. Some researchers argue

that mixed methods research is one of the major **research paradigms** (Qualitative Research, Quantitative Research, and Mixed Methods Research) used to conduct studies using either a qualitative approach or a quantitative approach as the main research technique and the other one as a validation technique [33]. Mixed method approaches can be labour intensive as they involve numerous stages of data collection. They provide a greater breadth of perspectives around certain issues, and combining approaches aids in overcoming deficiencies that arise when utilizing only one method [27].

I opted for a mixed method approach as it offered the flexibility required for gathering the intricacies of qualitative data gathered from the participants which I thought was useful for an exploratory study. This approach aided in data collection as it both captured intricacies found in student experiences, and helped reflect the nuances found in the barriers and benefits evinced in said student experiences.

3.2 Data Collection for my Study

Qualitative techniques can yield valuable, insightful, and rich data. Qualitative techniques can be used on their own or in conjunction with other research techniques depending on the nature of the research project [27]. An important goal of my exploratory study was to gather the experiences of the participants of my study which are specific to their interactions with developer social platforms during their university careers. Accordingly I decided to use *semi-structured interviews* followed by a *survey* for gathering data regarding student experiences with developer social platforms[1]. The semi-structured interviews helped me closely analyze student experiences with developer social platforms; whereas, the survey helped me quantify the findings I obtained from the interviews. Figure 3.1 shows the steps I followed.

3.2.1 Ethical Considerations

As a part of the design process of the data collection tools, I submitted a human research ethics board application[2] seeking the approval of my research protocol. This incfluded considerations regarding participant anonymity, data storage and disposal,

[1] https://apps.dtic.mil/sti/citations/ADA512853
[2] https://www.uvic.ca/userais/index.php

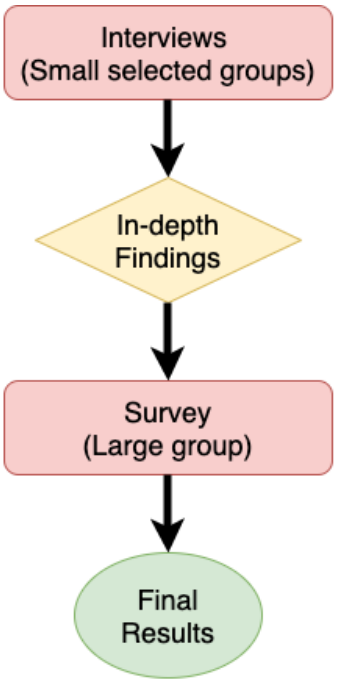


Figure 3.1: Flow of data gathering process for my study

and publication of the findings. The study was approved, with the reference number 19-0554-01, and as a part of this ethics application I submitted the email invitation scripts for both semi-structured interviews (Appendix A) and the survey (Appendix C), an implied consent form, and a copy of the semi-structured interview questions as well as the survey. This ethics package for the semi-structured interviews is attached in Appendices A and B, and the package for the survey is attached in Appendices C and D.

3.2.2 Semi-structured Interview Design

I designed an interview with open-ended questions that were targeted at gathering the nuances and details of individuals' experiences [25]. The main intent behind the use of semi-structured interviews as a data collection tool was to understand the backgrounds of the participants - such as if they are familiar with the usage of developer social platforms, if they have used these platforms at work or school, and whether they use these platforms outside the environments (jobs or school) where they are *required* to use them.

Real interest in participation, along with relevant data on barriers and benefits to participation, become more apparent where an individual *chose* to participate on the developer social platforms on their own volition. For example, if a student is required to complete a school assignment and their professor asks them to submit it on GitHub, then they have to use GitHub, irrespective of how they feel about it, in order to get assessed on the assignment. But in those cases where a student uses GitHub outside such work or school commitments: more insights into barriers and benefits can be gleaned. Hence it was important to understand their personal interactions with these platforms. To understand the above, I was required to design the interviews such that they can elicit data from students with varying kinds of educational and work experience backgrounds. Also, in order to extract experienced barriers and benefits to participation, the interview questions were adapted to the experience of the participant. For example, the questions covered the experiences of all the students, ranging from undergraduate to graduate, as well as from students who have never worked in the industry to people who have experience working in the industry. To achieve this, I used a branching technique which ensured that my questions adapted to the experiences of different kinds of participants. I discuss the details of the design of the interview process in chapter four, specifically section 4.1.

3.2.3 Survey Design

Surveys can be used as effective tools to evaluate the findings of a qualitative study by acquiring data from a larger population[9].

The results I obtained from the semi-structured interviews aided me in understanding student interactions with developer social platforms and revealed a number of the barriers and the benefits that said students experience. The results from the semi-structured interviews were based on the feedback of 20 participants. In order to quantify the results and see if the results were applicable to a larger population of participants, I designed a survey that was based on the findings of the semi-structured interviews. Most of the questions and options that participants could choose as a response in the survey were inspired by findings from the interviews. In chapter 5, I discuss in detail the design and implementation of the survey.

3.2.4 Population

The population I targeted for both the survey and semi-structured interviews were diverse in terms of:

- **Education level:** Students enrolled in undergraduate degrees in computer science or software engineering, graduate students in the computer science department, and the students who recently graduated from these programs.
- **Gender:** Students belonging to different gender identities participated in both semi-structured interviews as well as surveys.
- **Work experience in the software development industry:** Students who participated in the studies were a mix of people who have worked full-time or have participated in co-op programs or have never worked in software development related jobs.

Sample demographics

Participant samples for both semi-structured interviews and surveys represented the demographic diversity I mentioned above regarding the population.

Sample Size

The invite for the semi-structured interviews was sent to approximately 150 undergraduate students and 90 graduate students[3]. Semi-structured interviews demanded one hour of each participant's time which might have affected the response rate of the invitation. As per ethics requirements, a participant could decide to terminate the interview and in that event I had to discard the data gathered from them. 24 participants agreed to participate in total, out of which 4 participants did not complete their interview. The information for ethics related to semi-structured interviews is presented in Appendix B.

As regards the survey, the invite was sent to around 150 undergraduate students and 40 graduate students and I received 52 responses. As per the ethics requirements

[3]The invitation was sent to fewer people because I was working as a Teaching Assistant for two courses at the time and one of the research supervisors was teaching two courses. Due to the power relationship mentioned in the ethics application, I was unable to send invites to many undergraduate and graduate students

for the survey, if a participant did not complete the survey, I was to discard any data that I gather from them. Hence, I had to discard 12 responses as those 12 participants did not complete the survey. The information for ethics requirements related to the survey are presented in Appendix D.

The participant information for the semi-structured interview can be seen in Table 4.1, and the participant information for the survey can be seen in Table 5.1. I discuss the details of the selected participants in chapters 4 and 5.

3.3 Research Team

The dynamics of my research team had an effect on the procedural research choices I made and the data that I decided to collect. The team consisted of three members, and members of the CHISEL research group[4] provided constant feedback about the data collection instruments by piloting the studies. The first part of the thesis, where I conducted semi-structured interviews and identified barriers and benefits, was a team effort with an eye towards a research conference where we were to present the paper. The roles of the members are as follows:

- **Trishala Bhasin** - Principal researcher, Masters student. I conducted the majority of the research in all phases: from preliminary design to reporting the results. I proposed all of my research plans to the rest of the research team for discussion before acting on them, and changed plans where necessary to reflect the judgment of the team as a whole.
- **Dr. Adam Murray** - Co-supervisor and co-author of the research paper, the main consulting researcher throughout the research process. He did not conduct any research tasks himself but was involved in every decision that was made throughout the course of the research. He gave detailed feedback from the first drafts to the final drafts of the data collection instruments.
- **Dr. Margaret-Anne Storey** - Masters co-supervisor and co-author of the research paper, Professor. She was involved in all phases of the research project as my supervisor and acted as a consulting researcher who helped make key decisions during the research process, as well as helped direct the research project

[4]https://thechiselgroup.org/team/

in a way that would generate meaningful insights and have an impact on the software engineering research community.

- **CHISEL Group Members** - These people were both formally and informally involved in the study. Formally as the pilot group for different stages of my study, and informally through day-to-day discussions about thesis content and approaches.

  This collaborative environment helped to ensure that all of the decisions made throughout the research process were carefully considered and discussed before taking action in order to mitigate sources of individual researcher bias which is discussed in detail in Section 7.2.

3.4 Chapter Summary

In this chapter, I discussed the step-wise methodology of my study. I used a mixed methods approach that included a combination of quantitative and qualitative methods to derive findings. I also discussed the flow of my study from the data gathering steps for both of the data gathering tools (the semi-structured interviews and the survey) which included discussions regarding the population I targeted for both my study and the rationale behind my chosen research techniques.

In the next chapter, I discuss the details of the semi-structured interview process from participant recruitment to the results. Then in chapter 5, I discuss how I used the results from the semi-structured interviews to design the survey, and I also discuss the recruitment of the participants for the survey, and the results of the study.

Chapter 4

Semi-Structured Interviews

In this chapter I describe the semi-structured interviews process which was the first part of the data gathering step of my study. The second part of the data gathering process was the survey I conducted, which I discuss in the next chapter. Semi-structured interviews helped me understand the perception that students hold towards their interaction with GitHub and Stack Overflow, and since the interviews were open-ended, they helped me gather the details about the student interactions. I discuss how I identified barriers that students face and the benefits that students experience during their interactions with the developer social platforms under investigation.

The purpose of using semi-structured interviews for data collection was to gain information about personal experiences, attitudes, perceptions, and beliefs related to the topic of interest [14][28]. Then to use this information to answer my study's research questions about developer social platforms participation trends exhibited by students, and to identify the barriers and the benefits experienced by the students. This approach has been used in other similar studies such as the study to investigate barriers that women face in participating in Stack Overflow [21] where the interviews with different members of the community highlighted the issues that are being faced by women on the platform.

It is noteworthy that among all the popular social coding platforms, I chose to investigate student participation on GitHub. My decision to investigate GitHub out of all other platforms was influenced by many studies that indicated GitHub as the most widely used social coding platform. However, I wanted to see if this trend is also followed among the students, hence I asked the participants of my study about the platforms they actively use. 19 out of 20 participants in semi-structured interviews indicated that they used GitHub. Hence, I focused my questions on the student

participation on GitHub.

4.1 Question Formulation Process

The interview questions were a mix of open-ended and closed-ended questions. The questions were mainly centered around identifying the participants' interactions with open platforms that included asking the students who took the interview about their recent contributions on GitHub or their recent participation within online forums such as Stack Overflow.

4.1.1 Types of Questions

The questions that I asked from the participants during the semi-structured interviews contained four sections that gathered different kinds of data. Each of these sections aided me in building strong participant personas and carrying out the process of analyzing participation patterns effectively. Appendix A contains all of the questions that I asked the participants during the semi-structured interviews. The different sections of the questions that I asked from the participants during the interviews are as follows:

- Participant Information: This section asked for information about the participant, such as the degree that they are enrolled in, their current job, their gender, and their year of study. See Appendix A.1 for the full list of questions that fall under this section.

- Generic Questions on Interaction with GitHub: This section contained questions that were aimed at understanding how familiar the participant is with GitHub, and whether they have used GitHub in school or at their jobs. See Appendix A.2 for all the questions that fall under this section.

- GitHub Participation: The questions in this section revealed the extent to which students participate on GitHub outside of the required participation entailed by their school and job. See Appendix A.3 for the questions that fall under this section.

- Stack Overflow Interaction: This section contained questions related to the participant's interaction with Stack Overflow. See Appendix A.4 which contains

all of the questions that fall under this section.

Each section of the semi-structured interviews contained approximately 10 questions along with follow up questions based on the participants' responses. This meant that the follow up questions for a given question for each participant were different based on their experiences. None of the participants answered the same set of questions due to the layout of the questions and the differing experiences of the participants. This branching and openness in the questions aided in gathering responses from the participants that reflected individual experiences in greater detail.

4.1.2 Branching and Logic Design

I conducted interviews with different types of participants, some were active participants, while others were not. Hence, as mentioned above, the flow of questions differed for each participant. As such I implemented branching in the interview so that I did not ask questions that were not applicable in the case of a given participant.

Figure 4.1 represents the logic I followed in designing the interview questions. As seen in the figure, the questions that the participants answered were different for each one of them and were based on each participant's individual experiences.

As shown in Figure 4.1, The yellow and purple blocks indicate branching which means that while answering the questions related to general interaction with GitHub, if a participant is active, they answer the questions corresponding to a yellow box and if a participant is not active, they answer the questions related to purple blocks.

In the next two sections, I discuss the participant recruitment process and the details about how I gathered the data from the semi-structured interviews.

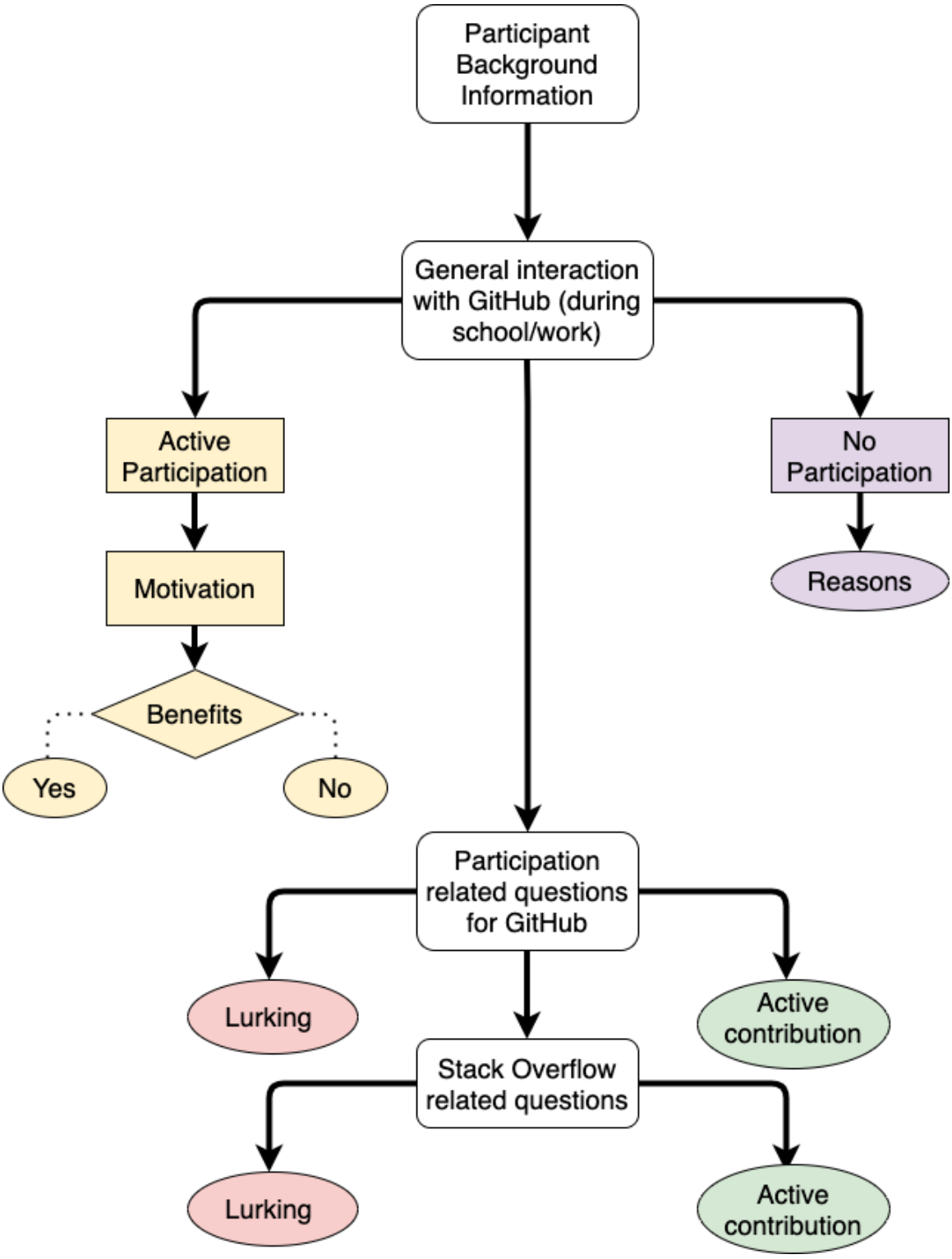


Figure 4.1: Flow chart presenting the logic followed for question formulation process

4.2 Participant Recruitment

I decided to conduct the interviews with more senior undergraduate, graduate, and newly graduated students as many of them would have had some co-op/internship experience(s) in their courses. Said experiences would have introduced the participants to the platforms in question, and may have lead to the participants being more inclined to use and contribute to the platforms.

4.2.1 Participant Selection Criteria

In order to be more inclusive and to collect results from students at different stages of their degree I sent out an invitation (as described below) to three cohorts of students:

- Cohort 1: Students in 3rd year or higher of an undergraduate Computer Science (BSc) or Software Engineering (BSEng) degree. This cohort spends most of their time on course work.
- Cohort 2: Graduate students taking a Master's (MSc) or Doctoral (PhD) degree in Computer Science. This cohort devotes time toward research and graduate projects.
- Cohort 3: Students on a paid co-op/internship or that recently graduated and are working for a software development company. This cohort spends a large amount of time on development tasks relevant to their employer's needs.

The invite was sent using mailing lists to all of the students registered in 3rd year, and above, of computer science and software engineering. This also included graduate students in computer science and new graduates from both software engineering and computer science. I received 45 responses, out of which I conducted interviews with 20 students based on their availability, job status, registration status as a full-time student, and willingness to share their in-depth experiences of the platforms. I recognize that the response rate was low (as I sent the invitation email to participate to more than 100 students), but as this was an exploratory study demanding one hour of their time and many students or new graduates might be busy, this low response rate was not surprising to me.

4.2.2 Selected Participants

Table 4.1 demonstrates the demographics of the participants that participated in the interviews. The range in the participant pool helped me gather barriers from different kinds of interactions with the developer social platforms.

Table 4.1: Information about our interview participants

Cohort	P No.	Degree	Gender
C1	P2	3rd Year, BSc, Computer Science	M
	P10	5th Year, BSEng, Engineering	M
	P12	5th Year, BSc, Computer Science	M
	P17	4th Year, BSc, Computer Science	NB
	P20	4th Year, BSc, Computer Science	-
C2	P1	4th Year, PhD, Computer Science	M
	P3	2nd Year, PhD, Computer Science	F
	P4	1st Year, MSc, Computer Science	M
	P5	2nd Year, MSc, Computer Science	M
	P6	2nd Year, MSc, Computer Science	F
	P7	2nd Year, MSc, Computer Science	-
	P8	4th Year, MSc, Computer Science	F
	P9	2nd Year, MSc, Computer Science	M
	P11	2nd Year, MSc, Computer Science	M
	P19	3rd Year, PhD, Computer Science	M
C3	P13	BSEng (Work Term), Engineering	M
	P14	BSEng (Work Term), Engineering	M
	P15	BSc (Full-Time Job), Computer Science	M
	P16	BSc (Work Term), Computer Science	F
	P18	MSc (Work Term), Computer Science	M

4.3 Data collection through Interviews

In this section, I discuss the steps I took in order to collect data from the participants I recruited. The interview process was intended to be conducted in person, but due to COVID-19 restrictions[1] that required social distancing I conducted interviews via Skype[2].

[1]https://www2.gov.bc.ca/gov/content/covid-19/info/restrictions
[2]https://www.skype.com/en/

4.3.1 Before the Interview

I began by sending the draft email to the administration members of the computer science and software engineering department that included the details of my semi-structured interview study and a request to get access to mailing lists for different cohorts of participants. In Appendix B.1 I have included the email that I sent to the administrative members of the departments.

In some cases, I gained access to an appropriate mailing list and for other cases, I sent out the draft recruitment email to the department administration and they sent the email on my behalf. In Appendix B.2, I have included the recruitment email that was sent to the students inviting them to participate in the interviews.

For those participants who responded to my recruitment email, I sent an email to the participants to set a time and date for the interview along with a consent form that discussed the confidentiality of the data I collect from them. Participants responded with their availability along with the signed consent form (the unsigned version has been included in Appendix B.4). I then shared a meeting invite that contained the Skype link with the participant.

4.3.2 During the Interview

The participant joined the video call and I read to them the verbal consent which I have included in Appendix B.3 in which I remind them about the anonymity of the answers they provide and their right to leave the interview. Once they agreed to continue with the interview, I began to ask the questions that constituted the substance of the interview. During the interview, I took notes on their responses and used a voice recorder to record their responses in full. I conducted the interview in the same way I would have done if it was an in-person interview.

4.3.3 After the Interview

Once all the interviews were completed, I transcribed the recorded interviews onto a Microsoft Word document. I removed all of the identifiable information of the participants from the transcribed documents, and uploaded the documents to Atlas.ti[3] from where I further analyzed the information.

[3]a qualitative data analysis tool https://atlasti.com/

4.4 Analysis of Semi-structured interviews

After receiving all of the participant responses, I further read and reflected on the data gathered. My objective was to conduct a 'thematic analysis' [16][2] of the transcribed interviews. I discuss the process of assigning the codes in detail in the upcoming sections. I briefly followed the approach described below to conduct interpretive thematic analysis which is as follows [18]:

1. Specify the objectives of the study
2. Read the transcripts completely at least twice
3. Select the coding unit (for example, a line of transcribed text)
4. Encode the units by their content
5. Group the units by codes (categories)
6. Identify the connections between categories (themes)
7. Select representative extracts of the text to exemplify categories and themes

Most of the analysis was conducted using the Atlas.ti tool.

