An Analysis of Human Aspects of Collaborative Group Members in OSS development

by

Sheik Shameer Sheik Saleem

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Open Source Software development is a collective activity that involves different software developers who may differ from each other. Although, previous researchers have focused on technical aspects like code factors, technology used etc, recently researchers have explored non-technical human aspects like personality, ethnicity and gender to measure various outcomes. This research assists the emerging state-of-the-art body on diversity research with an empirical study that analyzes how the personality and, the race and ethnic diversity of members in a collaborative group relates to their collaborative contributions in Open Source Software(OSS) development.

This research contains two parts - In the first part we analyse the collaborative group members' personalities and frequency of their collaborative contributions. In the second part we analyse the relationship between the diversity of collaborative group members' race and ethnicity, and the frequency of their collaborative contributions in GitHub. We infer collaborative groups within a project based on the collaboration between software developers in that project. Since previous studies have shown pull requests as the major contribution for a developer to be accepted as a group member, we measure the collaborative contributions of the group members by the number of pull request the group members have merged collaboratively.

Our results from the first part of our research, indicate that 1) the personality traits of collaborative group members does have a relationship with the frequency of their collaborative contributions. Specifically, the more conscientious and less extroverted the group members are, the more contributions that the group members merged. Furthermore, 2) groups that are more diverse with respect to Conscientiousness or Neuroticism have a negative relationship with the frequency of their collaborative contributions. Finally, 3) collaborative groups that are having a majority of highly open, conscientious, or neurotic developers have a positive relationship with their collaborative contributions as well.

Also, from the second part of our research, We observe that (1) a major part of the developer population are White developers; (2) homogeneous and heterogeneous collaborative groups, with respect to race and ethnicity of their group members, have a different distribution of collaborative contributions, with heterogeneous groups having more number of contributions than homogeneous groups and (3) Diversity of race and ethnicity of members in a collaborative group does have a positive statistically significant relationship with the frequency of collaborative group members' contributions.

Acknowledgements

I would like to thank my supervisor Prof Meiyappan Nagappan for his constant feedback and guidance without which the research would not have been possible. I also would like to thank Prof Shane McIntosh and Prof Jesse Hoey for agreeing to read my thesis and provide valuable feedback.

I also thank the members of the Software Architecture Group, especially Dr. Gema Rodriguez Perez for her continuous support, discussions, reviews and suggestions for the writing of this thesis. Lastly, I also thank my family and friends for their encouragement and support.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Software development is typically a collective process that concerns not only technical aspects of how to build and maintain software products, but also concerns human aspects like personalities, views and attitudes [41] as this process is done by developers. In Software Engineering (SE), diversity of developers can be desirable as it helps addressing a problem from different backgrounds, and also it helps designing more robust software products [39]. Also over the last few years big tech companies[1] [2] [3] have increased their efforts to create more diverse software groups. This research contains two parts, both parts focus on finding the relationship between a human aspect namely, personality, and the race and ethnic diversity of the collaborative group members and the frequency of their collaborative contributions.

OSS environments like GitHub are collaborative environments [118, 58, 85] that allow not only professionals but also volunteers to contribute to software projects. In this research, we consider a collaborative group of developers as a completely connected graph of three or more developers based on their collaborations with each other. This graph can be seen as a network in which each node of the network represents unique developers in a project and the edge and direction between nodes represents the collaboration between developers in a project [90, 40]. Thus, collaborative groups are formed when developers in the group have collaborated with all others in the group. Since developers collaborate with a varied number of developers in a project [40], analyzing the diversity of developers in such a collaborative group will provide the overall relationship between their different characteristics.

[1] https://diversity.google/

[2] https://www.microsoft.com/en-us/diversity

[3] https://diversity.fb.com/read-report/

In GitHub, pull requests provide a mechanism for developers to contribute to OSS projects. For example, when developers want to contribute to a project, they create a pull request and solicit its acceptance. Other developers interested in the pull request could comment on the pull request. Then, a developer of the OSS project accepts or rejects the pull request based on various factors. Previous studies have considered pull requests as the primary source of contribution to a project [118, 81]. Middleton *et al.* [81] show that pull request is the most important project interaction that makes an outside developer to be accepted as a team member. Tsay *et al.* [118] study the influence of social and technical factors when contributing to OSS projects using the pull requests as the primary means of contribution. In this study, we use the frequency of collaborative pull requests merged to the projects, that is pull request in which atleast two members of the group have collaborated as proxy to measure the contributions of collaborative group members in GitHub.