4.4.1 Generating Codes

I started with capturing the protocol into a MS Word document; then I exported the transcription to the Atlas.ti tool. Then I saved the transcription of the responses related to GitHub in a separate set of documents from the transcription of the responses related to Stack Overflow. The responses gathered for both GitHub and Stack Overflow analyzed separately; however, they were analyzed in the same way - the steps were as follows:

Step 1: I went through the transcription of the first interview and highlighted the statements that could resemble to being associated by one particular type of barrier. For example, when a question in the interview was asked about participation in a developer social platform[4] e.g., *"If a project is a private repository from one of the persons that you know, do you feel that you're more comfortable raising the issue and generating pull requests for fixes, vs. when it is open source? "* , a participant answered *"Yeah I would totally be more comfortable in a private repository as in a huge open-source project there are so many people and contributors it is very hard*

[4]The quoted language is presented as spoken

to get your contribution here as there are so many people working on their things." This response indicated a hesitation to participate when the projects are open but a comfort to participate when it is a private project, hence I highlighted this text.

Similarly, an example of highlighting a participant response from the data gathered from the questions related to Stack Overflow is *"when you are looking for a specific problem that you face, you find the answer on Stack Overflow and you come up with a better answer, do you post the improvement back?"*. The response was *"My code is not generic and is related to my specific problem, my code is customised for my own problem so it will not really help others"*. This response indicates a potential barrier so I highlighted it.

Step 2: Then I went over the highlighted instances that looked similar to each other and assigned a code to them. For example, the responses such as *"If fixing the issues is not taking much of my time, I will fix them, otherwise I will not"* and *"The projects were so big that if I started fixing one or two issues then it would take several days investing that much time is not possible for me at the moment"* are similar to each other as they both reflect the time consuming nature of the process of raising issues in open source projects: the participants do not contribute for this reason. I marked them with the same code, in this case it was *Time-related barriers*.

Step 3: To verify the coding scheme I used, I made use of a feature in Atlas.ti tool called 'Inter-coder Agreement Mode' which shows the degree of overlap between the coding of more than one team members which helps to monitor the agreement of codes of various team members. I provided a base set of codes to each team member, they had to add instances to the preexisting codes or add the missing codes independently. Then I held agreement sessions with my supervisors to update the existing codes. For consistency in the process, I created a coding handbook that I shared with the team members. I have presented it in Appendix F.

Step 4: While I was working on coding the data and finalizing the codes I made notes corresponding to each code in the Atlas.ti tool in order to reflect the perspective of the research team on the potential barriers. Figure 4.2 is the snapshot from the Atlas.ti tool which shows the comments made by the research team on all the barriers.

4.4.2 Clustering Codes and Identifying Themes

After obtaining the codes from the interview data, I wanted to identify the underlying patterns in the codes in order to draw insights. These underlying patterns would reveal the theme of the data being gathered. In order to derive the theme of the data, it was important to group the similar codes together to understand what these groups of codes collectively represent. The themes in the data later translated to the barriers and the benefits that I identified in my study. Hence, I conducted cluster analysis for qualitative data [36] in which I grouped the related codes that represented similar responses. The process of grouping was conducted separately for codes associated with GitHub and Stack Overflow. I undertook the following steps in order to derive the themes from the interview data:

Step 1: Since I had no previous grouping in mind I began with an open card sorting technique [45]. I used the Coggle.it tool to conduct the grouping process through digital card sorting. In the tool we enlisted all the codes openly along with the explanation associated with each group.

Step 2: I observed similarities in some of the codes such as *"I do not trust strangers online to collaborate in my projects"*, *"I do not trust stranger's algorithms directly to use in my projects"* reflect some underlying trust issues so we placed them under same group associated with trust issues. Using the same rationale, I completed the process of grouping the remaining data.

Step 3: To make sure our open card sorting is consistent across our research team, I conducted the process of grouping the data in a collaborative mode on Coggle.it, and I held regular agreement sessions with the research team that helped us reach agreement on the codes.

The final result of the grouping process is presented in Figures 4.3 and 4.4 for GitHub and Stack Overflow respectively.

In Figure 4.3, I list the 23 codes that emerged from my analysis that I placed in seven groups that would potentially point to barriers that I identified in GitHub.

In Figure 4.4, I list the barriers that the participants mentioned they experienced with with Stack Overflow. For Stack Overflow, 11 codes emerged from my analysis that I placed in 7 groups.

Name		Comment
I am comfortable asking questions from my known people only	2	It could come out of convenience, trust, or ease of communication.
I do not have time to create PR even though I want to, so I just raise issues	4	They feel that it is very time consuming to create a Pull request, so they just raise issues and describe what changes they would like to build in the code.
I do not have time to even raise issues	1	The participants feel they are too busy to even raise an issue for a bug they spotted. They avoid raising issues altogether.
I do not have time to find projects on OS and collaborate	1	It takes a lot of time to first find projects and then finding the areas or spots where I can participate. I do not want to invest that much effort, so I avoid participating
I do not report bugs because I'm not sure if it is actually a bug or a gap in my understanding	2	This means that they're not confident on their judgement of code. If they find mistakes in the code, they would have self-doubts and will not be confident enough to report the bug.
I do not trust strangers online to collaborate in my projects	2	They feel that since they do not know the strangers, they cannot trust their skills in a particular area. Letting them work on their code can not be as useful as finding a person they know who can produce more effect...
I find it faster to collaborate with people around me	1	The communication is quicker and faster when they're collaborating with people around them. Onboarding is easier and faster, less communication gaps tend to occur.
I have never seen any project in that much detail so I never felt like participating	1	It seems like the participants have not taken a concious effort in finding projects where they can collaborate. It reflects a lack of interest.
I raise issues only if it is not taking much of my time	2	If the process of raising issue is less time consuming, only then they would raise issues, otherwise they would let it be.
If creating a PR is a lot of work, I just raise issues	1	They feel that the process of creating a pull request on any project has a lot of steps involved. They do not have enough time or are not willing to put in that kind of efforts into stranger's repository. So they orefer to...
My school work keeps me occupied	2	They feel that they do not have time to participate in open source , because they are always busy with projects, assignments and other school work.
I am comfortable to contribute to stranger's repositories	2	They are not shy or refrain themselves from contributing to other poeple's repositories.
I participate in Hackathons	3	They participate in open competitions and hackathons, which does not have any direct benefit associated with them. The only possible reason could be either to enhance their coding skills or they like coding in thei...
I participate in Open source projects	4	They have participated in open source projects before, either during an internship. while developing something of my own or out of interest.
Peer Pressure affects me	4	If people around me are participating in stack overflow, it does affect me
Previous Work Experience	7	They have worked in a company before where the team used collaborative development process on with version control systems.
Self-score of my skills in github	7	That is how they rate their own skills on github on a scale of 10.
Time since using git	4	This is the duration for which they've been on GitHub.
I am more interested in participating when my work gets attention	1	They do have open projects on github, if the project that they intend to develop for opensource does not get any attention, they get bored and stop working on it.
I participate on only those open-source projects that I am pesonally using	2	They feel that if they are working on a project, it means they are already putting effort with the code there, hence they prefer to only participate in those kind of projects. They do not want to put any extra efforts in...
I am inexperienced, hence I do not participate	5	They believe that they do not have enough skillset to participate in stack overflow, or on GitHub. They also feel that there are many experienced people in the community already, their contribution is more significant...
I do not ask questions as the task of asking questions is too much work	1	They believe that asking questions is a lot of work as there is a format meant to be followed which requires efforts. They are not willing to make those efforts.
I do not trust stranger's algorithms directly to use in my projects	1	The person works in a specific fiels and does not trust the open-source code provider's mathematical ability as they do not know where are they coming from.
I know I will not get attention in any big project, so I don't participate	1	There are so many people already in open-source platforms, these people are more experienced hence they're more likely to get attention on big company projects. The participant feels that they will not get any att...
My problems are specific to my use case, so I do not ask questions on stack Overflow	2	They feel that most of the problems they face might arrise due to a gap in their understanding or because their problem is very specific to their use case. Hence, they do not ask questions on the stack overflow platf...
My projects are not good enough and people have done much better than me already	2	They believe that open communities are huge and people there have made much bigger global level projects, the participant feels that their projects are not as good as theirs.
My projects are too small to get attention from people	1	Partifipant believes that their projects are not very special, unique, or at a scale where it would grab attention from people at an open scale, so They never tried to seek outside collaboration. This reflects underlying...
Open participation is not my hobby	5	They would not prefer to work on open projects even if they have a lot of time. Maybe because they see programming as a way of gaining work experience, it is not their hobby that they want to pursue in their leasur...
If my projects are good enough, I post them online	1	They do not feel that all of their projects are good enough to be posted online, hence they need
I do not see any direct benefit to participation in open source	9	The participant expects to gain something when they are working on opensource projects, since they believe they are not getting anything in return like money, job etc from participating, so they do not wish to inves...
I don't care about other's code	3	The project owner should fix their own code. The participant just use it for themselves and if they make changes, they're for their own work. If those changes were needed, the owner would've done them himself.
It is not my responsibility to fix other's code	5	The participant feels that It is not their responsibility to fix other's code, there are so many people in the community who are much experienced, they should fix the code.
I type my problems on search engine	7	They prefer to type their problems on search engine instead of typing them on stack overflow, and if stack overflow is among the first few links to pop up, they visit it to get help with my code.
Long waiting time to get answers on stack overflow	3	The participant does not post questions on stack overflow because they feel it takes a lot longer to get a response there. They would rather find an answer themselves.
I am more comfortable collaborating within known circle only	9	It is much easier for the participant to contribute within known circle as they do not feel shy within a known group of people, communication can be done easily. and they are more confident in their code
I prefer to contribute to the repositories of people I know	1	They would rather fix code of people they know as they would give them appreciation in return.
I have never sought collaboration online	1	They prefer to collaborate with people around them, reason is unknown. It is coming from an active participant.
It is hard for me to get new people on board for my open projects	1	It is hard to explain to new poeple about what is going on in the project, It is easier to explain people around them though. They prefer to let known people participate in the project
I feel less motivated when I am working alone	1	participants feel that they like seeing people around them working in order to feel motivated. When they're working with a group of people, they can draw their own motivation from pwople around them.
My programming problems/requirements are my responsibility, hence I do not use open-sourc...	2	participant feels that the programming problems they're facing are their own responsibility. They're more comfortable in approaching people around them for collaboration

Figure 4.2: Comments on the codes in Atlas.ti

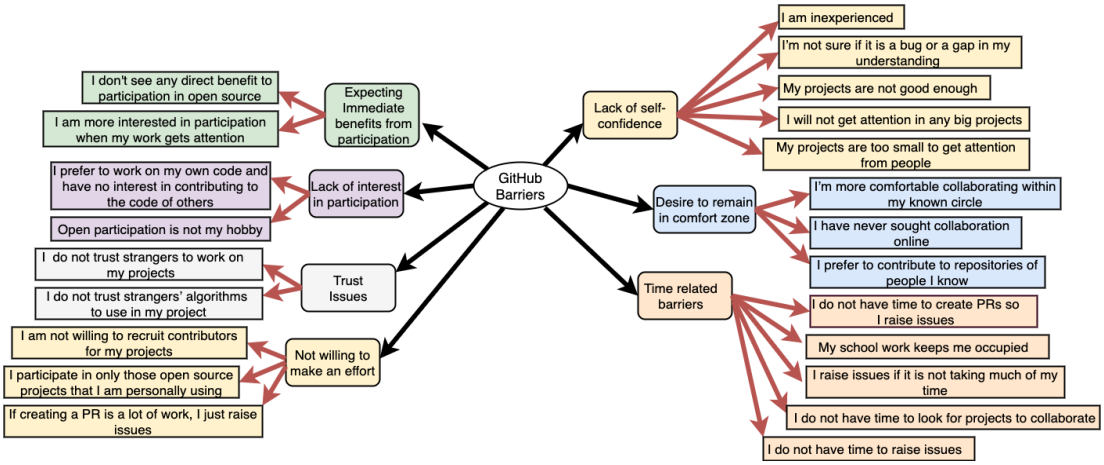


Figure 4.3: Grouping of the codes obtained from interview data associated with GitHub

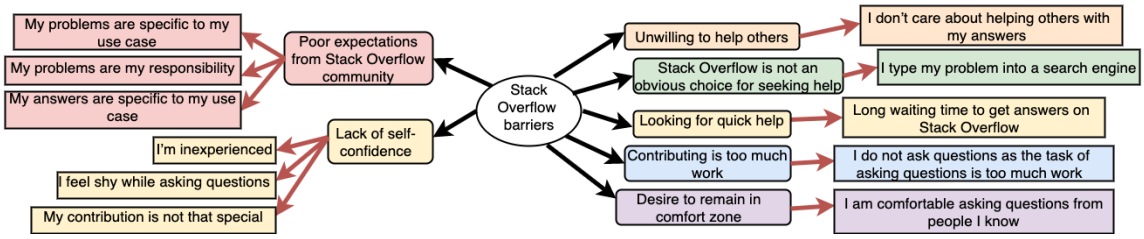


Figure 4.4: Grouping of the codes obtained from interview data associated with Stack Overflow

4.5 Results

In this section, I present the results I obtained from the semi-structured interviews I conducted with different cohorts of participants. I also answer the research questions through my results. I go into more depth about the barriers (RQ2) to help derive recommendations on how to address barriers, as this was the focus of my study, but I also confirm that students do experience the benefits I anticipated before my study.

4.5.1 RQ1: How do students participate and contribute to open content hosted on developer social platforms?

The majority of study participants used GitHub and Stack Overflow at some point during their degrees. As shown in Table 4.2, 19 out of the 20 people I interviewed had an account on either GitHub or Stack Overflow. Out of these, five participants shared that they are very active on both platforms. The table also displays if the participants indicated that they "contributed" to public content that is hosted on these platforms for purposes outside of their school work or co-op jobs.

For GitHub, participants either contributed to private repositories as part of their co-op job or course work, or collaborated with teams of other students for coursework. For Stack Overflow, all participants reported using the platform to find help with their code and development in their specific programming language. However, the degree of interaction with both GitHub and Stack Overflow varied across the participants as students indicated that they access GitHub more frequently as compared to Stack Overflow. I also asked them if they "contributed" content to projects hosted on GitHub or contributed to questions/answers on Stack Overflow.

While most interactions consisted of school or job-related work, some participants indicated contributions outside of these commitments to either open source projects on GitHub or to asking or answering questions on Stack Overflow.

After obtaining these results, I was interested to see whether students participate in developer social platforms outside their school and work commitments. If they do, how does this participation vary across students belonging to different demographic groups. Hence, I gathered this information via the survey and I discuss the results in the next chapter.

Table 4.2: Participant interactions with GitHub and Stack Overflow.

Cohort	P No.	Account Created	Contributed to Open Content
C1	P2	GitHub	GitHub
	P10	GitHub	GitHub
	P12	Both	Both
	P17	Both	None
	P20	GitHub	GitHub
C2	P1	Both	Both
	P3	Both	Both
	P4	Both	Both
	P5	Both	GitHub
	P6	Both	GitHub
	P7	GitHub	None
	P8	Both	Both
	P9	GitHub	GitHub
	P11	GitHub	GitHub
	P19	Stack Overflow	Stack Overflow
C3	P13	None	None
	P14	Both	Both
	P15	GitHub	None
	P16	Both	Both
	P18	Both	Stack Overflow
	Total	GitHub: 7 Stack Overflow: 1 Both: 11	GitHub: 7 Stack Overflow: 2 Both: 7

4.5.2 RQ2: Barriers Students Experienced with GitHub and Stack Overflow

Research has shown that every public-facing community [37] has barriers to entry and participation. To help us answer RQ2, our interviews highlighted the specific barriers our participants experienced with GitHub and Stack Overflow.

Barriers observed with GitHub

I discuss the barriers that I discovered in students' experiences in their interactions with GitHub. I also present the codes associated with each of these barriers along with an example response from the interview that pointed to the existence of the barrier.

1. **Lack of self-confidence.**

Some participants believed their programming skills were not as good as other people who were performing similar activities on GitHub such as hosting their projects for collaboration, contributing to other people's work, or posing questions on various repositories. Some indicated a feeling of inferiority with respect to their programming skills and felt that others will not understand their code. The codes and some of the quotes that reflected this barrier are presented in the table below. When I looked into the literature around lurkers in online communities [3], this barrier was also prevalent in other communities. Lurkers who feared receiving judgement from the community about their work did not trust the level of their own work and were consequently uncomfortable and reluctant to post online.

Code	**Example of the Participant Responses from the Interviews**
I am inexperienced	"There are online communities dedicated to the code bases. I know it is being taken care of by people who are more experienced than me." (P2)
I'm not sure if it is a bug or a gap in my understanding	"I never found a flaw I was sure about. I am never sure if it is bug or just my perspective." (P20)
The projects I created are not good enough	"I am sure there is a better version of my code somewhere there, someone else can do better than me." (P15)
I will not get attention in any big projects	"I am more comfortable collaborating privately as then my contribution matters. It is hard to get your contribution seen in bigger and more open projects." (P3)
My projects are too small to get attention from people	"I posted my class work once but ever got any stars or other activity so I figured I need to post something big and important." (P5)
If my projects are good enough I post them online	"None of my existing work is good enough to be open." (P15)

2. **Desire to remain in comfort zone.**

Some participants indicated being more comfortable when interacting with peo-

ple within their known social circles. They noted that they avoid interacting with people outside their circles as they feel *shy* with outsiders which is also observed in a study on social barriers faced by newcomers [48]. This may be due to the perception that within a known circle they can "be wrong", a feeling they are uncomfortable sharing with outsiders. Some participants also displayed introversion as they were not comfortable interacting with outsiders even in the form of comments.

Some participants indicated being more comfortable when interacting with people within their known social circles. They noted that they avoid interacting with people outside their circles as they feel *shy* with outsiders.

Code	**Example of the Participant Responses**
I'm more comfortable collaborating within my known circle	"I don't like to interact much, I hesitate a lot and it is in my nature. The thought of having to interact with so many people at the same time scares me." (P2)
I have never sought collaboration online	"I don't remember going to Twitter and publishing 'I need collaborators for this project' I usually ask people I know." (P1)
I prefer to contribute to repositories of people I know	"It is hard to fill my knowledge gap for repositories of strangers if their documentation is not complete. It's not like I can just ask them like I would if I knew them." (P4)

3. **Time-related barriers.**

Some participants indicated an interest in participating in open-source communities but could not do so because of other commitments. The undergraduate students who mentioned this were occupied with their school work. Some participants with more experience, such as the MSc and PhD students, indicated being unable to finish contributing to projects due to time but they pose questions and add comments instead, which is less time consuming than creating a pull request. It is noteworthy that this does not appear to come from a lack of interest, but rather the participants worried they lacked the time to ensure their

code contributions were of sufficient quality. This barrier also emerged prominently in a systematic literature review of 20 studies that were aimed at understanding the barriers faced by contributors of developer social platforms [50].

Code	Example of the Participant Responses
I do not have time to create PRs so I raise issues	"I participate in open source but I don't have much time so I usually report issues, and sometimes I would create a branch and start working on it and maybe create a PR. Then I stop following it." (P1)
My school work keeps me occupied	"Usually during an academic term I have a heavy load of mid-terms and assignments and I cannot do anything else." (P13)
I raise issues if it is not taking much of my time	"If the issue is small enough for me to explain, only then I raise it." (P1)
I do not have time to find projects on open source and collaborate	"I'm busy with my research and collaborating in other people's repositories requires time and effort that I cannot put in at the moment. Maybe after I graduate I might." (P18)
I do not have time to even raise issues	"I am always running on a tight schedule, I have never bothered to even point out any issue. After my Master's thesis is done I will have less stress and more active time there." (P6)

4. **Expecting immediate benefits from participation.**

   Actively participating and collaborating with people can help grow one's professional and technological skills. When one participates in online communities, they can enhance their coding skills, contribute across a diversity of projects, develop collaborations with others regardless of geographical boundaries, have an opportunity to learn from people with different expertise, and enhance collaboration skills. However, some participants expected an immediate return on investment from their contributions and failed to recognize the long-term benefits. This was also observed in a study conducted on Microsoft online communities[23]. The participants in the Microsoft study mainly perceive incentives to include situations where an employee completes tasks and in return is remunerated or where a student works on class projects and in return receives

grades.

Code	Participant Responses from the Interview
I don't see any direct benefit to participation in open source	"It is not giving any immediate value in return as compared to some other development activities like competitive coding. It is not beneficial to me in short or long run." (P9)
I am more interested in participation when my work gets attention	"When you don't get recognition from other people, it gets boring. I have released many projects and when they don't get any attention, I get bored and stop working on it." (P11)

5. **Lack of interest in participating.**

   Some students indicated a lack of interest in participating in open-source projects hosted on sites like GitHub. If they had free time, they would rather participate in other activities, such as hobbies. They see the act of working on projects only as an activity that would add to their professional profiles and aid them in getting jobs. They mostly had no intent of or motivation for contributing back to these open-source projects. This lack of interest is also displayed by other people who use the content that is hosted on the developer social platforms but do not post anything of their own [37].

Code	Example of the Participant Responses
Open participation is not my hobby	"Contributing to open source has always been an extra step towards my career. It is not something I would do for fun as it is not one of my main hobbies." (P9)
I have never seen any project in that much detail, so I never felt like participating	"I have never felt like working on other people's projects on GitHub. I'd rather make my own projects and work on that." (P11)

6. **Not willing to make an effort.**

   The interviewees whose responses pointed to this barrier lie somewhere between 'time-related barriers' and 'lack of interest in participating': they want to par-

ticipate but are not interested in taking the extra steps needed to contribute. Some noted they would only contribute to software projects that they're personally using. These participants would occasionally make pull requests or leave comments on the repositories they're already accessing as it required minimal extra effort. They did not always share their projects online and they preferred to collaborate with people they could meet in person as they felt communicating through issues and comments on repositories was too much work.

Code	**Example of the Participant Responses**
I participate in only those open-source projects that I am personally using	"I only fix bugs on projects if it affects me in some way like I need to use the code for my own project." (P14)
I am not willing to find people to get them on board with my projects	"Online collaborations require a certain way to communicate like reviewing PR, raising and resolving issues which is extra work that I have failed to keep up with in the past." (P3)
If creating a PR is a lot of work, I just raise issues	"Making a PR means following community guidelines and then writing code, waiting for comments to edit it again. I'd rather just report bugs." (P1)

7. **Trust issues.** Some of the participants in our study do not share their code on these platforms (particularly algorithmic issues they encounter) as they do not trust the expertise of people they have never met. These students prefer discussing and developing their code with peers they trust. They also feel that the problems they're facing require qualifications, such as coding proficiency, that are hard to verify in people they have never collaborated with before.

Code	Example of the Participant Responses
I do not trust strangers to work on my projects	"I do not know if a stranger will be able to understand my project and there are chances that my work can be misinterpreted." (P8)
I do not trust strangers' algorithms enough to use them in my project	"With algorithms you have to understand math behind it and the theory behind it. I don't know what kind of knowledge they have in my project domain so it is hard for me to open my projects for contribution." (P10)

Barriers observed with Stack Overflow

Some of the themes that emerged in this section are similar to those observed with GitHub, however, barriers on both platforms are different.

1. **Stack Overflow is not an obvious choice for seeking help.**

   When some of our interviewees discover issues in their code, they type their problems into search engines and use whatever links appear on the first page of search results. If a link to Stack Overflow is in the search results, then they will try this first. For them, Stack Overflow is just another website that might contain an answer to their question at that moment, and not a place to routinely search for answers and make regular contributions. This barrier is also highlighted in a study that analyzes the responses and the respondents of the questions being posted on Stack Overflow [59].

codes	Participant Response from the Interview
I type my problem in a search engine	"I only go to SO if I have a very specific request that I cannot solve after say 5-7 days." (P9)

2. **Looking for a quick answer.** Some participants indicated that they visit Stack Overflow to look for solutions to the problems they encounter in their code, but they never post a question if they do not find the exact solution they are looking for. They feel that it takes too long to post questions and then wait for responses from other members of the community. These participants prefer

to read various resources and determine the solution themselves. This barrier is also observed among the people who are new users of the Stack overflow platform [48].

Code	Example of the Participant Responses
Long waiting time to get answers on Stack Overflow	"I am looking for quick help and I know SO will take time to give an answer to any question I post there." (P7)

3. **Contributing is too much work.**

   Some interviewees were interested in participating on Stack Overflow, but felt that doing so was too much work due to community standards and the formatting required to post questions, even the responses to questions need to follow formatting rules.

Code	Example of the Participant Responses
I do not ask questions as the task of asking questions is too much work	"I noticed that there was a kind of standard way of asking questions. If it is not that way then you get down-voted very fast, there have been times when my questions got closed because of not being appropriate. Over time I realized that asking a question is not enough, you also need to say what you have already tried and you need to give references and parts of the source code and different links to show people what you tried." (P19)

4. **Desire to remain in their comfort zone.**

   This barrier is similar to the GitHub barrier 'Desire to remain in comfort zone'. Some students mentioned they are too shy to participate in such communities as they prefer not to share their questions and answers outside their known circle. Some try to avoid interactions with people they do not know which is reported as a barrier to participation in the studies that are aimed at understanding the participation on Stack Overflow [52].

Code	Example of the Participant Responses
I am comfortable asking questions from my known people only	"I am afraid of the response I will get if I post a question. I was once being told that my question is absurd when I posted it on stack overflow so I stopped posting." (P8)

5. **Poor expectations from the Stack Overflow community.**

   Some participants felt their issues tended to be specific to their projects, and that if they post questions on Stack Overflow, they would not make much sense to anyone else. This misalignment prevents these participants from posting questions at all. Other participants shared a hesitation towards seeking help with their problems as they believed they should do more research themselves, they felt that their questions may look absurd to the community. Hence, they avoided relying on Stack Overflow to get help with their code or even sharing a response they might have to another user's question. This barrier also emerged in a study on new comers to open source platforms [48].

Code	Example of the Participant Responses
My problems are specific to my use case	"Last time I asked a question was around four years ago and at that time my question was very case sensitive." (P16)
My problems are my responsibility	"How will I learn if I see the answer instantly?" (P17)
My enhancement of answers are specific to my use case	"I think they're suitable for my particular use case. Their answer is more generic and my answer is tailored to meet my needs." (P14)

6. **Lack of self-confidence.**

   Some participants felt they were not skilled enough to answer others' questions, and as a result, they refrained from participating. A similar theme appeared in response to contributions on GitHub.

Code	Example of the Participant Responses
I'm inexperienced	"SO is a very specific question answer place, sometimes I do not even understand whether context of the question is what I am thinking." (P20)
I feel shy while asking questions	"I don't know, maybe I do not want to deal with interaction that follows the asking of a question there." (P11)
My contribution is not that special	"Even if I come up with an improvement, I feel like it is not something special or new that nobody would have thought of." (P15)

7. **Unwilling to help others.**

   This barrier is similar to one of the reported GitHub barriers. Some students do not participate as they have no desire to contribute to Stack Overflow and give back to the community; they believe that reading and using answers is enough for them. They only see the responses they receive as a search result and not as a place where they could potentially interact with others by answering questions and/or asking questions of their own. This mindset is explained in another study that identifies differences between active users and lurkers [23].

Code	Example of the Participant Responses
I don't care about helping others with my enhanced answers	"I don't care enough to share my improvement of an answer." (P11)

4.5.3 RQ3: Benefits to Participation

I drew examples from the interviews that show the benefits that the participants of my study experienced from making active contributions on the developer social platforms under investigation. I then categorized the examples into codes which I further organized into clusters that represent a certain type of benefit. I describe these benefits below, noting which platforms they apply to (GH, SO, or both). Quotes from the interviews are included as examples, and I note which interviewee is responsible for each quote with (Px).

Benefits from participating in social coding platforms like GitHub

1. **Learning from the Community**

   Developer social platforms offer a space for people in the development field to share their code, collaborate with each other, and get help with their problems. This sense of belonging to a community is reported as a driving factor to contribute to the content that is hosted on open source in the existing literature [26].

   - **Gaining experience in collaborative development**

     Participants shared that as they contribute to repositories, they gain experience working with a team. This helped them grow the soft and technical skills needed for jobs where they were expected to work collaboratively. Some statements from the interviews that point to this benefit include: *"There is always an opportunity to learn a lot from other people, sometimes you see varied expertise, so you get chances to learn." (P3)*

   - **Learning from other people's work**

     Some participants shared that whenever they were looking for ways to accomplish a certain task, they looked at other people's code on these platforms. The fact that code is available publicly and can be accessed easily gives people an opportunity to explore other development styles and gain insights from other people's work. One of the interviewees noted that *"There is always an opportunity to learn a lot from other people, sometimes you see varied expertise, so you get chances to learn." (P19)*

   - **There is always support available**

     Some participants stated they enjoy the development support that is always available whenever they get stuck on a problem. When asked about the benefits they experienced while developing projects collaboratively, one of our interviewees responded with: *"Huge community support is available readily without geographical boundaries, no centralization of power, sense of equality."(P14)*

2. **Altruism**

Some of the more experienced students that participated in our interviews stated they like to share their knowledge with everyone for public good.

- **I like it when others find my code useful**

  I asked the participants who share public repositories of their coding projects how it feels getting recognized by other people. One noted that *"There are two projects that I published recently, I noticed that students are using for their class assignments, they email me questions and report errors. I feel that it is cool that some people find it useful." (P1)*

- **I report bugs on projects as it makes the product better for everyone**

  This was noted by participants who use open-source communities for their own work. They appear to not only use the code but also care about the code base as a whole: they usually report any bugs they spot so that the project can be improved. As one interviewee mentioned, *"There is a framework that is relatively new, I'm using that and it is very good and useful. Every time I come across a bug, I report that as I feel it will make the product better for everyone." (P4)*

However, in some of the previous work, altruism is reported to be arising from an intrinsic motivation to seek validation on one's programming abilities [23].

3. **Increased confidence from displaying skills online**

Some participants experienced increased confidence when they shared their code online. These benefits that are being experienced by the individuals help them increase their value as a human capital that can be beneficial for the student as I discussed in the section 2.4 of the thesis [6].

- **I use my online repository as a portfolio**

  Many participants reported using their GitHub accounts as their online development portfolios, keeping them up to date with their recent work and referencing them on their resumes. They actively browse other public repositories and contribute to them (with PRs or comments) to show they're regular members of the software development community. Some examples of this include: *"My GitHub has a lot of projects I worked on,*

and I mention my GitHub on my resume" (P2), "Once a company asked me for my GitHub page, and they didn't select me as I didn't have much going on there. Since then I make sure to stay active there" (P15).

- **I like the attention I get when people use my code**

  Some participants reported that they feel good when they are recognised on developer social platforms. This feeling affirmed their skills and helped motivate them to work harder. One of our interviewees mentioned, *"If you get the attention, recognition, it is always nice. Makes me feel good about time I spend developing the project." (P1)*

Benefits from Stack Overflow

1. **Learning from the community**

   - **Learning new ways of solving a problem**

     Some participants indicated during the interviews that their interaction with Stack Overflow is not limited to finding answers to their programming problems, but they also use Stack Overflow to enhance the programming logic that they use while writing code. This enhances the quality of their work. For example *"With new libraries being developed for Python for data analysis, it is a good place to find new ways to write scripts" (P18).*

   - **There is always support available**

     This is a very popular benefit that is being offered by Stack Overflow as it is a popular choice for seeking help with coding problems. For example *I always find what I am looking for" (P19)*

2. **Altruism**

   Some of the more experienced students that participated in our interviews stated they like to share their knowledge with everyone.

   - **I like helping fellow coders by answering their questions**

     Some participants stated they like to answer questions as it is their way of giving back to the community they use for help. They also felt that responding to questions takes little effort compared to how much an answer

might help someone who needs it. For example: *"If I know the solution, why not post it? I know how frustrating it is to find answers." (P1)*

3. **Increased confidence from displaying their skills online**

   Some participants experienced increased confidence when they shared their code online. This is observed in the previous work aimed at understanding the behaviours of the contributors to the developer social platforms [52].

   - **I like to improve my reputation** (SO)

     Some participants liked to maintain a reputation on their Stack Overflow profile by collecting badges. They achieve these by actively participating on the site: they up-vote and down-vote answers, respond to questions, and help moderate the site. These activities make them feel confident in their skills and are referenced on their resumes to show prospective employers. As one interviewee noted, *"I try to post an answer next to other answers telling them that well, other answers are correct as well but I came up with this one maybe you wanna take a look at this. It might receive up-votes that increases my reputation." (P4)*

4.6 Summary

In this chapter, I discussed the semi-structured interview process along with the details about the participant recruitment, how I conducted the interviews, and how I used qualitative data analysis techniques to draw results. I used the results I obtained in semi-structured interview to design a survey and gathered insights from a larger population. I discuss in detail, the survey design and distribution process in the next chapter along with the results I obtained from the survey study.

Chapter 5

Survey

After obtaining the results the from semi-structured interviews with 20 participants, I designed the survey firstly to gather more information on the participation of students on developer social platforms, and secondly to see how my findings from the interviews applied to a larger number of participants. The survey addressed the research questions of this study that were aimed at student participation, and identifying the barriers and benefits experienced by the students when they interacted with developer social platforms.

It is important to note that even though the questions I administered in this survey were aimed at understanding general participation activities that are carried on all the social coding platform (such as sharing projects on public or private repositories, cloning projects, raising issues, creating pull-requests), I was curious to see how popular other social coding platforms such as GitLab and bit bucket were among the students as compared to GitHub which is the most popular social coding platform [10]. Similar to semi-structured interviews, GitHub emerged as the most popular social coding platform as all the participants indicated that they had an account on GitHub. The language I used as I discuss the findings of the survey uses the term 'social coding platforms' to denote the activities that students performed on GitHub as most of the student mentioned that they use GitHub to share their code and barely use other platforms. I share more information about this in the upcoming sections.

5.1 Survey Design and Piloting

The survey was designed to integrate findings from the interviews whilst leaving room for broader insights that could be drawn from the survey itself. Hence, the questions I included in the survey were very similar in nature to the questions I asked during the interviews, except, the responses to the questions were mostly closed ended. The survey was designed using Survey Monkey[1] which is a popular tool for designing adaptive surveys[2] and is easy to use.

5.1.1 Survey Logic Design

In this section, I discuss the categories of questions available in the survey and the branching used in the survey along with the logic of the appearance and organization of the questions based on a participant's previous response. Appendix C contains the survey questions. The categories of questions that I asked the participants are given below:

1. **Demographic information of the participant**: This part of the survey sought out basic information about the participants such as *"What gender do you identify as?"*, *"What degree are you enrolled in/or have recently completed"*, and *"Have you worked in a full-time job related to software development before"*. Appendix C.1 contains the questions related to this section.

2. **Social coding platforms and their usage in school**: In this part of the survey I placed the questions that inquire into whether the participant has used social coding platforms as a part of their school curriculum, or whether they have used social coding platforms as a part of their assignments. Appendix C.2 contains the questions related to this section.

3. **Personal interaction with social coding platforms**: In this part of the survey I included the questions that helped me identify how students use social coding platforms for their own personal work. Appendix C.3 contains the questions related to this section.

[1] https://www.surveymonkey.com

[2] Schouten, B., Peytchev, A., Wagner, J. (2017). Adaptive survey design. CRC Press. Statistics in the Social and Behavioral Sciences

4. **Interaction with Stack Overflow**: This part of the survey contained questions related to student interaction with Stack Overflow. The questions ranged from asking whether participants had Stack Overflow accounts, whether or not they actively contribute, and, if they do, how they contribute. Appendix C.4, C.5 contains the questions related to this section.

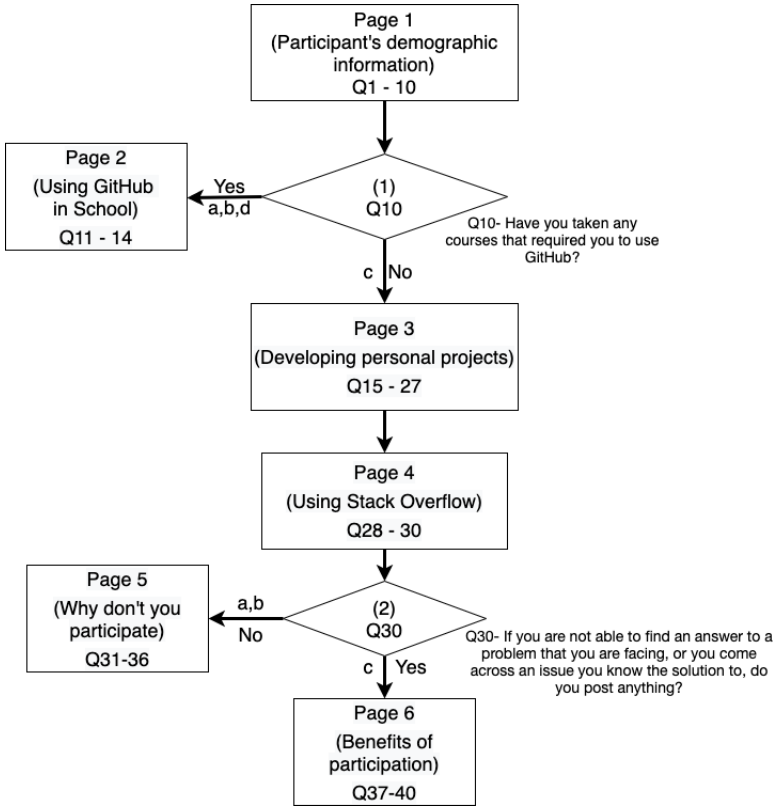


Figure 5.1: This flow chart presents the logic of the appearance of questions based on the options(a,b,c,d) that the participant select in the survey

As presented in Figure 5.1, the survey contained six pages:

- Page 1 contained the demographic questions for the participants that each participant answered.
- Page 2 contained questions about the usage of social coding platforms such as GitHub and Stack Overflow during school, and participants can only reach the question on this page if they indicate that they use social coding platforms in their schooling.

- Page 3 contained questions about the personal interactions of students with social coding platform excluding school and work.
- Page 4 contained general information about participant relationships with Stack Overflow such as whether the participant has an account or not.
- Page 5 contained information related to participation on Stack Overflow which was only accessed by the participants if they indicated on Page 4 that they use Stack Overflow.
- Page 6 contained the last question that asked participants if they wanted to share any additional information.

5.1.2 Pilot Surveys

The steps I followed to conduct the survey are shown in Figure 5.2 and explained below.

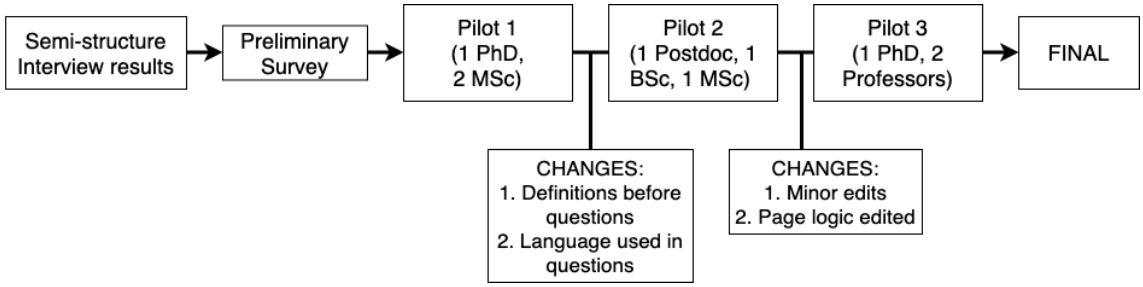


Figure 5.2: Flow chart presenting the steps undertaken when piloting the survey

1. I began writing the survey after analysing the results from my interviews.
2. The first pilot was sent to one PhD student, and two MSc students from my lab who were not involved with my research. They attempted the survey, and provided feedback. I made revisions to the questions based on the feedback received from them.
3. Next I sent out the revised version of my survey to one post-doc researcher, one undergraduate student, and one MSc student. They timed their attempts of the survey and provided feedback. This time, there were no major revisions identified, just minor edits in the wording of the response options in the survey.

4. Along with accommodating the feedback I received, I also analyzed the data gathered from the pilot surveys to make sure that I receive the information I require to answer my research questions.

5. After tweaking my survey and getting my changes approved by my supervisors, I sent out the survey for the third pilot round to a PhD student and the lab coordinator for their feedback.

6. This time there were no issues identified, so I proceeded to share my survey with my supervisors for their approval followed by distributing the survey to the participants.

5.2 Survey Dissemination and Data Collection

In this section, I describe the details of the recruitment of the participants, the demographic information of the selected participants, and I also explain how I collected data. My goal was to disseminate the survey to a large number of participants and gather responses from a diverse range of students.

5.2.1 Participant Recruitment

The process of recruiting the participants for this survey was similar to the recruitment process for semi-structured interviews. I sent out invitations to participate to the three cohorts of students that I discussed in Section 4.2. I received replies to the invitations from the interested participants and sent them a link to my survey along with a consent form that contained information related to data and the handling of the responses. The survey took 12 minutes on average to complete.

5.2.2 Data Collection Process

The survey was distributed via email and used the Survey Monkey tool. I sent the invitation to participate in the survey to approximately 200 students (approximately 40 graduate students in the computer science department and 160 students enrolled in bachelor of computer science and software engineering degrees). The tool showed how many people responded to the survey, how many people completed the survey, and the participant drop-out rate. For my analysis, I only considered the data collected

from the participants who completed the survey. I kept the links for participants to attempt the survey open for 14 days (2 weeks) and I received 46 responses. After 14 days from having sent the invitations, I closed my survey by deactivating the link. The Survey Monkey tool offers visual representation of the questions asked, displaying the number of responses per section. As such, I used that visualization to acquire demographic information of my respondents and laid out a basic understanding of the types of responses I received. Then I exported a CSV file of the survey responses in order to analyze the data using python libraries for data analysis. I have explained the details of the data analysis process in the next section.

5.2.3 Participant Demographics

The survey responses came from a diverse group of participants. 46 students attempted the survey out of which I dropped six sets of responses as they did not complete the survey. As per the ethics application, participants were free to not complete the survey, and in that case, the data that was gathered from them has to be deleted. Finally, I was left with 40 respondents whose responses I used for my study. I present detailed demographics of the participants of the survey in Table 5.1.

Table 5.1: Demographic information of the survey respondents

CATEGORY	RESPONSE	UG	Grad	Total
Gender	Male	22	5	27
	Female	5	7	12
	Other	1		1
Work Experience	Co-op	11	1	12
	Full-Time	3	4	7
	Both	10	6	16
	Never Worked	4	1	5
Account Created on Developer Social Platforms	Stack Overflow	13	8	21
	GitHub	28	12	40
	GitLab	24	5	29
	Google Code	3	1	4
	Bitbucket	4	4	8

In the upcoming sections of this chapter, I use the information presented in Table 5.1 as I discuss the results of they survey.

5.3 Analysis of the Survey

In this section, I describe the techniques I used to capture the intricacies of participant responses and synthesize them into cohesive findings that accurately reflect the data.

5.3.1 Data Processing and Exploration

In order to conduct the quantitative analysis, I started with exporting the survey data as a CSV [32] file. Then I loaded the CSV file on a computational notebook [41] in the Python programming environment, and imported all of the python packages I wanted to utilize for analyzing and visualizing the data. I took the following steps to analyze the data on the computational notebook:

1. **Data Extraction:** I began with extracting the data from the Survey Monkey platform, loading it on the python computational notebook, and transforming it in order to draw insights. This is known as ETL, which stands for Extract, Transfer and Load. It is an important step for pre-processing the data before its analysis.

   Python was used for, among other things, discarding data of participants who did not complete the survey, and normalizing the values of columns when I quantified the responses.

2. **Creating environment for data analysis:** I imported various packages [29] that facilitate the data analysis process using the python language. The initial CSV file that I obtained from the survey was comprised of 40 rows of data with 103 data points. I discarded the columns that would not contribute to any results, such as

   ```
   Respondent ID
   Collector ID
   Start Date
   End Date
   IP Address
   Email Address
   First Name
   Last Name
   ```

```
Custom Data 1
```

3. **Data segregation:** The responses in my survey were a mix of open-ended and closed-ended questions, I created a table that contained all of the demographic information of the participants such as the gender of the respondents, the degree each respondent is enrolled in, how many of them have an account on different developer social platforms, and whether they have used social coding platforms for GitHub and Stack Overflow.

   I also created separate tables for participant information corresponding to usage of the relevant social coding platforms during school, during work, and during their personal use.

4. **Dealing with sampling bias:** As shown in Table 5.1, the majority of participants of the survey are male undergraduate students (22 participants), which makes the sample size of other populations small. To overcome this problem, I normalize the absolute values and present my results using percentages.

5.3.2 Type of Responses

The survey contained three type of questions that required different data analysis techniques. The types of responses that could be quantified are shown below.

- **Select-one response:** For a given question, the participant could select only one response which was the most applicable to them. For example, for the question *How many private repositories containing your personal projects do you own on GitHub?*, the options for this question were: 0 , 1-5, 6-10, >10; Here the participant can have only one response. Such responses helped in understanding the demographic distribution of respondents on a given question.

The type of responses that were analysed using qualitative data analysis techniques are shown below.

- **Select all that apply type response:** For some questions, the respondents could select multiple responses. For example, as regards the questions that would ask participants the reasons for not contributing to a project hosted on GitHub, there can be multiple reasons for why a participant is not participating. hence, the responses for such questions were multiple select type.

- **Open-ended response:** For some questions in the survey, I provided the participants with open text fields where they could type their answers out. Such responses helped me gather additional information that was not directly covered in the response options I provided. I analysed such responses with qualitative data analysis techniques.

In the upcoming sections, I use this classification of the survey questions to interpret the data I gathered from the survey and answer my research questions. It is noteworthy that the survey is a pilot study for future research work, and the results I obtained were based on a small number of participants. The data trends might change if the same study is conducted with a larger participant group.

5.3.3 Limitation of Data Gathered

Since the data I gathered in the survey was tailored towards capturing student experiences, the data did not translate well into a framework for statistical data analysis. The reasons for the lack of statistically significant data analysis are as follows:

- Since the samples sizes for different demographics was not balanced, statistical tests would not produce accurate results. For example, in Table 5.1, I shared that five participants have never worked in software development industry before as compared to 35 participants who have worked in the industry before. Comparing the data gathered from 5 participants with the data gathered from 35 will require using statistical techniques such as sampling 5 participants out of 35 and then making comparisons, or comparing their mean values which can introduce sampling bias [24].

- The nature of the data gathered is highly qualitative in nature. If I sample a smaller number of participants from an unequal population, I might miss out on unique information related to participation from the participants who were left out.

- Many questions in they survey were *select-all that apply* style, which allowed participants to select multiple responses for a given question. Also there were a few open-ended questions where the participants typed their responses. Such information is hard to quantify.

- Lastly, when it comes to identifying human perception, it is possible that a given person can experience barriers and benefits simultaneously or the same person can experience multiple barriers or benefits at the same time. Hence quantifying such information could fail to produce accurate information as the same population will be sampled for different kind of barriers or benefits at the same time.

Since it was a pilot survey which was a follow-up of a qualitative study, the design of the survey questions was not suitable for the gathering of data that could be statistically tested accurately. Hence I analyzed data using qualitative data analysis techniques to accurately capture student experiences. I discuss more about it in Chapter 7 where I discuss the limitations and the threats to validity of my study.

5.4 Data Interpretation: Platform Usage Trends

In this section, I explain how students use GitHub or if they use any other social coding platforms, and how this usage differs among students with different levels of education and work experience in the software development industry. I also explore whether there is a gender-wise difference in interactions with developer social platforms among the participants.

I begin by sharing information about GitHub along with other social coding platforms followed by sharing information about Stack Overflow. Here I present the data from the interviews in tables followed by text describing the data interpretation which will provide insights into the data presented.

In the next section of this chapter, I discuss how this interpretation of data answers the research questions posed.

5.4.1 Most Popular Social Coding Platform

Table 5.1 shows the list of developer social platforms on which the participants of the survey have an account. All of the participants of the survey indicated that they have an account on **GitHub**, which makes it the most popular social coding platform among the participants [38]. Another popular social coding platform as indicated by the participants in the survey was GitLab. 29 participants (out of which 83% were enrolled in undergraduate degrees) indicated that they have an account on GitLab

(This could also be because many programming courses at our university require students to use GitLab to submit their assignments).

5.4.2 Role of Social Coding Platforms in Job Search

Many employers today, when screening job applicants look at an applicant's social coding profile as a part of the hiring process [22]. Many times job application portals add an optional field for students to add a link to their GitHub profiles. Hence, students who are looking for jobs try to be active on these platforms to increase their chances of getting hired. They use their profiles as their work portfolio (I discuss this in detail in the next section). In this section I discuss how many of the participants share their social coding profiles on their resumes or with employers at the time of applying for jobs. I divided this information based different groups I identified in the survey.

Undergraduate Vs Graduate students

Table 5.2 shows the percentage of students who share a link to their GitHub profiles on their resume. The table divides all of the participants of the survey based on the degrees they are enrolled in.

Table 5.2: Job applications and Developer Social Platforms

QUESTION	RESPONSE	UG(27)	Grad(13)
Sharing GitHub profile on Resume	Yes	74%	84%
	No	26%	16%
Job application requires Github link	Yes	67%	46%
	No	33%	54%

The results of the survey revealed the following information:

- More than 70% of the undergraduate as well as graduate students who took the survey share a link to their GitHub profiles on their resume.
- However, I observed in my survey that when applying for jobs, undergraduate respondents were asked to provide said link on their resume more often in job applications as compared to graduate students.

Gender-wise distribution

Table 5.3 shows the gender-wise distribution of the respondents who share the link of their social coding platform profiles on their resume and whether they have been asked to share their social coding platform profile with the potential employers.

Table 5.3: Job applications and Developer Social Platforms

QUESTION	RESPONSE	Male(27)	Female(12)	Other(1)
Sharing GitHub profile on Resume	Yes	81%	75%	100%
	No	19%	25%	0
Job application requires Github link	Yes	67%	50%	0
	No	33%	50%	100%

It can be observed that:

- 78% percent of the participants irrespective of their gender identity share their social coding profiles on their resume.

Work experience in the software development industry

Table 5.4 divides the participants of the survey based on whether they have worked in the software development industry before or not. It also compares how sharing of social coding profiles on resume and with potential employers differ based on previous work experience.

Table 5.4: Job applications and Developer Social Platforms

QUESTION	RESPONSE	Worked(35)	Never Worked(5)
Sharing GitHub profile on Resume	Yes	83%	60%
	No	17%	40%
Job application requires Github link	Yes	66%	0%
	No	34%	100%

5.4.3 Using social coding platforms for development in jobs and schooling

Integrating the use of social coding platforms in regular curriculum is a common practice in curriculum related to software development [19]. It is more common in second and third year undergraduate courses as compared to graduate level courses as the

earlier courses are focused on teaching the software development process and version source control; furthermore, collaborative development and source code sharing are integral parts of the process.

Also, in jobs related to software development, the process of generating the code takes place using social coding platforms. Understanding the background of the participants and their interaction with social coding platform during their schooling or working life is important to understand whether the barriers to participation in social coding platforms are not due to unfamiliarity with the platforms.

In this section, I present information on how students interact with social coding platforms during their school work and their jobs.

Undergraduate vs Graduate Students

Table 5.5 contains information about the participants enrolled in undergraduate and graduate degrees, and how they have used social coding platforms during their course work or during their jobs.

Table 5.5: Social Coding Platform usage in Job and School

QUESTION	RESPONSE	UG(27)	Grad(13)
Social coding platform used in companies	GitHub	37%	46%
	Other	26%	46%
	Did not use any	17%	8%
Social coding platform in in courses	GitHub	74%	100%
	GitLab	22%	0
	No	4%	0

Based on the responses we received, we saw a trend that:

- Survey respondents enrolled in graduate degrees reported using social coding platforms more than the undergraduate students during their jobs.
- The data also shows that when it comes to courses, graduate students use GitHub more than undergraduate students.
- The survey respondents who were enrolled in undergraduate degrees stated that they use GitLab in their courses, this may be because the university curriculum requires students to use GitLab to submit assignments related to programming.

5.4.4 Personal interactions of students with social coding platforms

For students who are enrolled in software engineering and computer science degrees, a big part of their curriculum involves interaction with social coding platforms [58]. They also have to use social coding platforms in the software development industry. For most of these activities, they have external motivations to use these platforms such as good grades at school or monetary benefits at work [23]. Real barriers and benefits come into the picture when students have to invest their time into using social coding platforms and build projects that are not giving them anything in return. Hence in order to find barriers and benefits, it is important to understand personal interactions that the students have with these social coding platforms.

I discuss how different participant groups interact with the social coding platforms when they are developing their personal projects.

Undergraduate vs Graduate Students

Table 5.6 shows how development of personal projects on social coding platforms and collaboration varies for students enrolled in undergraduate and graduate degrees.

Table 5.6: Personal interactions with social coding platforms

QUESTION	RESPONSE	UG(27)	Grad(13)
Number of personal projects on social coding platforms	GitHub	85%	76%
	Multiple platforms	0	15%
	None	15%	1%
Public repos on GitHub	Yes	81%	77%
	No	19%	23%
Public repos on other platforms	Yes	11%	16%
	No	89%	84%
Contribute to repos of strangers?	Yes	30%	61%
	No	70%	39%
Sought collaboration on personal projects?	Yes	11%	37%
	No	67%	38%
	Only known people	22%	25%

I also observed in the data from the surveys that:

- More than 75% of undergraduate students as well as graduate students who took the survey stated that they have personal projects on social coding platforms.

However, graduate students have hosted more of their projects on the public repositories as compared to the undergraduate students.

- None of the respondents of the survey enrolled in undergraduate degrees shared their personal projects on multiple social coding platforms (when it came to sharing their personal projects, GitHub was the only platform they used).
- Out of all of the respondents of the survey, graduate students have contributed more to strangers' repositories as compared to the undergraduate degree students. Also, graduate students have been more open to seeking collaboration from strangers on their work as compared to the undergraduate students.
- The respondents enrolled in graduate degrees are more comfortable interacting with strangers as compared to undergraduate students.

Gender-wise distribution

Table 5.7 shows the gender-wise difference of interaction of users on social coding platforms.

Table 5.7: Personal interactions with social coding platforms

QUESTION	RESPONSE	Male(27)	Female(13)	Other(1)
Personal projects on social coding platforms	GitHub	81%	75%	100%
	Multiple platforms	4%	0	0
	None	15%	25%	0%
Public repos on GitHub	Yes	89%	58%	100%
	No	11%	42%	0
Public repos on other platforms	Yes	11%	8%	100%
	No	89%	92%	0%
Contribute to repos of strangers?	Yes	48%	17%	100%
	No	52%	83%	0
Sought collaboration on personal projects?	Yes	26%	8%	0
	No	37%	84%	100%
	Only known people	37%	8%	0

It can be observed that:

- More than 75% of the participants of the study irrespective of their gender host their personal projects on social coding platforms.

- While other genders did not have any personal projects on the platforms, a few male participants did host their projects on other platforms.
- Even though female participants indicated that they have hosted their projects on GitHub, about 40% of them prefer keeping their projects private.
- Participants irrespective of their gender do not host their projects on other social coding platforms as much as they do on GitHub.
- While the participants identifying as male and other genders are more comfortable contributing to the work of people outside their known circle, female participants indicated that they are not comfortable interacting with strangers.
- Male participants are more comfortable seeking collaboration on their work outside their known circle as compared to other genders.

Work experience in software development industry

Table 5.8 shows how personal interaction on social coding platforms differs for the participants who have worked previously in the software development industry and the participants who have not worked before in the software development industry.

Table 5.8: Personal interactions with social coding platforms

QUESTION	RESPONSE	Worked(35)	Never Worked(5)
Personal projects on social coding platforms	GitHub	80%	60%
	Multiple platforms	6%	0%
	None	14%	40%
Public repos on GitHub	Yes	83%	60%
	No	17%	40%
Public repos on other platforms	Yes	14%	0%
	No	86%	100%
Contribute to repos of strangers?	Yes	46%	20%
	No	54%	80%
Sought collaboration on personal projects?	Yes	37%	20%
	No	16%	60%
	Only known people	47%	20%

It can be observed that:

- Respondents who have worked in the software development industry are more likely to develop personal projects on social coding platforms than those who haven't worked in said industry.
- Respondents who have previous relevant work experience tend to own more public facing projects as compared to students who have not worked in the software development industry before.
- In the same way, respondents who have worked before are more collaborative in nature as compared to those who have not worked before.

5.4.5 Student Interaction with Stack Overflow

Interactions on Stack Overflow are different from the interactions on social coding platforms like GitHub. Students heavily access Stack Overflow to get answers to their programming problems. Interactions on Stack Overflow vary from just accessing the content on the platform to making contributions such as up voting or down voting answers, posting questions, and responding to other people's questions.

I observed in related studies, and in my study, that very few students make contributions on Stack Overflow as compared to social coding platforms. In this section, I present the data I gathered around student interactions on Stack Overflow and compare how these interactions vary for different population among the students who took the survey.

Undergraduate versus Graduate Students

Table 5.9 shows the contributions made by undergraduate and graduate students on Stack Overflow. It shows whether any of the students made contributions in terms of asking or answering questions on Stack Overflow.

The observation that can be drawn from the table is:

- Overall, participants who have undergraduate degrees are more active on Stack Overflow as compared to graduate students, but the graduate students who are active contribute more as compared to undergraduate students on Stack Overflow.
- Less than 50% of the respondents of the survey post few questions on Stack Overflow and answer even fewer number of questions.

Table 5.9: Students and their participation on Stack Overflow

QUESTION	RESPONSES	UG(27)	Grad(13)
Contributed to to any activity on SO	Often	4%	16%
	Few times(2-10)	44%	23%
	Once	4%	0
	Never	48%	61%
Post a question on SO	Yes	41%	38%
	No	59%	62%
Answered a question on SO	Yes	33%	23%
	No	67%	77%

- Some of the undergraduate students indicated that they made a contribution once, whereas none of the graduate student indicated that they made a contribution once.

Table 5.10 shows how undergraduate and graduate students access Stack Overflow, and whether their interactions are affected by badges and reputation awarded within the Stack Overflow platform as an incentive. The table also shows if Stack Overflow is just another search result for them or not, by stating whether the students directly go to Stack Overflow or not.

Table 5.10: Students and their use of Stack Overflow

QUESTION	RESPONSES	UG(27)	Grad(13)
When you access SO are you always logged into your account?	Yes	11%	39%
	No	30%	31%
	Maybe	37%	15%
	I access SO but I do not bother logging in	22%	15%
Directly go to SO for your coding problem	Yes	11%	31%
	I type my problems on search engine and go to SO from there	89%	69%
Collect badges or build reputation on SO	Yes	4%	8%
	It is not my goal but since I am active there I have badges and reputation	11%	31%
	No	85%	61%

It can be observed that:

- Graduate students are more likely to be logged into their account when accessing content on Stack Overflow as compared to undergraduate students.

- While more than 50% of graduate and undergraduate students do not go to Stack Overflow directly with their programming problems, graduate students are more likely to go directly to Stack Overflow as compared to undergraduate students.

- Graduate students who took the survey held a higher reputation than the undergraduate students who took the survey on Stack Overflow, but it is mostly because they are further along in their careers and participate more as compared to undergraduate students on these platforms and happen to have reputation. It is not their motive.

Gender-wise Distribution

Table 5.11 shows the activities different genders perform on Stack Overflow.

Table 5.11: Stack Overflow participation based on gender

QUESTION	RESPONSES	M(27)	F(12)	Other(1)
Contributed to to any activity on SO	Often	7%	8%	0
	Few times(2-10)	48%	17%	0
	Once	0	0	100%
	Never	44%	75%	0
Post a question on SO	Yes	44%	33%	0
	No	66%	67%	100%
Answered a question on SO	Yes	37%	8%	100%
	No	63%	92%	0

It can be observed that:

- None of the gender groups contribute to Stack Overflow often, but the male participants evince a higher usage rate under the category of those who have used the platform a few times.

- Overall participation among male participants is higher than other genders on Stack Overflow.

- Males have indicated that they are more comfortable asking questions as compared to participants belonging to other genders.

- Males tend to answer more questions as compared to other genders.

Table 5.12 shows more details about participation of students on Stack Overflow and how it varies according to gender.

Table 5.12: Use of Stack Overflow based on level of education

QUESTION	RESPONSES	Male(27)	F(12)	Other(1)
When you access SO are you always logged into your account?	Yes	19%	25%	0
	No	19%	50%	100%
	Maybe	41%	8%	0
	I access SO but I do not bother logging in	21%	17%	0
Directly go to SO for your coding problem	Yes	11%	25%	0
	I type my problems on search engine and go to SO from there	89%	75%	100%
Collect badges or build reputation on SO	Yes	7%	0	0
	It is not my goal but since I am active there I have badges and reputation	78%	25%	0
	No	15%	75%	100%

It observed that:

- As shown Table 5.12, even though overall female participation is lower than the male participation on Stack Overflow, if female participants have an account on Stack Overflow, they tend to be logged in every time when they use the platform as compared to males. A majority of the participants irrespective of their gender do not directly go to Stack Overflow.
- Male participants have more badges as compared to other participants on the platform.
- Male participants do not care much about being aware whether they are logged into their social coding account or not when they access the respective sites.

Work Experience Distribution

Table 5.13 shows the differences in the interaction of students with work experience and without work experience with Stack Overflow.

It can be observed that:

Table 5.13: Students and their participation on Stack Overflow

QUESTION	RESPONSES	W(35)	NW(5)
Contributed to to any activity on SO	Often	10%	0
	Few times(2-10)	37%	40%
	Once	1%	0
	Never	52%	60%
Post a question on SO	Yes	40%	20%
	No	60%	80%
Answered a question on SO	Yes	31%	20%
	No	69%	80%

- Overall, people with more work experience are more active on Stack Overflow as compared to people who have not worked before.

Table 5.14: Student and their use of Stack Overflow

QUESTION	RESPONSES	W(35)	NW(5)
When you access SO are you always logged into your account?	Yes	31%	0
	No	26%	20%
	Maybe	26%	40%
	I access SO but I do not bother logging in	17%	40%
Directly go to SO for your coding problem	Yes	15%	0
	I type my problems on search engine and go to SO from there	85%	100%
Collect badges or build reputation on SO	Yes	6%	0
	It is not my goal but since I am active there I have badges and reputation	20%	20%
	No	74%	80%

5.5 Data Interpretation: Barriers and Benefits

In this section, I analyze the data associated with the barriers and benefits that the students experience when participating on the developer social platforms. The responses from the survey that revealed the information associated with barriers and benefits were *'select-all that apply'* and *'open-ended'* in nature. Since a single

participant could select multiple options as a response to the question being asked or could write a response, I analyzed the data using qualitative data analysis techniques on the data I gathered (similar to the semi-structured interview data gathering that I discussed in Section 4.4). The steps are as follows:

1. I began with exporting the responses to the questions that were aimed at identifying barriers and benefits in an MS excel[3] sheet.

2. I conducted closed-card sorting [45] on the survey data where I categorised the survey responses into existing codes from the semi-structured interviews (I have explained the coding process in detail in Section 4.4.2 and the clustering process in detail in Section 4.4.3). I did so by placing the barriers and benefits along with their codes on a fresh coggle.it document.

3. Then I assigned the responses to the existing codes while looking for the responses that do not fall under any existing code. If that was the case then I assigned a new code to those responses (which did not happen in this study).

4. I assigned frequency to each of the codes by counting the number of times the responses associated with the codes were selected by the respondents.

5. Since I reused the codes from the semi-structured interviews, the codes were organized into barriers and benefits along with the assigned frequencies.

6. By adding the frequency assigned to the codes, I assigned ranks to the barriers and benefits based on the sum obtained by adding the frequencies (higher the value of sum, higher the rank). Then I arranged the barriers and benefits based on ranking from most prominent to least prominent.

7. Finally I counted the number of participants who experienced a given barrier or benefit.

5.5.1 Responses for Social-coding Platforms

Table 5.15 presents the data associated with motivations and road blocks that students are subject to when they start participating on social coding platforms.

[3]https://www.microsoft.com/en-ww/microsoft-365/excel

Table 5.15: Students experience when sharing personal projects on developer social platforms

QUESTION	RESPONSES	Total
Motivation to upload your projects **related to school or job** on social coding platforms	I use these platforms to keep a record of my work	28
	Any public-facing projects add to my work portfolio	26
	I upload my projects on social coding platforms to use VCS[4]	16
	Other*	1
Motives behind creating **personal projects** that are not related to your school or job	To enhance my skills and learn related new languages	27
	To add to my resume	22
	Programming is my hobby	21
	I only make projects for my career	8
	Other**	5
Motives behind publicly sharing your personal projects on social coding platforms such as GitHub	Any public-facing project adds to my work portfolio	27
	To gain reputation in the community	19
	I like keeping my code open to use	17
	I like developing projects and sharing good ones publicly	14
	I keep my personal work public so that people can collaborate	13
	Other***	7
Reasons for hesitation to publicly share projects on social coding platforms	I have ongoing projects in private that I will share when they are ready	19
	My projects are not good enough to be shared publicly	14
	Maintaining public projects is big work	9
	No time for public projects	7
	There are no benefits to sharing public facing projects	4
	Other****	5

Similarly, Table 5.16 shows the barriers and benefits students experience when they seek contribution from or contribute to other people's work.

After gathering the responses from the participants along with the corresponding counts, I used a card sorting technique which is same as the technique I used on data from my semi-structured interviews in order to present the quantified responses obtained from the survey. For social coding platforms such as GitHub, there were 61 responses in total that included 33 close-ended responses and 28 open ended responses that were categorized into barriers and benefits that students experienced. Similarly,

QUESTION	RESPONSES	Total
Sought collaboration on projects you publish on social coding platforms	No, I never thought about finding collaborators on my projects	10
	Yes, but only within my known circle not strangers	9
	Yes, I link my project repository on various forums to find collaborators	8
	No, finding collaborators is too much work	5
	Why would anyone collaborate with me? There are other better projects	2
	Other*	4
Contribute to repos of users that you do not know (strangers)	Yes, I contribute if I come across something interesting	12
	No, It is very time consuming	11
	No, my skills are not sufficient	7
	No, I am not interested in contributing to stranger's code	6
	Yes, if it does not take much of my time	4
What stops you from collaborating with people outside your known circle	When I gain more skills and experience, I will participate	16
	Collaboration is a lot of work, if I have time and it is easy, I will contribute	12
	I am not comfortable collaborating with people I do not know	7
	Collaboration will not benefit me in any way	2
	Other**	5

Table 5.16: Barriers and benefits experienced by our participants by collaboration on developer social platforms

as regards Stack Overflow, there were a total of 26 responses that included 16 close-ended responses and 10 open-ended responses that were categorized into barriers and benefits.

This time, I reused the codes that I created while working on the semi-structured interviews data in order to categorize the responses we gathered in the survey. During the coding process I was looking for the emergence of new codes, but none emerged.

5.6 Discussion of the Survey Results

In this section, I discuss the results I obtained from the survey I conducted as a part of my study. I answer the research questions based on the data interpretation I conducted in Section 5.4 and 5.5. The interpretation presented in section 5.4 answers research question 1, and the interpretation conducted in Section 5.5 answers research

questions 2 and 3.

5.6.1 RQ1. How do students use GitHub and Stack Overflow?

Based on the data gathered from 40 students: all students use various developer social platforms at different stages of their degree.

As shown in Table 5.1, all of the students have an account on at least one social coding platform and all of the students have an account on GitHub. When it comes to Stack Overflow, around half of the respondents have an account on Stack Overflow. The degree and the intent of interaction differs for both of the platforms. For instance, students often use social coding platforms to develop projects, not just to use the community aspect of it, but to use the features and functionality provided by the platform such as keeping the track of the code changes, and storing different versions of the projects. In the same way, students also use platforms such as Stack Overflow to find answers to their problems. The cumulative data presented in Table 5.5 shows that more than 95% of the students have used social coding platforms during their courses, and a majority of students who have worked in the industry before have used social coding platforms during their job.

Different demographics of participants interact with platforms differently. Below, I present the results from Section 5.4 on how different demographics interact with the platforms differently.

Does one's level of education have an effect on the amount or degree of interaction with developer social platforms?

1. Based on the responses I obtained, I speculate that undergraduate students are more likely to be asked for their social coding profile link, as compared to graduate students, by potential employers. This is likely because graduate students are seen as more experienced than undergraduate students.

2. Figure 5.3 shows that 30 out of 40 students have used social coding platforms at school, 28 indicated that it was mandatory to use a platform and 11 students said they used it anyway because it is easy for them to develop projects on a social coding platform.

   When it comes to the school or work environment, respondents enrolled in

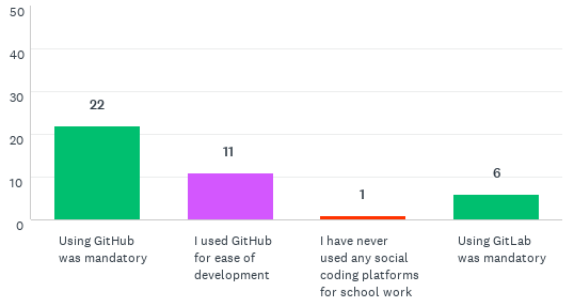


Figure 5.3: Using social coding platforms at school

graduate degrees used these platforms more than the respondents enrolled in undergraduate degrees.

3. When it comes to personal development activities, I inferred that graduate participants are more likely to share their work publicly and to collaborate outside of their known circle as compared to undergraduate participants.

4. The results of the survey also showed that when it comes to Stack Overflow, even though undergraduate students participate more on the platform, the graduate students who do participate are more dedicated and they have a better reputation than undergraduate students.

Based on the information shared above, I observed that graduate students are more aware of their interests, hence their participation is more frequent and active than undergraduate students. Also, graduate students are slightly more likely to collaborate on developer social platforms than undergraduate students.

Are there any noteworthy differences across different genders when it comes to interactions with developer social platforms?

1. The survey results indicated that females tend to keep their projects in private repositories more than other genders.

2. Males are more likely to collaborate with people outside their known circle than other genders. However, as per the survey, the number of male respondents who collaborate was not significantly larger than other genders.

3. Similarly, male participants reported to being more comfortable asking questions and answering questions on Stack Overflow compared to other genders.

I observed that male participants are more open to sharing their work with strangers which shows that they are more comfortable with their skills than other genders.

Do students who have worked before in the software development industry use social coding platforms differently than students who have never worked?

1. Students who have worked in the software development industry before share the link to their social coding profile on their resume more than those who have not worked.

2. Students who have worked before are more open to sharing their code online and collaborating outside of their known circle as compared to people who have not worked.

3. The same trend can be seen as regards participation on Stack Overflow.

5.6.2 RQ2: Barriers Students Experience with GitHub and Stack Overflow

Here I discuss the quantification of the barriers to participation I identified from the analysis of the responses I gathered from 40 participants who completed the survey. I also present the top two most prominent codes for each of these barriers along with their frequency among the 40 participants of the survey.

All of the barriers and the benefits I present below are ranked from most prominent to least prominent based on the number of participants experiencing them. I also share the top two codes associated with each barrier.

Barriers to participation in GitHub

The quantified barriers that students experienced on social coding platforms like GitHub are presented below:

1. **Lack of self-confidence**

This is a prominent barrier that students at all stages of their education face. The top codes associated with this barrier are shown below.

Popular codes	**Participant Count**
I am inexperienced	23
My projects are not good enough	19

With **33** participants experiencing this barrier, this barrier was also faced by more than 50% of the participants of the previous study (semi-structured interviews) I conducted as a part of this research.

2. **Desire to remain in comfort-zone**

   This is the second most popular barrier that I identified in the survey which was also seen among the participants of the semi-structured interviews. **27** participants pointed to the existence of this barrier.

Popular codes	**Participant Count**
I am more comfortable collaborating within my known circle	17
I never sought collaboration online	12

   This barrier is not limited to social coding platforms like GitHub as it also applies to forums like Stack Overflow.

3. **Not willing to make an effort**

   23 participants indicated that they are not open to applying effort to make contributions on social coding platforms.

Popular codes	**Participant Count**
Seeking collaboration / contributing to other people's code is too much work	26

   In order to be active on social coding platforms it is important to make some effort to share one's work publicly and, at the same time, seek collaboration from people around you to give feedback.

4. **Time related barriers**

Lack of time is another important barrier that was reported multiple times in the responses I received.

Popular codes	Participant Count
I do not have time at the moment to collaborate	22
My school work keeps me occupied	3

When someone says that they do not have time for a certain activity, it means that the particular activity is very low on their priority list. **21** participants who took the survey faced this barrier.

5. **Lack of interest in participation**

   12 participants indicated that they are not interested in making contributions to social coding platforms such as GitHub.

Popular codes	Participant Count
Open participation is not my hobby	10
I prefer to work on my own projects over collaborative projects	6

6. **Expecting immediate benefits from participation**

   This barrier was common among the graduate students, and the instances associated with this barrier were selected 9 times. A total of **7** participants from our study have faced this barrier.

Popular codes	Participant Count
I do not see any benefit to participation in open source	6
I am only interested in participating when my work gets attention	3

   This barrier is not common among undergraduate students, but I found this barrier among students in the semi-structured interview as well.

Barriers in Stack Overflow

Stack Overflow is a popular platform where developers seek help with their programming problems. All of the participants of the survey, as well as the semi-structured interviews, have indicated that they do access content on Stack Overflow. The barriers that students face in their participation on Stack Overflow are presented below.

1. **Stack Overflow is not an obvious choice for seeking help**

   A majority of the participants who took the survey indicated that if they encounter a coding problem, they type their problem on search engines and only visit Stack Overflow if it appears in their search results.

Popular codes	**Participant Count**
Stack Overflow is just another search engine result, I never go there directly	30
I prefer to use other resources over Stack Overflow	5

   27 participants have indicated that the Stack Overflow platform is not their choice for seeking help.

2. **Looking for quick help**

   This is another popular barrier that exists in the Stack Overflow community as it was selected 18 times by the participants. They do not feel like sharing their questions on Stack Overflow as they feel that there is an onerous delay before they get their questions answered.

Popular codes	**Participant Count**
Long waiting time to get answers on SO	11
Answering questions is a big task	7

   14 participants pointed to the existence of this barrier, hence they avoid sharing questions on the platform.

3. **Lack of self-confidence**

Lack of self-confidence is the second most popular barrier selected by participants as regards the Stack Overflow platform.

Popular codes	Participant Count
I am inexperienced	21
My contribution is not that special	5

People in the programming field tend to underestimate their own skills which is evident from their interactions with question answer forums like Stack Overflow [40]. **18** participants indicated that this was a barrier for them.

4. **Contribution on Stack Overflow is too much work**

The codes associated with this barrier were selected 22 times by the participants of the survey. This barrier was also discussed by the participants of the semi-structured interviews.

Popular codes	Participant Count
Asking questions in a particular format is too much	15
Answering questions is a big task	7

13 participants have expressed the presence of this barrier based on their survey responses.

5. **Desire to remain in comfort-zone**

The codes associated with this barrier were selected 13 times by the participants. Students indicated that they feel uncomfortable interacting with people they do not know and ask them their programming problems. Some students see the act of asking questions a way of manifesting insecurity which many of the participants were not ready to do with strangers.

Popular codes	Participant Count
I am comfortable asking questions only from people I know	13

However, only **9** participants of the survey displayed this barrier.

6. **Unwilling to help others**

The codes associated with this barrier were selected 12 times by the participants. Students indicated that mostly they do not have an intention to contribute to the platform even if they know the answer to a question because they do not care enough.

Popular codes	**Participant Count**
I don't care about helping others by knowledge sharing	12

The presence of this barrier was seen in **6** participants who took the survey but this barrier was also highlighted by the participants of the semi-structured interviews.

5.6.3 RQ3. Benefits to Participation

In this section, I talk about the benefits that the students experience when they participate on developer social platforms.

Benefits to participation in GitHub

While a lot of the participants of the survey pointed to experiencing barriers, for those who have made active contributions, they experienced benefits too. By highlighting the benefits that students experience while using social coding platforms, like GitHub, I hope that students will be encouraged to participate and enjoy some of the benefits.

1. **Interest and skill-development**

   30 people have claimed that they gained skill and experience through using social coding platforms frequently. They got a chance to get feedback from the community as well as an opportunity to see the code of other developers. This helped them develop a further interest in developing their skills.

Popular codes	Participant Count
I find collaborative development interesting	33
Skill development	27
Programming is my hobby	25

2. **Increased confidence from displaying skills online**

   27 participants of the survey said that they gained recognition on GitHub which made them feel good about their programming skills. Seeing their repositories gaining starts and being cloned by other users of GitHub gave them a sense of validation on the quality of projects they developed. Many participants of the survey have experienced this benefit at some point with their interaction with social coding platforms.

Popular codes	Participant Count
I use my social coding(GitHub) profile as my portfolio	40
I like showing my work online	21

   A majority of the participants have indicated that they use their GitHub profiles as their work portfolios and further that they use their profile content to show their skills to potential employers.

3. **Altruism**

   15 participants who took the survey indicated that they like to help people by making their own code accessible or by contributing to the open source projects so that the project becomes useful for everyone. This attitude of doing something that can benefit other people is a big source of motivation for them to participate and contribute on open communities.

Popular codes	Participant Count
I like it when others find my code useful	14
I report bugs to make products better for everyone	1

4. **Learning from the community**

   2 participants indicated that their participation in open communities has helped them learn from the community. The communities that are formed on the social coding platforms are comprised of learned people who have strong coding experience as well as experience working in accordance with strong industrial practices. By collaborating with these individuals, students got a chance to enhance their skills and improve their coding style by acquiring inspiration and knowledge from such experienced developers.

Popular codes	**Participant Count**
Learning from the community	2

Benefits to participation in Stack Overflow

Here, I will describe the benefits that students experience using platforms like Stack Overflow.

1. **Community reliance**

   12 participants have indicated that the Stack Overflow community is very reliable as it makes them feel that they are not alone in their problems.

Popular codes	**Participant Count**
There is always support available	12

2. **Altruism**

   11 respondents who have participated on the Stack Overflow platform have done it for altruistic reasons. They feel that the responses on the platform have helped them a lot with their work, and that is only possible because some individual decided to make an effort to share an answer. By doing the same, they get an opportunity to pay back to the Stack Overflow community, and help someone else who might get stuck in a problem that they themselves once faced.

Popular codes	**Participant Count**
I like helping fellow coders	7
Paying back to the community	6

3. **Increased confidence from displaying skills online**

   11 participants of the survey have indicated that when they see their code being used by people on Stack Overflow, they feel good about their skills. Also, they like to build reputation on Stack Overflow as it feels to them like a measure of their knowledge of coding. Knowing that many experienced people use the platform, having good reputation makes them feel as capable as other people on the platform.

Popular codes	**Participant Count**
I like to gain Stack Overflow reputation	7
I feel more confident when people use my code	6

4. **Sense of involvement**

   5 participants indicated that being an active member of the Stack Overflow community, by answering asking questions, gives them a sense of belonging to the community, which helps motivate them to be more active in the community too.

Popular codes	**Participant Count**
Asking and answering questions make me feel more involved in the community	5

5.7 Summary

In this chapter, I discussed the design, distribution and analysis of the survey study which helped in extending the results obtained from the semi-structured interviews I discussed in chapter 4. The survey provided an in-depth information about student participation on GitHub and Stack Overflow, I also found that GitHub is widely used by the students out of all the social coding platforms available.

In the next chapter, I combine the results from the survey as well as the semi-structured interviews to discuss the overall findings of my study. I also discuss how the barriers I identified as a part of this study resonate with the existing work on the barriers to participation in the literature and how the findings of my study be used

to create recommendations for different stakeholders who can help increase student participation.

Chapter 6

Discussion

In this chapter, I discuss the implications of the findings collectively from semi-structured interviews and the surveys. I also discuss how the barriers I identified in the study add to the existing work in the area of participation on developer social platforms. I also provide recommendations for various stakeholders who can benefit from the findings of this study.

6.1 Synthesizing the Interview and Survey Results

The 20 students who participated in the semi-structured interviews did not participate in the survey. Since the design of the survey was based on the findings of the semi-structured interviews, recruitment of the same participants would have introduced a bias. This bias would manifest due to the questions of the survey already aligning with the experiences of the participants of the semi-structured interviews as the survey questions were derived from their experiences. Introducing a new set of participants for the survey mitigated this bias and helped in validating the insights drawn from the interviews. In this section, I discuss the findings of my research by combining the results from semi-structured interview study and the survey.

6.1.1 RQ1: How do students participate and contribute to the content hosted on developer social platforms?

The semi-structured interviews revealed different activities that the students from different demographic platforms perform on the developer social platforms. However, due to the semi-structured nature of the interviews, it was difficult to gather

nuanced information on the activities for different demographic groups. The survey did not just help in obtaining the details of different activities performed by different demographic groups, it also helped in comparing the activities performed by different demographic groups in order to obtain more insights that I discussed in Section 5.4.

I found from my study that all of the participants were aware of GitHub and Stack Overflow and used these platforms in some capacity, either in courses or professionally. Most of our participants indicated that they have accounts on developer social platforms: almost all have a GitHub account, many have accounts on other social coding platforms such as GitLab, and over half have Stack Overflow accounts. However, not all students who have an account contribute, and not all contributions are equally substantial.

Regarding platforms such as GitHub, I still observed contributions from nearly all students as they publish projects there for classes as a matter of course, so they have at least a rudimentary understanding of how to make contributions when developing outside of schoolwork. Also all the students have an account on social coding platforms as it is integral part of their school and career. However, in case of platforms like Stack Overflow, there are fewer students who actually contribute there. Adding contributions to Stack Overflow is definitely an extra step that the student need to take in order to get their contribution in. As they do not have to participate as a part of their degree or job. They can easily excel in their career without participating in Stack Overflow, which shows that the people who are actually passionate about collaboration participate in Stack Overflow.

6.1.2 RQ2: What are the barriers that students face in their participation on GitHub and Stack Overflow?

The main focus of this study was to uncover **barriers** that students may face and to see if these barriers were a significant impediment. A prominent barrier which was reported several times in previous studies on understanding user interactions with developer social platforms is **lack of self-confidence**. This barrier emerged prominently in my study too which may indicate that low self-confidence of some users of the developer social platforms starts at an early stage (when these users are students and still learning about collaborative development) of their interaction with

platforms like GitHub and Stack Overflow. Such users can build their confidence at their student stage by understanding the benefits of contributing to the public facing content on these platforms, get motivated to contribute and gain appreciation by other users of the platforms.

I also learned that students often had low expectations about using Stack Overflow - which may indicate that more training could help students use this tool more effectively. Finally, many of the students lacked the requisite motivation to contribute, and exhibited poor awareness of the benefits they may gain.

While the data gathered from the survey did not reveal any new barriers not reported in the semi-structured interviews, it helped bring new insights in terms of assigning ranks (based on the frequency of occurrence among the participants) to the barriers I identified in the semi-structured interviews.

The major barriers observed with **social coding platforms** are:

- Lack of self-confidence among the students regarding their skills and the work they develop which stops them from participating.
- Students prefer to stay in their comfort zone as it is easy for them to interact with people who are either physically around them or the people that the students know personally. This deprives them from an opportunity to explore the opportunities beyond their known circle or meeting people who are in different part of the world but hold similar interests.
- When publishing a public facing project on platforms like GitHub, it is important to create on-boarding documentation for people who might be interested in collaborating. Many students who share their projects on these open platforms are not willing to do these extra steps of creating on-boarding documentation, or if they come across something interesting, they are not interested in putting extra efforts to read through the project's documentation to make contribution.
- Students feel overwhelmed with their school work and are not able to take time out for their hobby software development, let alone the development on the social coding platforms. This is a very prominent barrier among the students who are enrolled in undergraduate degrees.

In addition to the barriers I discussed above, there are some barriers that are not prominently presented in the existing literature on developer social platforms. These barriers were not reported as heavily as the barriers that I discussed above, but they were still reported heavily by the participants of my studies. These barriers are:

- Some students actively avoid participation on the platforms as it is not in their line of interest. The only interactions entered into are for school or work which are unavoidable for them.
- When it comes to doing any work outside their hobbies, some students expect immediate returns from the work they're doing. They feel that as compared to the returns they will get from personal participation, the amount of effort they would put in making personal contributions on the developer social platforms is too high.

The last barrier that was not reported by any of the participants of the survey, but it appeared prominently among the participants of the semi-structured interviews. Also, out of all the participants of the semi-structured interviews, this barrier was not observed among the students who were enrolled in undergraduate degrees but was sometimes seen in the senior students such as masters and PhD students in case of social coding platform participation is as follows:

- Not trusting the coding skills of people that the students do not know personally on social coding platforms, hence refraining from interacting with these people.

I speculate that these barriers when worked upon can increase the participation of students on social coding platforms.

Similarly the major barriers with **Stack Overflow** platform are as follows:

- Many students do not recognize Stack Overflow as a place to seek help which deprives them from understanding the working of the platforms.
- Mostly students are looking for instant solutions to the coding problems they encounter when they access Stack Overflow. If they do not find the response they are looking for, they do not post questions as they feel that it will take long for their question to be answered.

- Lack of confidence is not limited to social coding platform, this barrier also stops the participants from participating on Stack Overflow.
- A lot of students indicated that even when they come across a question they can contribute to, the amount of work that goes in creating the answer (providing code snippets, explaining the rationale behind the answer, and following the expected format of the website) makes them not want to make such efforts.

6.1.3 RQ3: For the students who participate, what benefits do they experience from their active participation on developer social platforms?

In terms of **benefits**, I found that students experience similar benefits to those benefits enjoyed by other developers as reported in the previous literature. However, some benefits may be of particular importance to students. As students have a greater need to rapidly learn new material as a part of their degree curriculum as compared to other developers who are past their learning stage, they can benefit from connecting with others, and the visibility of their contributions can be used to build their portfolios and boost their confidence.

In my research involving semi-structured interviews and the survey, I evaluated the interactions of students with two different kinds of developer social platforms. Github offers a space to share the project source code and collaborate, whereas Stack Overflow is a platform where one can ask and answer to the questions that one encounters while programming. Thus the nature of benefits experienced for each of these platforms is different.

However, I identified some benefits that were common to both GitHub and Stack Overflow, these are as follows:

- Developer social platforms offer a space where when people share their code publicly, they not only apply their programming skills but others can see their skills too. Hence, they feel good when their work gets recognized as this feeling resembles to gaining appreciation on an achievement. This benefit is experienced by students in case of their participation on both social coding platforms and Stack Overflow.

- Also, since using the content available on social coding platforms is very common for students, they feel that contributing to the content on the platform is their way of giving back to the community.

Benefits to participation on **GitHub** are as follows:

- Social coding platforms offer a space where people who hold interest in programming can meet other such people and they can gain skills in collaborative development along with programming.
- There are people on the community that are very skilled and who publish their work. It offers a chance for students to learn from their development style.

Benefits to participation on **Stack Overflow** are as follows:

- Stack Overflow offers a space where students can find help whenever they get stuck.
- Also, by participating on Stack Overflow, students get a sense of involvement with the content they access on these platforms which makes taking help feel good.

Even though the survey helped in quantifying the benefits that the students experience through their participation on the developer social platforms, it did not reveal any new benefit that was not reported by the semi-structured interviews. The survey helped in determining the major benefits that the students who actively participate experienced based on the frequency of occurrence of those benefits among the students.

6.2 How do the barriers I identified resonate with existing barriers in the community?

In Figure 2.1, I summarized the barriers that exist in the communities formed around developer social platforms. These barriers are based on existing work related to evaluation of the participants of the developer social platforms. As seen in Figure 2.1, some people are interested in participating but hesitate to do so. Such people face barriers like:

- Lack of confidence of an individual on their programming skills.
- Misconception about the way in which making a contribution works on the developer social platforms.
- Poor time management skills.

On the other hand, some individuals are not interested in participating. The barriers they face are:

- Not passionate about collaboration.
- Selfish motives.

In the study I conducted, I reveal more nuanced barriers that are faced by students who are a part of the community formed around the developer social platforms.

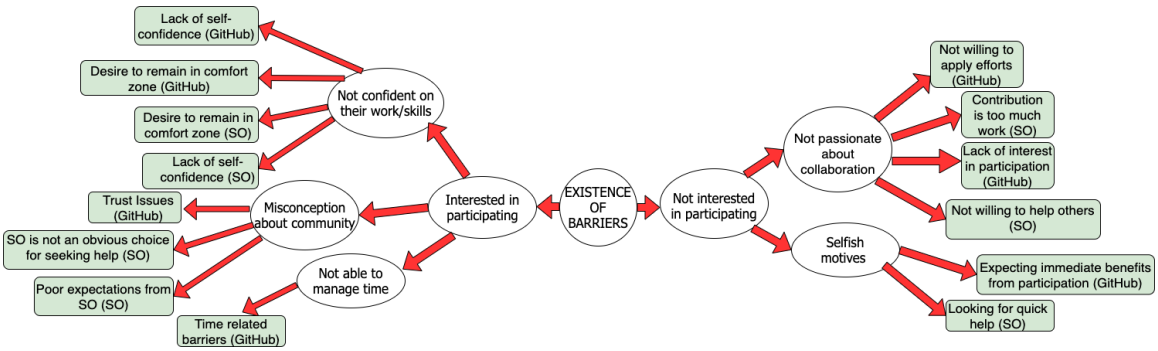


Figure 6.1: Barriers found in the literature extended to include barriers discovered in our study.

Figure 6.1 shows how the findings of my research can be extended to the existing barriers in the community. I believe that the barriers that I identified, and that fall under the 'Interested in Participating' category, can be worked upon and improved easily as they are perceived in nature. Perceptions can change easily, but in the case of other barriers such as *Not willing to apply effort* or *Contribution is too much work*, the person is actively avoiding the act of participation. In order to tackle these kind of barriers, more work needs to be done which I discuss in Chapter 8.

While conducting my research, I got an opportunity to read literature on different aspects of the developer social platforms and different group of people who have

a direct or indirect impact on these platforms. In the upcoming sections of this chapter, I discuss the steps that different stakeholders such as: students, educators and platform designers could take to help increase student participation.

6.3 Recommendations for Students

The idea of collaborative development and the existence of developer social platforms is a multifaceted concept with different actors involved. After gaining a nuanced perspective on student participation, in this section I include recommendations for students on the steps they could take to create a more welcoming environment for the students who are hesitant to participate.

Before I make recommendations for students, I discuss an important phenomenon knows as peer parity and peer influence that can play an important role in the lives of the students.

6.3.1 Using Peer Parity and peer Influence to Increase Student Participation on Developer Social Platforms

Peer parity is defined as when an individual can identify with at least one other peer when interacting in a community [20]. Peer parity can exist across and within races, genders, experiences, career positions and more.

Peer parity gives rise to **peer influence** which is a well discussed phenomenon in the field of human behavioral studies. Students, like any other people, are the members of groups, for example, the students who are taking same courses during an academic term form groups to discuss their work, or graduate students who conduct their research under the same supervisor are a part of a research group. All the groups (as distinguished from arbitrary categories) have direct or indirect influence over their members [35]. Peer group plays a large role in the social and academic development of the students as students tend to observe their peers perform certain activities and get curious (and sometimes motivated) to perform the same activities [4].

As a part of my semi-structured interview study which I presented in Chapter 4, I asked the participants a question about whether they will be influenced if they see people around them contributing to the developer social platforms. Out of 20 respondents, 17 indicated that they will be curious to know what activities their

peers are doing and they would like to give it a try too.

By presenting the results I obtained about the benefits that are experienced by the students when they participate on developer social platforms, a curiosity to participate can be created among the students who are facing barriers and refrain from participating on the developer social platforms.

6.3.2 Recommendations

I recommend students to invest time on the developer social platforms, and contribute more. I also encourage students to manage their time in a way that they are able to complete their work related to school in a reasonable time so that they get extra time to explore developer social platforms with an intention to contribute. Once they start accessing the content hosted on the developer social platforms with an intent to collaborate, over time the students might experience:

1. Stronger programming skills by contributing to other people's code as every individual has their own programming style and by contributing to other people's work, students can learn new ways of approaching problems.

2. More confidence on their programming problem solving abilities as a result of making their code available for public to read. This can help students overcome their hesitation and if their work gains recognition from other users, they will feel validated that leads to an increased sense of competence [54].

3. High activity on the platforms that may grab attention of future employers as high activity on platforms like GitHub reflect high collaborative skills which is highly appreciated in software development companies [22].

4. Stronger work portfolios.

I also recommend students to contribute to big open projects like different distributions of the Apache project[1] or Grunt.js[2]. This will help them with:

1. Understanding how to collaboratively develop projects.

2. Learning the integration of other developers' work into their own.

[1]https://github.com/apache/xx
[2]https://github.com/gruntjs/grunt

3. Gaining hands-on experience with concepts like merge conflicts, raising flags on project modules, running test cases, issue resolution which will help them during their jobs.

This recommendation is reported as very effective in the studies focused at overcoming the barriers to participation that newcomers face with their participation on developer social platforms [49].

I encourage students to ask questions and to post answers in Stack Overflow because:

1. Asking a question in the right manner is an important skill as asking the right questions helps an individual to articulate their thoughts.

2. By answering to the questions, students get an opportunity to share their knowledge with others and get a sense of clarity on their own concepts [8].

Provide feedback to the content they access on the developer social platforms as:

1. If it is an open project, the product owner can make the product better for everyone.

2. If it is an answer on Stack Overflow, the response gets validated or can be improved for everyone else to use.

6.4 Recommendations for other Stakeholders

Other than students, there are other stakeholders who could help increase student participation.

6.4.1 Recommendations for Educators

Firstly, I recommend educators to increase awareness among the students about the participation on the developer social platforms. This will help in increasing the participation by:

1. Educators play an important role in influencing the students [53]. If the students hear from the educators about benefits of participation, they will be motivated.

2. Educators can also increase awareness that other students and even more senior developers also face challenges such as poor self confidence in using social coding platforms.

Educators can design their study material such that there is more emphasis placed on making actual contribution on participating on the developer social platforms.

1. They can give such assignments to students that involve collaborating on projects hosted on developer social platforms [13].

2. They can provide grade incentives if the students respond to the questions posted on Stack Overflow.

3. Including industry collaborations as a part of the curriculum can help students gain interest in the participation in open source project development [10].

6.4.2 Recommendations for Platform Designers

I recommend platform designers to consider different groups of users when they are designing or evaluating their platforms. They can design their platforms to ensure that the projects that are being hosted give detailed installation instructions along with explaining the project dependencies. This can help new programmers understand the software development process better and they do not feel as intimidated by big projects. In addition to ensuring that the content hosted is well explained for new developers, they can also the the following steps:

1. In addition to the issues, platform designers can create a space where users can post their feedback on the platform.

2. Platform designers should moderate the platforms such that if a new user is making contributions, harsh responses can be blocked.

They can also offer greater support for learning how to use these tools and provide students with support when they get stuck.

1. Visual pedagogy such as video tutorials can be created for the students so that they can grasp the functionality of the platforms faster.

2. Creating peer-mentor programs to facilitate guided help for new comers can be used more widely to help students with their participation on the developer social platforms [43].

6.5 Summary

In this section, I summarized the findings from the semi-structured interviews and they surveys. I also discussed how the barriers I identified in my study resonate with the barriers that exist already in the community. I also provide recommendations for different stakeholders such as: Students, Educators and Platform Designers who could help increase participation on the platforms. In the upcoming Chapter, I discuss the limitations and the threat to validity of my study.

Chapter 7

Limitations

As I mentioned previously throughout the thesis, no study is perfect and each methodological choice has both benefits and drawbacks for overall research quality. In this chapter, I identify limitations and threats to validity associated with this research, why the benefits of these choices justify the risks and the measures I took to minimize potential issues and biases throughout the research design.

7.1 Threats to Construct Validity

As this is an exploratory study and relies heavily on qualitative data gathered from students there are limitations on how this study can be applied and understood in the computer science research domain.

7.1.1 Purely Technical Research

As this study was created based on semi-structured and open-ended responses gathered from students, it does not easily capture the intricacies of more technical and mathematical research. As a sociotechnical domain, related work shows that, overwhelmingly, this kind of research is conducted using statistical and mathematical modelling because of the technical nature of the research itself. My data collection methods and analysis does not capture the intricacies of statistical aspects (like calculated population sizes, correcting the numbers based on sampling differences [24]) of such studies, as it is concerned primarily with understanding human interactions with developer social platforms rather than focusing on the platforms and counting the human activities on them. Hence, one limitation of this study is that it may

not be the best choice for discussing the statistics of the student involvement in the communities formed on various developer social platforms. In that case I would recommend to study the research that is based on a mathematical model for analysis of data as the results in that study would be statistically derived.

7.1.2 Question Order Bias

Question order bias, or "order effects bias", is a type of bias where a respondent may react differently to questions based on the order in which questions appear in a survey or interview. The semi-structured interviews and the surveys that I conducted involved asking questions from the students. While semi-structured interview question progression was open and varied on the student responses, the progression for the questions I administered in the surveys were fixed. When a participant is asked similar kind of questions repeatedly, it can create a *norm* in their head and the participant can overstate a given situation. This could be an issue in the survey if all the questions related to barriers to participation were grouped together and the questions related to benefits were grouped together. In such case, a participant could overstate the presence of barrier in their participation which could affect the results.

To mitigate the 'question order bias', I grouped the questions by topic to unfold in a logical order. For example, the questions related to use of developer social platforms in school were grouped together and the questions that focused on barriers and benefits were made adaptive as shown in Figure 5.1. With this way of placement of the questions in the survey, the participants did not encounter the same kind of questions in progression and I was able to minimize this bias.

This bias can be further mitigated by randomizing the questions being asked in the survey in order to ensure that the participant is not being pressed to answer about a single topic.

7.2 Threats to Internal Validity

As many of the tasks associated with this research were conducted alone by a single researcher, there are a few threats to internal validity associated with different forms of bias.

7.2.1 Potential for Researcher Bias

One limitation of this study is the potential for researcher bias during the drafting of the questions featured in the semi-structured interviews, the responses to which were a motivation for the questions of the survey. Also, some bias may have occurred in the follow-up questions I asked during the interviews. There may be other possible biases such as the order of questions asked and the specific wording of questions asked. In anticipation of this issue, I took some measures to mitigate this possibility such as conducting several rounds of pilot interviews.

7.2.2 Perceived Barriers and Benefits

As with any study involving participants, this study involving semi-structured interviews and a survey has certain limitations associated with participant perception. Oversampling of extremes (highly positive self-image, highly negative self-image) is a common limitation of such studies. In this case, an oversampling of the participants who put extra efforts in to illustrate their activity on developer social platform without acknowledging issues they have may have helped identify some benefits; or oversampling the participants who suffer with severe "Imposter syndrome" which is common in the field of software development may have helped identify some barriers that may not have emerged had the sample included more neutral students.

Additionally, regarding surveys which have more structured sets of questions that is the questions with closed-ended responses, they can result in findings that may be biased as some students are unlikely to respond in a way that reflects poorly upon them. This is a common limitation of surveys and may have affected the responses to a few questions in particular. For example, when asked about reasons for not posting questions on Stack Overflow where the participants could select a response varying from not having enough time to participate to feeling that their question is not worth posting, they chose the response that says that they want to post a question but they do not have time, when in reality they are afraid of their question not being good enough to be posted there. Also, I speculate that some participants may have underestimated or overestimated their interactions with developer social platforms. Thus, the findings should be considered as limited to the responses of the participants *relative* and to their responses for other categories in the same question, rather than considering broader applicability giving rise to overemphasizing the individual

response itself.

7.3 Threats to External Validity

As this research focused on a particular group of the people (students) who use developer social platforms, there are a number of potential threats to external validity associated with this work.

7.3.1 Participant Selection

One limitation to this research is the selection of students who are third year and up in their undergraduate degrees along with graduate students. Choosing a different sample likely would have produced different results for both the semi-structured interviews and the survey, and below I discuss how a different research sample would have affected the findings.

Among the students enrolled in undergraduate degrees, by selecting students from third year and up, it is not possible to understand the participation of students in the early years of their degree. Had I recruited students from first and second year, the ranking of the barriers may be different as the students who are in their third year and up have already completed several courses requiring interactions with developer social platforms. By recruiting students who are in the first and second years of their degrees, it would have been possible to also explore the barriers to entry that students experience. Though, there were a number of reasons for doing recruiting the participants who were in their third year and above:

- My intent was to have participants who understand the functioning of the developer social platforms and have had a chance to use these platforms during their degrees and jobs.
- I wanted to compare the platform usage at different level of education, and it would not be fair to compare first year undergraduate student with a PhD student.
- I also wanted to include at least some participants who have worked in the software development industry before in order to understand the role of work experience in the interaction with developer social platforms. Students in their first or second year of degrees usually do not have that kind of work experience.

Second, there is a limitation associated with conducting the study with the students of only the University of Victoria. The curriculum, the semesters where a students can enroll in an internship, and the length of the program varies for different universities. For example, some universities do not have cooperative education programs such as the ones that the University of Victoria offers, hence the students of other universities might not have any work experience. This means that recruiting participants from different universities may show a difference in the findings when comparing the participants with different levels of education in terms of their industrial experience using social coding platforms.

7.3.2 Participant Recruitment

During the recruitment phase for both the semi-structured interviews and the survey, I explicitly mentioned my interest in understanding how students use GitHub and Stack Overflow, and shared my goal to identify barriers and benefits they may experience. This information may have influenced who participated in the semi-structure interviews and the survey: in particular, I expected that students that do not use these tools much would not have participated.

Chapter 8

Conclusion

In this chapter, I suggest possible avenues for future work, as well as conclude the thesis by reiterating the important takeaways from this research.

8.1 Future Work

This research investigated student interactions with developer social platforms such as GitHub and Stack Overflow, and identified barriers and benefits that the students experience while using these platforms. There has not been a lot of research dedicated to students as a sub-community of users contributing to and accessing content from developer social platforms. Also this research has some limitations, hence there are a number of possible areas for future work and follow-up studies.

One interesting study would be to weigh different factors that leads to the barriers that students experience and how different barriers and benefits are correlated. This will help researchers understand the core barriers that indirectly give rise to other barriers, and what benefits help in overcoming what barrier. Many participants of this study indicated that if they see their peers performing a certain activity, they will get motivated to do the same. It will be interesting to calculate the effects of peer activities on the students when it comes to interaction with developer social platforms and to conduct qualitative studies to explore the effect of peer parity on the student participation trends.

Also, in Section 5.4, I present some data about students who contributed to Stack Overflow once and never returned. There can be studies conducted to understand what was their initial motivation to make the first contribution and how that changed

after they made their first contribution. This study could unravel some barriers to entry and can help in making Stack Overflow more accessible for students.

Through this research, I learned the student side of the story when it comes to interacting with developer social platforms. An extension of this study would be to conduct the survey with different stakeholders to understand their perception on student participation, what different stakeholders feel will be important steps to increase student participation and how those steps mitigate the barriers I identified with this study. Such study will help in making the developer social platforms more accessible for students and it will also give an opportunity to stakeholders like platform designers or the senior community members to think about mitigating the barriers to entry and this will result in the growth of different communities around the developer social platforms.

Lastly, as I discussed in Chapter 7, one of the major limitation of this study is a lack of statistical analysis of barriers. It will be interesting to create mathematical models based on the survey results that would help creating a better design for follow-up surveys.

8.2 Conclusion

Developer social platforms encompasses both social as well as technical aspect of software development. These platforms grow only when people decide to open the code they create for public and collaborate with other users. Motivated by investigating student interaction with this platform and observing a lack of students involvement on the developer social platform, my goal was primarily to investigate the barriers that the students face when interacting with developer social platforms.

The first major contribution of this thesis is a detailed account of a mixed methods study to understand user participation and identifying barriers and benefits to participation. The exploration process through semi-structured interviews was a team effort involving several round of pilot interviews to identify primary set of participation trends of students, and the barriers and benefits they experience on developer social platform. Then I designed a survey using the results obtained from the semi-structured interviews which, in addition to extending the study to a bigger population also validated my primary results, giving a list of barriers and benefits experienced by students.

These findings lay a foundation for understanding the mindset of students and will help different stakeholders such as future researchers, employers, educators, and students themselves learn more about these experiences and act on them.

Bibliography

[1] Ifeoma Adaji and Julita Vassileva. Towards understanding user participation in stack overflow using profile data. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10047 LNCS(August 2018):3–13, 2016.

[2] Babatunde Femi Akinyode and Tareef Hayat Khan. Step by step approach for qualitative data analysis. *International Journal of Built Environment and Sustainability*, 5(3):163–174, 2018.

[3] Yair Amichai-Hamburger, Tali Gazit, Judit Bar-Ilan, Oren Perez, Noa Aharony, Jenny Bronstein, and Talia Sarah Dyne. Psychological factors behind the lack of participation in online discussions. *Computers in Human Behavior*, 55:268–277, 2016.

[4] Filade and Bankole Adeyemi. Peer group influence on academic performance of undergraduate students in Babcock University, Ogun State. *African Educational Research Journal*, 7(2):81–87, 2019.

[5] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Discovering value from community activity on focused question answering sites: A case study of stack overflow. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 850–858, 2012.

[6] Gary S Becker. Investment in human capital: A theoretical analysis. *Journal of political economy*, 70(5, Part 2):9–49, 1962.

[7] Avijit Bhattacharjee, Sristy Sumana Nath, Shurui Zhou, Debasish Chakroborti, Banani Roy, Chanchal K Roy, and Kevin Schneider. An exploratory study to find motives behind cross-platform forks from software heritage dataset. pages 11–15, 2020.

[8] A. Bosu, C. S. Corley, D. Heaton, D. Chatterji, J. C. Carver, and N. A. Kraft. Building reputation in stackoverflow: An empirical investigation. In *2013 10th Working Conference on Mining Software Repositories (MSR)*, pages 89–92, 2013.

[9] Caroline Bulsara. Using a mixed methods approach to enhance and validate your research. *Brightwater Group Research Centre*, pages 1–82, 2015.

[10] Xiufeng Cheng, Ziming Zhang, Yue Yang, and Zhonghua Yan. Open collaboration between universities and enterprises: a case study on github. *Internet Research*, 2020.

[11] Jailton Coelho and Marco Tulio Valente. Why modern open source projects fail. *Proceedings of the ACM SIGSOFT Symposium on the Foundations of Software Engineering*, Part F1301:186–196, 2017.

[12] Eleni Constantinou and Georgia M. Kapitsaki. Identifying Developers' Expertise in Social Coding Platforms. *Proceedings - 42nd Euromicro Conference on Software Engineering and Advanced Applications, SEAA 2016*, (1):63–67, 2016.

[13] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. Social coding in github: Transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, page 1277–1286, New York, NY, USA, 2012. Association for Computing Machinery.

[14] Melissa DeJonckheere and Lisa M. Vaughn. Semistructured interviewing in primary care research: A balance of relationship and rigour. *Family Medicine and Community Health*, 7(2):1–8, 2019.

[15] Risqi Puspa Dewi, Mira Kania Sabariah, and Veronikha Effendy. Design Recommendation Information Architecture of Hospital Website Using Bottom-Up Approach on Card Sorting Method. *International Journal on Information and Communication Technology (IJoICT)*, 5(2):20, 2020.

[16] Steve Easterbrook, Janice Singer, Margaret-Anne Storey, and Daniela Damian. *Selecting Empirical Methods for Software Engineering Research*, pages 285–311. Springer London, London, 2008.

[17] Noella Edelmann. Reviewing the definitions of "lurkers" and some implications for online research. *Cyberpsychology, Behavior, and Social Networking*, 16(9):645–649, 2013.

[18] Victoria Elliott. Thinking about the coding process in qualitative data analysis. *Qualitative Report*, 23(11):2850–2861, 2018.

[19] Joseph Feliciano, Margaret Anne Storey, and Alexey Zagalsky. Student experiences using GitHub in software engineering courses: A case study. *Proceedings - International Conference on Software Engineering*, pages 422–431, 2016.

[20] Denae Ford, Alisse Harkins, and Chris Parnin. Someone like me: How does peer parity influence participation of women on stack overflow? *Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC*, 2017-Octob:239–243, 2017.

[21] Denae Ford, Justin Smith, Philip J. Guo, and Chris Parnin. Paradise unplugged: Identifying barriers for female participation on stack overflow. *Proceedings of the ACM SIGSOFT Symposium on the Foundations of Software Engineering*, 13-18-Nove:846–857, 2016.

[22] R. G.U.S. Gajanayake, M. H.M. Hiras, P. I.N. Gunathunga, E. G. Janith Supun, Anuradha Karunasenna, and Pradeepa Bandara. Candidate selection for the interview using GitHub profile and user analysis for the position of software engineer. *ICAC 2020 - 2nd International Conference on Advancements in Computing, Proceedings*, pages 168–173, 2020.

[23] Alexander Hars and Shaosong Ou. Working for free? Motivations for participating in open-source projects. *International Journal of Electronic Commerce*, 6(3):25–39, 2002.

[24] Tung-Hsing Hsiung, Stephen Olejnik, and Carl J. Huberty. Comment on a Wilcox Test Statistic for Comparing Means When Variances Are Unequal. *Journal of Educational Statistics*, 19(2):111–118, 1994.

[25] Carruthers John. A Rationale for the Use of Semi-structured Interviews. *Journal of Educational Administration*, 28(1), jan 1990.

[26] Karim R Lakhani and Robert G Wolf. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. 2003.

[27] Sean Lochrie, Ross Curran, and Kevin O'Gorman. Qualitative Data Gathering Techniques. *Research Methods for Business and Management*, 2015.

[28] Joseph e Mcgrath. Methodology Matters: Doing Research in the Behavioral and social sciences. In Ronald m Baecker, Jonathan Grudin, William s Buxton, and Saul Greenberg, editors, *Readings in Human–Computer Interaction*, Interactive Technologies, pages 152–169. Morgan Kaufmann, 1995.

[29] Wes McKinney. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc.", 2012.

[30] Catherine McLoughlin and Mark J.W. Lee. Social software and participatory learning: Pedagogical choices with technology affordances in the web 2.0 era. *ASCILITE 2007 - The Australasian Society for Computers in Learning in Tertiary Education*, pages 664–675, 2007.

[31] Christopher Mendez, Hema Susmita Padala, Zoe Steine-Hanson, Claudia Hilderbrand, Amber Horvath, Charles Hill, Logan Simpson, Nupoor Patil, Anita Sarma, and Margaret Burnett. Open source barriers to entry, revisited: A sociotechnical perspective. *Proceedings - International Conference on Software Engineering*, pages 1004–1015, 2018.

[32] Johann Mitlöhner, Sebastian Neumaier, Jürgen Umbrich, and Axel Polleres. Characteristics of open data csv files. In *2016 2nd International Conference on Open and Big Data (OBD)*, pages 72–79. IEEE, 2016.

[33] David L. Morgan. Mixed methods research. *The Cambridge Handbook of Sociology*, 1:153–161, 2017.

[34] Lorraine Morgan and Patrick Finnegan. Benefits and drawbacks of open source software: An exploratory study of secondary software firms. *IFIP International Federation for Information Processing*, 234:307–312, 2007.

[35] Theodore M Newcomb. Student peer-group influence. 1962.

[36] Rose Khavari Nicholas Dias, Yung Peng. HHS Public Access. *Physiology & behavior*, 176(3):139–148, 2017.

[37] Blair Nonnecke, Dorine Andrews, and Jenny Preece. Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research*, 6(1):7–20, 2006.

[38] Saya Onoue, Hideaki Hata, and Ken Ichi Matsumoto. A study of the characteristics of developers' activities in GitHub. *Proceedings - Asia-Pacific Software Engineering Conference, APSEC*, 2:7–12, 2013.

[39] S.Pli R.For. Qualitative and Quantitative Sampling. (1981):220–239, 1987.

[40] Adam Rosenstein, Aishma Raghu, and Leo Porter. Identifying the prevalence of the impostor phenomenon among computer science students. *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*, pages 30–36, 2020.

[41] Adam Rule, Aurélien Tabard, and James D Hollan. Exploration and explanation in computational notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[42] Robert M. Sauer. Why develop open-source software? The role of non-pecuniary benefits, monetary rewards, and open-source licence type. *Oxford Review of Economic Policy*, 23(4):605–619, 2007.

[43] Jefferson Silva, Igor Wiese, Daniel M. German, Christoph Treude, Marco Aurélio Gerosa, and Igor Steinmacher. A theory of the engagement in open source projects via summer of code programs. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, ESEC/FSE 2020, page 421–431, New York, NY, USA, 2020. Association for Computing Machinery.

[44] Leif Singer, Fernando Figueira Filho, Brendan Cleary, Christoph Treude, Margaret Anne Storey, and Kurt Schneider. Mutual assessment in the social programmer ecosystem: An empirical investigation of developer profile aggregators. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, pages 103–116, 2013.

[45] Donna Spencer and Todd Warfel. Card sorting: a definitive guide. page 2, 2004.

[46] StackOverflow. Stack overflow report 2019. 2019. [Online; accessed 19-April-2020].

[47] Wouter Stam. When does community participation enhance the performance of open source software companies? *Research Policy*, 38(8):1288–1299, 2009.

[48] Igor Steinmacher, Tayana Conte, Marco Aurélio Gerosa, and David Redmiles. 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*, CSCW '15, page 1379–1392, New York, NY, USA, 2015. Association for Computing Machinery.

[49] Igor Steinmacher, Tayana Uchoa Conte, Christoph Treude, and Marco Aurélio Gerosa. Overcoming open source project entry barriers with a portal for newcomers. In *Proceedings of the 38th International Conference on Software Engineering*, ICSE '16, page 273–284, New York, NY, USA, 2016. Association for Computing Machinery.

[50] Igor Steinmacher, Marco Aurelio Graciotto Silva, Marco Aurelio Gerosa, and David F. Redmiles. A systematic literature review on the barriers faced by newcomers to open source software projects. *Information and Software Technology*, 59:67–85, 2015.

[51] Na Sun, Patrick Pei Luen Rau, and Liang Ma. Understanding lurkers in online communities: A literature review. *Computers in Human Behavior*, 38:110–117, 2014.

[52] Bogdan Vasilescu, Vladimir Filkov, and Alexander Serebrenik. StackOverflow and GitHub: Associations between software development and crowdsourced knowledge. *Proceedings - SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013*, pages 188–195, 2013.

[53] Justin W. Vollet, Thomas A. Kindermann, and Ellen A. Skinner. In peer matters, teachers matter: Peer group influences on students' engagement depend on teacher involvement. *Journal of Educational Psychology*, 109(5):635–652, 2017.

[54] Zhendong Wang, Yi Wang, and David Redmiles. Competence-confidence gap: A threat to female developers' contribution on github. In *2018 IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, pages 81–90, 2018.

[55] Xin Xia, Lingfeng Bao, David Lo, Pavneet Singh Kochhar, Ahmed E. Hassan, and Zhenchang Xing. What do developers search for on the web? *Empirical Software Engineering*, 22(6):3149–3185, 2017.

[56] Deheng Ye, Zhenchang Xing, and Nachiket Kapre. The structure and dynamics of knowledge network in domain-specific q&a sites: a case study of stack overflow. *Empirical Software Engineering*, 22(1):375–406, 2017.

[57] Yue Yu, Gang Yin, Huaimin Wang, and Tao Wang. Exploring the patterns of social behavior in Github. *1st International Workshop on Crowd-Based Software Development Methods and Technologies, CrowdSoft 2014 - Proceedings*, pages 31–36, 2014.

[58] Alexey Zagalsky, Joseph Feliciano, Margaret Anne Storey, Yiyun Zhao, and Weiliang Wang. The emergence of GitHub as a collaborative platform for education. *CSCW 2015 - Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing*, pages 1906–1917, 2015.

[59] Haoxiang Zhang, Shaowei Wang, Tse-Hsun Chen, and Ahmed E. Hassan. Reading answers on stack overflow: Not enough! *IEEE Transactions on Software Engineering*, pages 1–1, 2019.

Appendices

Appendix A

Semi-structured Interview Questions

A.1 Participant Background Information

Please fill the information provided below.

Degree:

Year of Study:

Gender:

Occupation (If working):

A.2 General Questions related to GitHub

Q1. Have you taken any courses previously where you created projects or assignments that involve programming?

Q2- Have you been involved in paid projects outside school assignments, for example during an internship or a job that involved programming?

Q3- Are you involved in programming projects or applications which are not a part of paid job or assignments?

If the answer is yes:

Q3.1- What kind of projects do you usually participate in, when you do not have any benefits associated with it (For example open competitions or open source projects aside from the assignments)?

Q3.2- Can you please explain in detail about your area of interest, in terms of project development? (Optional)

Q3.3- If you're given an opportunity to develop an open source project, do you prefer to work alone on these projects or do you like to collaborate with other people?

Q3.4- Do you publish your work online to let others view what you're working on or do you like to keep the projects to yourself?

Q3.5- How do you like to save your project's source code? Do you save different versions of your code, how do you do that?

Q3.6- Have you tried to find collaborators for your project?

Q3.7- If you do not find people whom you already know, do you go outside your known circle of people? What communication channel do you use to collaborate with your project partners?

Q3.8- How do you find time for development activities that involves contributing to the projects?

If no:

Q3.1- Have you considered participating in the open source projects? Given that you do not participate, what stopped you from participating?

If they say, we do not have time:

Q3.1.1- What all activities (as in courses, club related activities, teaching assis-

tance related activities to name a few) which involves school and work commitments do you participate in throughout the week that consume the majority of your time?

Q3.2- Do you have any hobbies outside Computer programming? (Don't tell me your hobbies if you're not comfortable as it is a yes or no question)

Q3.3- Do you get a chance to pursue your hobbies outside computer science, given the busy schedule that you typically have?

Q3.4- If you are on a vacation from work or school, would you participate in development of projects for yourself or collaborate with others?

Turning back to the course work:

Q4- How do you store the work that you created as a part of the Assignment and project?

Q5- What platform do you use in order to store your work online?

Q6- Please describe a personal experience with the platform you use?

A.3 Participation related Questions for GitHub

Q7- Do you have a Github account? If not, then do you have any other social coding account?

Q8- How long have you been using the version control system you mentioned before?

Q9- Have you taken any course at school which involves using git in projects?

Q9.1- Was the task meant to be done by a group of students or as an individual?

Q9.2- Can you describe the task assigned during the course to you, that lead to interaction with git. And what was the most challenging part?

Q10- Are you aware of the process of generating pull request for projects, and have you generated a pull request in an existing project?

Q10.1- Did you know the owner of the repository?

Q10.2- Was that pull request a part of a course, a part of a job, or was a part of the project you were developing out of interest?

Q10.3- What did you contribute to the project? (For example: edited the code, added documentation, fixed a bug, suggested an improvement?)

Q10.4- Was your pull request accepted?

Q11- Do you use command line?

Q12- On what level of expertise do you rate yourself on GitHub?

Thinking back to your last few responses:

Q13- Has your pull request ever been rejected? What was the reason given to you?

Q14- While browsing an existing project on Github, or when you're browsing the project for personal use and you come across a bug or improvement, do you raise an issue?

If yes:

Q14.1- How did you come to the conclusion that there is a reportable issue existing in the repository?

Q14.2- Was your pull request merged into the main repository?

If no:

Q14.1- If you felt that there is some problem with the code, then why did you not raise an issue?

Q14.2- If the project was not opensource, but was from a private repository, would you still not raise an issue?

Moving on to more open ended questions:

Q15- Have you ever resolved an issue on any repository?

If yes:

Q15.1- How did you come to the conclusion that the solution you are proposing could enhance the working of the project or is a suitable solution to the problem?

Q15.2- Did you share your code snippet in the solution to the issue, or did you add a comment your proposed solution?

If no:

Q15.1- If it was not an open project, but was privately owned with limited audience, would you still not share your solution?

Q15.2- If the project belonged to your friend and was placed in a private repository, how would that be different from the opensource case?

Q16- Do you own a public repository on Github with an open project for which you're seeking collaboration?

If yes:

Q16.1- How do you feel about getting pull requests on your repositories?

Q16.2- How do you feel about resolving issues for your repositories?

Q16- What benefits have you experiences while working on Open Source projects?

Q17- Have you ever forked or cloned an existing repository, made constructive changes to it and generated a pull request in order with suggested changes? (IF NO) What is the reason for not suggesting changes?

Q18- What according to you are the benefits of working on projects in a collaborative environment?

Q19- Do you mention your successful open project participation on your resume?

Q20- Have you ever participated in companies' open source projects?

If yes:

Q20.1- What benefits did you get out of it?

A.4 Questions related to Stack Overflow

Q1- What is your go-to website when you need help with your code?

OR

Q1- Which forum do refer to the most, while fixing your answers?

Q2- Do you go directly to the stack overflow with your code problem or do you type your problem in the search engine and then pick stack overflow to view the answers?

Q3- What kind of problems do you usually encounter in the code that you search for help online? (Checking for syntax, fixing errors, improving solutions. . .)?

Q4- Which computer programming forums do you have an account on?

Q5- Do you have a stack overflow account?

Q6- In the Past 3 months, why did you visit stack overflow? (Example: exploring some code, and potential improvements out of curiosity, stuck with an error, and wanted to know the solution, I was looking for a specific piece of code for logic)

Q7- Do you usually get a good answer when you are looking for by searching on stack overflow?

Q8- If you do not find the exact question then, do you post one question yourself? If no then why not?

Q9- In the past 3 months did you participate in any way with stack overflow? (Eg: upvoted / downvoted an answer, uploaded a question, answered, was just there looking)

Q9.1- Do you feel that you are a part of the stack overflow community?

Q9.2- Which of these activities that were being mentioned before would make you feel more included in the community?

Q10- Have you ever reused a code snippet from stack overflow, or modified it to use it for yourself and came up with a better answer?

Q11- If no to the above question then why did you not post the improvement back?

Q12- If you see a lot of classmates posting on the slack overflow, how would that make a difference for you?

Appendix B

Semi-structured Interview Recruitment Documents

B.1 Email to Organizers

Dear CONTACT NAME,

I am PRINCIPAL APPLICANT. I am conducting a research study as a Principal Applicant with my supervisors PRINCIPAL INVESTIGATOR 1, and PRINCIPAL INVESTIGATOR 2 as a part of my Master's of Science (Computer science) Thesis titled "Investigating Barriers to Student Participation in Open Communities: An Exploratory Study". We are a part of Human Computer Interaction and Software Engineering Lab (CHISEL) at the department of Computer Science at the University of Victoria. We are writing to request your participation in the research project. We are looking for adults (19 years or older) who are students enrolled in Computer Science and Software Engineering degree in their Third Year and above, and the Graduate Students in Computer Science Department.

Due to your involvement in Administration, we would like you to help us invite the above mentioned students to take part in the study aimed at understanding how the students participate in Open platforms such as Gitlab, Github, Stack overflow and identifying any hindrances they face while making a contribution to these community platforms. If you agree, I would like to request you to post an invitation to participate in the study in a mailing list of your choice.

Participation is completely voluntary. I will not be contacting anyone in the group

directly. It's up to the individual to decide to participate, and they may withdraw at any time without any consequences or any explanation. If they withdraw from the study, the data related to their participation will be deleted.

All information collected will be anonymized and confidential and the researchers alone are the only ones who will have access to this data. All data will be stored on protected computers located in a research lab at the University of Victoria. Any personally identifiable information will be deleted from the data.

Study data will be kept for five years (starting the beginning of this study) and at the end of this time, data and backups will be deleted from the project repository, and any physical record will be destroyed as appropriate. Data collected and responses to questions will only ever be presented via pseudonym or in aggregate form, as appropriate.

You may verify the ethical approval of this study or raise any concerns you might have by contacting the researchers directly or the Human Research Ethics Office at the University of Victoria.

Please do not hesitate to contact me if you have any questions about this study. Thanks, and have a wonderful day!

B.2 Invitation to Participate

Hi!

I'm PRINCIPAL APPLICANT. I am conducting a research study with my supervisors PRINCIPAL INVESTIGATOR 1, and PRINCIPAL INVESTIGATOR 2 where we're interested in understanding how students participate in Open Communities and identify barriers to their participation on platforms such as GitHub, GitLab or Stack Overflow. The study is a part of my Master's of Science (Computer Science) Thesis titled "Investigating Barriers to Student Participation in Open Communities: An Exploratory Study". We are looking for adults (19 years or older) who use Open Source platforms. We would appreciate speaking with you to learn how you participate in these communities.

We would be delighted if you would be willing to participate in an interview and share your experiences by:

Answer questions about how you use Open communities like Github and Stack Overflow. Answer questions about the barriers you have faced while making contributions in these open communities. Answer questions about how your participation

in these Open Communities has helped you, directly or indirectly.

If you're willing to participate, the study will be conducted in person during a time you find convenient. We estimate that this will take 60 minutes of your time. Please let me know if you would be willing to participate and what time slot works best for you.

We will handle your responses and data confidentially. This is a purely academic study. If you consider this email to be spam, I'm very sorry! There will be no followup to bug you. We will openly publish the results so everyone can benefit from them, but will anonymize everything before doing so.

Regards, PRINCIPAL APPLICANT CHISEL Group (thechiselgroup.org) / Computer Science Department, University of Victoria PRINCIPAL APPLICANT'S EMAIL

B.3 Verbal Consent

At the time of the interview:

I'm PRINCIPAL APPLICANT at the University of Victoria. I'm interested to learn more about how you participate in Open source platforms like Stack Overflow, GitLab or GitHub and how you use these community forums so that I can identify the barriers to participation faced by you and other students.

In general, I'll be asking questions about how often you visit these community platforms, how you use the information being stored there, how you contribute in the community, and how it helps you with your work, directly or indirectly.

This interview is designed to take around 60 minutes of your time. However, if we go over time but you want to share more insights with me, feel free to do so as your contributions will be highly valuable for my research. Also, I want to remind you that this interview is being (audio) recorded for future analysis, your participation is completely voluntary, and you can withdraw at any time. Is that ok with you?

If participant agrees: Great! Now please tell me a bit about yourself.

B.4 Signed Consent

I am PRINCIPAL APPLICANT. I am conducting a research study as a Principal Applicant with my supervisors PRINCIPAL INVESTIGATOR 1, and PRINCIPAL INVESTIGATOR 2 as a part of my Master's of Science (Computer science) Thesis

titled “Investigating Barriers to Student Participation in Open Communities: An Exploratory Study”. We are a part of Human Computer Interaction and Software Engineering Lab (CHISEL) at the department of Computer Science at the University of Victoria. We are looking for adults (19 years of age or older) who are students of the Department of Computer Science or Software Engineering at the University of Victoria, or are on a Coop work term at the moment, or graduated in the year 2019 from Computer Science or Software Engineering department and are working in the industry. The participants should have experience with the use of Open Source Platforms (Gitlab/ Github/ Stack Overflow etc.) to host and collaborate on software projects or to find answers to programming problems.

You are being invited to participate in this study and we would appreciate the opportunity to gather data on how you interact in the Open Communities, understand your communication pattern and identify barriers to participation that students face. We would be delighted if you would be willing to:

(A) participate in an interview where we will ask some questions related to your experience with the various open communities;

(B) be willing to let us record your responses on a voice recorder as well as on paper; and (C) answer questions related to challenges you faced while participating in the community.

We estimate that this will take no more than 60 minutes of your time. This is a purely academic research project with no commercial interests. We will openly publish the results so everyone can benet from them, but will anonymize everything before doing; your responses will be handled confidentially. Please note that you are not obligated to participate in the study. If at some point during the survey you want to stop, you are free to do so without any negative consequences and any data collected up to that point will be discarded.

Thank you for participating!

Participants Selection

You are being asked to participate in this study because you have been identied as a current or recent computer science or software engineering student, and have indicated that you are willing to assist with our study by agreeing to be interviewed.

What is involved

If you consent to voluntarily participate in this research, your participation will include:

During this study, the principal applicant will ask you some questions about how

you participate in Open communities and what barriers do you face in the participation process. With your permission, the session will be audio recorded and written notes will be taken. A transcription will be made from the audio recording.

Risks

There are no known or anticipated risks to you by participating in this research. The study will be completed in a private laboratory with only the principal applicant present.

Benets The potential benets of your participation in this research include benets to the state of knowledge, particularly in software development.

Voluntary Participation

Your participation in this research must be completely voluntary. If you do decide to participate, you may withdraw at any time without any consequences or any explanation. If you do withdraw from the study your interview responses will be deleted, and the voice recording will be deleted. The material will not be used in any form in our current (and future) study.

Anonymity

In terms of protecting your anonymity the interview will not ask for any identifying information. Any private information given to us will be removed in the nal analysis, stripped and replaced with anonymous identiers.

Condentiality

Your condentiality and the condentiality of the data will be protected by access restricted and password protected computers located in a research lab at the University of Victoria. Only the principal applicant will be able to access the complete data. After conducting the interviews, the data will be stripped of any condential information and stored in a private Github Repository. Dr. Margaret Anne Storey will only access the anonymized data. Please be advised that this research study includes data storage on the private Github repository of the CHISEL lab. As such, there is a possibility that the anonymized information about you that is gathered for this research study that is stored on Github may have servers located in the U.S.A. and may be accessed without your knowledge or consent by the U.S. government in compliance with the U.S. Patriot Act.

Researchers' Relationship with Participants

The investigators have no relationship with any of the participants, The Principal Investigator will not know who has enrolled in the study or has declined to participate. She will not participate in recruitment or send out emails to students in the course

if she is the course instructor. PRINCIPAL APPLICANT (Principal Applicant) will solely conduct the recruitment process and will remove any identifiable record from the data collected except for the group they belong to in order to study the differences in the responses between groups, before presenting data to the Principle Investigator and other members in the research team, hence identity of the participants will remain concealed.

Dissemination of Results

It is anticipated that the results of this study will be shared with others in the following ways: thesis, dissertation, class presentations, published articles, chapters or books and presentations at scholarly meetings. Also the thesis for which this study is intended for will be made publicly available online at the website "Uvic Space" and at the personal website of the "Principal Applicant" (Trishala Bhasin).

Disposal of Data

Data collected in the study will be disposed of a maximum of 5 years after the beginning of the study. All data collected in the study and backups will be electronically deleted. All the physical information will be destroyed as appropriate (paper will be destroyed using a paper shredder).

Contacts

Individuals that may be contacted regarding this study include:

PRINCIPAL APPLICANT'S EMAIL PRINCIPAL INVESTIGATORS 1 and 2 EMAIL

In addition, you may verify the ethical approval of this study, or raise any concerns you might have, by contacting the Human Research Ethics Ofce at the University of Victoria.

Your signature below indicates that you understand the above conditions of participation in this study, that you have had the opportunity to have your questions answered by the researchers, and that you consent to participate in this research project.

Name of Participant

Signature

Date

A copy of this consent will be left with you, and a copy will be taken by the researcher

Appendix C

Survey Questions

C.1 Demographic Information

Social coding platforms are online tools where you can share your code with different users, develop projects collaboratively and use Git versioning in your projects. This survey will take 10-15 minutes of your time. I am gathering information as to how you, as a student, interact with social coding platforms, and question/answer forums like Stack Overflow.

While you are not required to answer questions, please answer as many questions as you can in order to help with data collection.

Please contact trishalabhasin@uvic.ca if you face any problems. Thank you so much for participating!

Q1. What degree are you enrolled in / have you recently completed?

Q2. What gender do you identify as?

Q3. Have you completed any co-op term(s) during your current/previous degree(s)?

Q4. How much time have you spent working as a full-time employee in a software development-related job, excluding co-ops / internships?

Q5. What social coding platforms do you have an account on? (Choose all an-

swers that are applicable to you)

Q6. Do you have a link to any of your social coding platforms on your resume?

Q7. Have you ever been asked for a link to your social coding profile when applying for a job?

Q8. Have you ever used GitHub in any software development company you have worked for?

Q9. Have you ever taken any courses that required you to use GitHub?

C.2 GitHub and School

Q10. When building projects for your school courses, which activities have you performed on social coding platforms? (Choose all answers that are applicable to you)

Q11. Have you ever published projects on a social coding platform that you created as a part of your school assignments?

Q12. What motivates you to upload your projects on social coding platforms? (Choose all answers that are applicable to you)

C.3 Personal Interaction with GitHub

This sections contains questions related to your personal interaction with GitHub (and other similar platforms) not relating to school or job-related work.

Please answer these questions purely on the basis of your personal development activities.

Q13. Do you create projects involving programming outside of your school assignments / job tasks?

Q14. What are your motives behind creating personal projects that are not related to your school or work? (Choose all answers that are applicable to you)

Q15. Do you post your personal projects on social coding platforms?

Q16. How many public repositories that contain your personal projects (outside school / work) do you have on GitHub?

Q17. How many private repositories that contain your personal projects (outside school / work) do you have on GitHub?

Q18. How many public repositories that contain your personal projects (outside school / work) do you have on social coding platforms other than GitHub?

Q19. How many private repositories that contain your personal projects (outside school / work) do you have on social coding platforms other than GitHub?

Q20. (Please answer this if you have ever published a project in a public repository on a social coding platform.) What are your motivations for publicly sharing your personal project(s) on social coding platforms such as GitHub? (Choose all answers that are applicable to you)

Q21. (Please answer this if you have ever hesitated to publish your code on a social coding platform.) What stops you from publicly sharing your projects on social coding platforms like GitHub? (Choose all answers that are applicable to you)

Q22. Have you sought collaboration on a project you published on a social coding platform?

Q23. Do you contribute to repositories of users that you do not know (strangers)?

Q24. If you do not contribute to the repositories of people outside your known circle, what is stopping you? (Choose all answers that are applicable to you)

Q25. (Please answer this question only if you do not use GitHub but do use other

platforms for development activities.) Why do you prefer using platforms other than GitHub for your development activities?

C.4 Stack Overflow related Questions

This section contains information related to Stack Overflow.

Q26. When you access Stack Overflow, are you always logged into your account?

Q27. Do you directly go to Stack Overflow site for your coding problems?

C.5 Participation related Questions

This section contains questions related to the participation on Stack Overflow.

Q28. Have you ever contributed to any activity (asking/answering questions, upvote/downvote etc.) on Stack Overflow?

Q29. What activities have you performed on Stack Overflow? (Choose all the answers that are applicable to you)

Q30. If you are not able to find an answer to a problem that you are facing, do you post a question yourself?

Q31. What motivates you to post questions on Stack Overflow? (Choose all the answers that are applicable to you)

Q32. If you're not able to find the solution to a problem then why don't you post a question yourself? (Choose all the answers that are applicable to you)

Q33. Have you ever answered any question(s) on Stack Overflow?

Q34. What motivates you to answer questions on Stack Overflow? (Choose all the answers that are applicable to you)

Q35. What stops you from answering the questions that you might know the answer to on Stack Overflow? (Choose all the answers that are applicable to you)

Q36. Do you collect badges or build reputation on Stack overflow?

Q37. Is there any additional information or feedback that you would like to share with us?

Appendix D

Recruitment Details for Surveys

D.1 Email to Organizers

Dear CONTACT NAME,

I am PRINCIPAL APPLICANT. I am conducting a research study as a Principal Applicant with my supervisors PRINCIPAL INVESTIGATOR 1, and PRINCIPAL INVESTIGATOR 2 as a part of my Master's of Science (Computer science) Thesis titled "Investigating Student Experiences with GitHub and Stack Overflow: An Exploratory Study". We are a part of Human Computer Interaction and Software Engineering Lab at the department of Computer Science at the University of Victoria. We are writing to request your participation in the research project. We are looking for adults (19 years or older) who are students enrolled in Computer Science and Software Engineering degrees in their Third Year and above, and the Graduate Students in Computer Science Department.

Due to your involvement in Administration, we would like you to help us invite the above mentioned students to take part in the study aimed at understanding how the students participate in Github and, Stack overflow and identifying any hindrances and benefits they face while making a contribution to these community platforms. If you agree, I would like to request you to post an invitation to participate in the study that is a 15 minute Survey in a mailing list of your choice.

Participation is completely voluntary. I will not be contacting anyone in the group directly. It's up to the individual to decide to participate, and they may withdraw at any time without any consequences or any explanation. If they withdraw from the

study, the data related to their participation will be deleted.

All information collected will be anonymized and confidential and the researchers alone are the only ones who will have access to this data. All data will be stored on protected computers located in a research lab at the University of Victoria. Any personally identifiable information will be deleted from the data.

Study data will be kept for five years (starting the beginning of this study) and at the end of this time, data and backups will be deleted from the project repository, and any physical record will be destroyed as appropriate. Data collected and responses to questions will only ever be presented via pseudonym or in aggregate form, as appropriate.

You may verify the ethical approval of this study or raise any concerns you might have by contacting the researchers directly or the Human Research Ethics Office at the University of Victoria (CONTACT INFORMATION).

Please do not hesitate to contact me if you have any questions about this study. Thanks, and have a wonderful day!

D.2 Invitation to Participate

Hi PARTICIPANT-NAME

I'm PRINCIPAL APPLICANT. I am conducting a research study with my supervisors PRINCIPAL INVESTIGATOR 1, and PRINCIPAL INVESTIGATOR 2 where we're interested in understanding how students participate in GitHub and Stack Overflow, and identify barriers to their participation on these platforms. The study is a part of my Master's of Science (Computer Science) Thesis titled "Investigating Student Experiences with GitHub and Stack Overflow: An Exploratory Study". We are looking for adults (19 years or older) who are aware of the usage of GitHub and Stack Overflow. We would be delighted if you would be willing to participate in a Survey and share your experiences by:

Answer questions about how you use social coding platforms like Github and Stack Overflow. Answer questions about the barriers you have faced while making contributions in these open communities. Answer questions about how your participation in these Open Communities has helped you, directly or indirectly.

If you're willing to participate, please contact me at PRINCIPAL APPLICANT'S EMAIL. We estimate that this will take 10-15 minutes of your time.

We will handle your responses and data confidentially. This is a purely academic study. If you consider this email to be spam, I'm very sorry! There will be no followup to bug you. We will openly publish the results so everyone can benefit from them, but will anonymize everything before doing so.

Regards, PRINCIPAL APPLICANT CHISEL Group (thechiselgroup.org) / Computer Science Department, University of Victoria PRINCIPAL APPLICANT'S EMAIL

D.3 Signed Consent Form

I am PRINCIPAL APPLICANT. I am conducting a research study as a Principal Applicant with my supervisors PRINCIPAL INVESTIGATOR 1, and PRINCIPAL INVESTIGATOR 2 as a part of my Master's of Science (Computer science) Thesis titled "Investigating Student Experiences with GitHub and Stack Overflow: An Exploratory Study". We are a part of Human Computer Interaction and Software Engineering Lab (CHISEL) at the department of Computer Science at the University of Victoria. We are looking for adults (19 years of age or older) who are students of the Department of Computer Science or Software Engineering at the University of Victoria, or are on a Coop work term at the moment, or graduated in the year 2019 from Computer Science or Software Engineering department and are working in the industry. The participants should be aware of the use of Github/ Stack Overflow to host and collaborate on software projects or to find answers to programming problems. You are being invited to participate in this study and we would appreciate the opportunity to gather data on how you interact on GitHub and Stack Overflow, understand your communication pattern and identify barriers and benefits to participation that students face. We would be delighted if you would be willing to: (A) participate in a survey that contains questions related to your experience with GitHub and Stack Overflow;

(B) be willing to let us record your responses on a voice recorder as well as on paper; and (C) answer questions related to challenges you faced while participating in the community. We estimate that this will take no more than 15 minutes of your time. This is a purely academic research project with no commercial interests. We will openly publish the results so everyone can benefit from them, but will anonymize everything before doing so; your responses will be handled confidentially. Please note that you are not obligated to participate in the study. If at some point during the survey you want to stop, you are free to do so without any negative consequences and

any data collected up to that point will be discarded. Thank you for participating! Participants Selection You are being asked to participate in this study because you have been identified as a current or recent computer science or software engineering student, and have indicated that you are willing to assist with our study by agreeing to be interviewed.

What is involved

If you consent to voluntarily participate in this research, your participation will include: Answering to close-ended multiple choice questions related to your interactions with Github and Stack Overflow. Risks

There are no known or anticipated risks to you by participating in this research. Benefits The potential benefits of your participation in this research include benefits to the state of knowledge, particularly in software development.

Voluntary Participation

Your participation in this research must be completely voluntary. If you do decide to participate, you may withdraw at any time without any consequences or any explanation. If you do withdraw from the study your survey responses will be deleted. The material will not be used in any form in our current (and future) study.

Anonymity

In terms of protecting your anonymity the survey will not ask for any identifying information. Any private information given to us will be removed in the final analysis, stripped and replaced with anonymous identifiers.

Confidentiality

Your confidentiality and the confidentiality of the data will be protected by access restricted and password protected computers located in a research lab at the University of Victoria. Only the principal applicant will be able to access the complete data. After receiving your survey response, the data will be stripped of any confidential information and stored in a private Github Repository. Dr. Margaret Anne Storey, and Dr. Adam Murray will only access the anonymized data. Please be advised that this research study includes data storage on the private Github repository of the CHISEL lab. As such, there is a possibility that the anonymized information about you that is gathered for this research study that is stored on Github may have servers located in the U.S.A. and may be accessed without your knowledge or consent by the U.S. government in compliance with the U.S. Patriot Act. Researchers' Relationship with Participants The investigators have no relationship with any of the participants, The Principal Investigator Dr. Margaret Anne Storey, and co-supervisor Dr. Adam Mur-

ray will not know who has enrolled in the study or has declined to participate. She will not participate in recruitment or send out emails to students in the course if she is the course instructor. PRINCIPAL APPLICANT (Principal

Applicant) will solely conduct the recruitment process and will remove any identifiable record from the data collected except for the group they belong to in order to study the differences in the responses between groups, before presenting data to the Principle Investigator and other members in the research team, hence identity of the participants will remain concealed.

Dissemination of Results

It is anticipated that the results of this study will be shared with others in the following ways: thesis, dissertation, class presentations, published articles, chapters or books and presentations at scholarly meetings. Also the thesis for which this study is intended for will be made publicly available online at the website "Uvic Space" and at the personal website of the "Principal Applicant" (Trishala Bhasin).

Disposal of Data

Data collected in the study will be disposed of a maximum of 5 years after the beginning of the study. All data collected in the study and backups will be electronically deleted. All the physical information will be destroyed as appropriate (paper will be destroyed using a paper shredder).

Contacts

Individuals that may be contacted regarding this study include: PRINCIPAL APPLICANT'S EMAIL PRINCIPAL INVESTIGATOR 1'S EMAIL PRINCIPAL INVESTIGATOR 2'S EMAIL In addition, you may verify the ethical approval of this study, or raise any concerns you might have, by contacting the Human Research Ethics Office at the University of Victoria. Your signature below indicates that you understand the above conditions of participation in this study, that you have had the opportunity to have your

questions answered by the researchers, and that you consent to participate in this research project.

Name of Participant

Signature

Date

A copy of this consent will be left with you, and a copy will be taken by the researcher

Appendix E

Additional Data from the Survey

(This page deliberately left blank to fit the table in the next page)

E.1 Tables for Qualitative Data Analysis

QUESTION	RESPONSES	UG	Grad	Total
Motivation to upload project related to school or job on social coding platforms	Any public-facing projects add to my work portfolio	19	7	26
	I use these platforms to keep a record of my work	18	10	28
	I upload my projects on social coding platforms to use VSC	11	5	16
	Other*			1
Motives behind creating personal projects that are not related to school or job	Programming is my hobby	18	3	21
	To enhance my skills and learn new languages	22	5	27
	To add to my resume	17	5	22
	I only make projects for my career	2	6	8
	Other**			5
Motives behind publicly sharing personal projects on GitHub	I like developing projects and sharing good ones publicly	14	5	19
	I keep my personal work public for people to collaborate	11	2	13
	I like keeping my code open to use	10	4	14
	Public-facing projects add to my work portfolio	20	7	27
	To gain reputation in the community	12	5	17
	Other***			7
Reasons for hesitation to publicly share projects on social coding platforms	My projects are not good enough to be shared publicly	11	3	14
	Maintaining public projects is big work	4	5	9
	No time for public projects	3	4	7
	There are no benefits to sharing public facing projects	4	0	4
	I have projects in private that I will share when ready	14	5	19
	Other****			5

Table E.1: Barriers and Benefits that students experience when sharing personal projects on developer social platforms

QUESTION	RESPONSES	UG	Grad	Total
Sought collaboration on projects you publish on social coding platforms	Yes, I link my project repository on various forums to find collaborators	6	2	8
	Yes, but only within my known circle not strangers	8	1	9
	No, finding collaborators is too much work	1	4	5
	No, I never thought about finding collaborators on my projects	5	5	10
	Why would anyone collaborate with me? There are other better projects	2	0	2
	Other*			4
Contribute to repos of users that you do not know (strangers)	Yes, I contribute if I come across something interesting	9	3	12
	Yes, if it does not take much of my time	3	1	4
	No, It is very time consuming	7	4	11
	No, my skills are not sufficient	7	0	7
	No, I am not interested in contributing to stranger's code	6	0	6
What stops you from collaborating with people outside your known circle	Collaboration is a lot of work, if I have time and it is easy, I will contribute	8	4	12
	I am not comfortable collaborating with people I do not know	5	2	7
	Collaboration will not benefit me in any way	0	2	2
	When I gain more skills and experience, I will participate	13	3	16
	Other**			5

Table E.2: Barriers and benefits experienced by our participants by collaboration on developer social platforms

Appendix F

Coding Handbook for Semi-structured Interviews

Coding Handbook

for the study titled 'Investigating Barriers to Student Participation in Open Communities: An Exploratory Study'

Thank you so much for participating in the reviewing process and agreement sessions for the above study. In order to participate in the study's review process, you will be provided with a transcript of an interview along with a list of 38 codes divided into 10 groups. You have to go through the transcript, which is an interview conducted by me with the participant about their participation habits around open community participation. While going through the transcript, you have to highlight the line(s) that you feel are corresponding to one of the barriers from the list, and add a comment to indicate what barrier do you feel is the line an instance of. If you feel a particular line is indicating an important piece of information, please feel free to mark it and add a note. The investigators will add the code or enhance the code description.

You can use .docx file or atlas.ti to go through the transcript. The process of coding is described in the following sections.

1. **Viewing the transcript**

If using Word or .docx, the file will be provided to you. If using the atlas.ti tool, these are the steps to be followed:

Launch Atlas.ti > Project > Import Project > *open the project being sent to you*

Once the project is loaded, go to Documents > Documents Manager > *Click on the transcript that is to be coded*

2. **Viewing the codes**

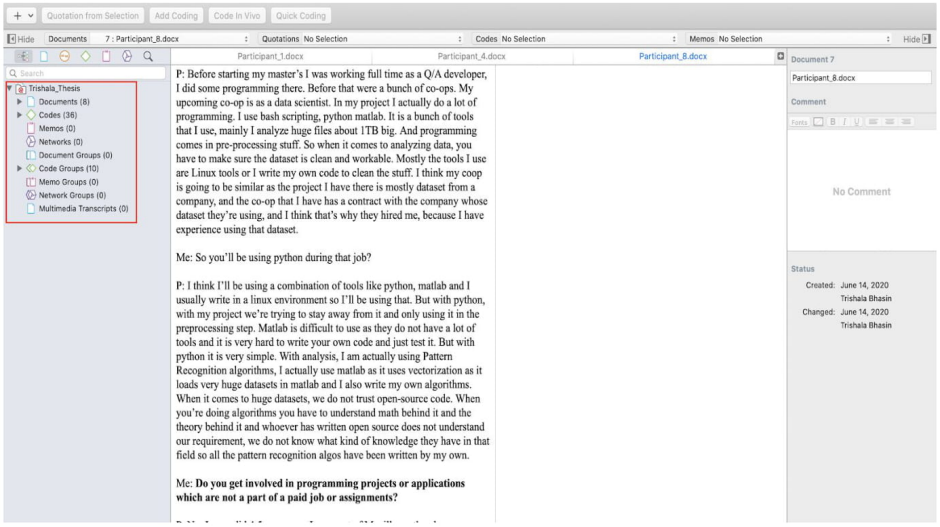
If not using atlas.ti, we will provide a pdf page of all the codes segregated in groups that can be referenced while marking instances on the .docx file.

If using Atlas.ti: Codes > Code Manager

This will display all the codes on the right, groups on the left and comments associated with each code, if you want to understand the code better.

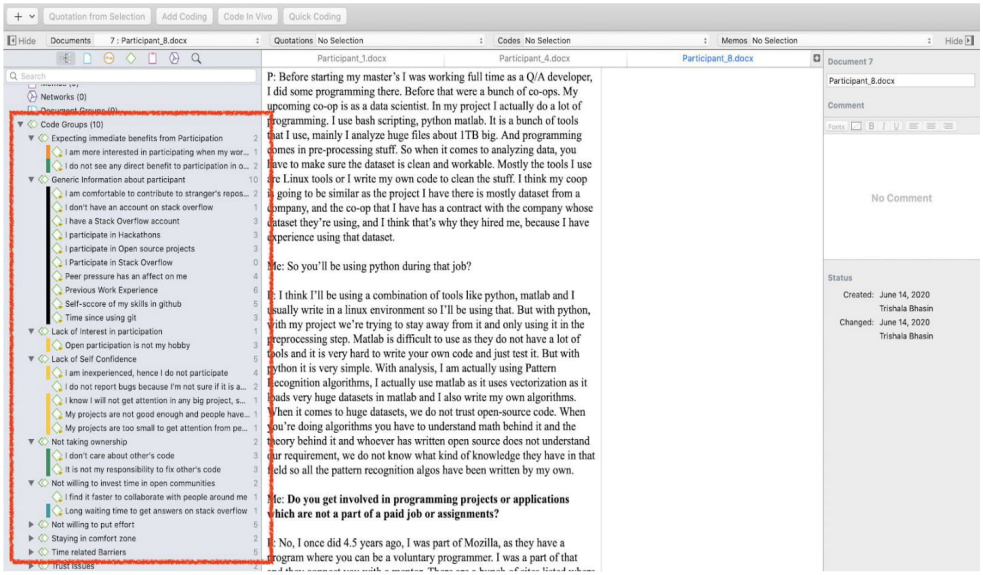
3. **Viewing transcript and coding on Atlas.ti**

When you're on the document manager screen, you will see some tabs on the left on your screen



Here, you can expand the 'code groups' tab. It will display all the code groups available. Then you can expand each group to see all the groups. It will be best to expand all the groups so that you can read and refer to all the codes easily.

It should look like this:



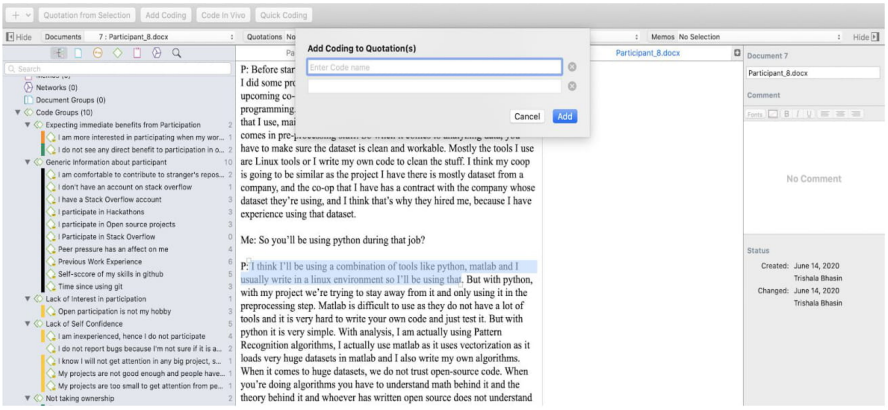
You can click on individual codes to read the researcher's description of the codes provided in the comment section.

4. **Marking the transcript**

If using .docx , use the simple process of selecting the line that you want to assign a code to, right click and add a comment with the name of the code provided in the pdf.

If using the atlas.ti, the process is even simpler!

Select the line you want to assign code to, right click and select 'Add Code option'. A window will pop up which will have a bar where you can start typing the code. As you type the code, a list of codes will appear and you can select the code you wanted to assign from there to your highlighted segment.



Your changes will be saved automatically.

5. **Exporting the project to send it back to me**

In order to send the code back to me through atlas.ti:

Project > Export project > *Save the zip file and transfer it to me via email or slack*

Thank you so much for participating in my study and feel free to ask me more questions on the coding process.