Personality seeks to make predictions about what people do in given circumstances [23]. Existing personality tests exhibit both reliability [123, 53, 110] and validity [110, 53, 75]. In Software Engineering, the personalities of the software developers has been found to have a relationship with their social and technical activities in a project [91]. Rastogi and Nagappan [98] studied the relationship between the personality traits of about 400 active GitHub developers and their contributions. Their results showed that developers with different levels of contributions have different personality traits. Iyer *et al.* [58] analysed the influence of personality traits on the pull request evaluation process in GitHub. They found that pull requests from more open, more conscientious, and less extravert developers have a higher chance of approval. They also found that pull requests closed by more conscientious, more extravert, and more neurotic developers have a higher likelihood of acceptance.

Race is often related to phenotype/physical features, individual notions of self- defined identity, and racial ancestry [104]. Ethnicity is often related to cultural expressions (i.e., religion, beliefs, and customs) and identification [28]. Vasilescu *et al.*'s GitHub Survey [121] highlighted that GitHub developers can be aware of the race and ethnicity of other developers in a collaborative environment; and that 30% of GitHub developers have felt sometimes negative experiences due to diversity in terms of national origin, language, and ideology. Nadri *et al.* [85] analyzed whether there is evidence of bias based on perceived race in the written comments of rejected contributions in GitHub. Although they did not find any explicit racism in the written comments left by GitHub developers, their results indicate that there may be a bias against perceived non-white races and ethnicities. Moreover, their initial analysis leads them to believe that there may exist an unconscious bias against individual developers with ethnicity perceptible as Non-White [84].

This research adds to the existing human aspects research in software engineering by analyzing the relationship between the personality, race and ethnicity of the collaborative group members and the frequency of their collaborative contributions. We believe that this research provides valuable insights that could help foster a healthier OSS collaborative environment. For this research, we define our thesis statement as follows.

Thesis Statement: *Personality and race and ethnicity of collaborative group members will have a relationship with the frequency of their collaborative contributions in OSS environment.*

In this thesis, we conduct different experiments to verify if our thesis statement is true. We verify that personality of collaborative group members have a relationship with the frequency of their collaborative contributions by answering the following research questions.

- **RQ1: Does the personality of collaborative group members have a relationship with the frequency of their collaborative contributions?**
- **RQ2: Does diversity in personality of collaborative group members have a relationship with the frequency of their collaborative contributions?**
- **RQ3: Does proportion of developers with a specific personality trait in a group have a relationship with the frequency of their collaborative contributions?**

Furthermore, we also verify that race and ethnicity of collaborative group members have a relationship with the frequency of their collaborative contributions by answering the following research questions

- **RQ1: What is the distribution of collaborative group members races and ethnicities ?**
- **RQ2: Do homogeneous and heterogeneous collaborative groups with respect to race and ethnicity have a different distribution of their collaborative contributions ?**
- **RQ3: Does racial and ethnic diversity of collaborative group members have a relationship with their collaborative contributions?**

1.1 Contributions

In detail, our contributions from this thesis are as follows:

- We show that the personality of collaborative group members is related to the frequency of their collaborative contributions in GitHub.
- We show that diversity in personality of the collaborative group members is related to the frequency of their collaborative contributions in GitHub.
- We show that majority of developers of a particular personality trait in a group have a relationship with the frequency of their collaborative contributions in GitHub.
- We empirically observe that the collaborative groups with all or with a majority of White members are the most commonly found population in GitHub.
- We show that homogeneous and heterogeneous collaborative groups with respect to the race and ethnicity of their members have different distribution of collaborative contributions. Heterogeneous groups have a higher median contributions than homogeneous groups.
- Our results provide empirical evidence that having diversity of the race and ethnicity of members in a collaborative group is related to more collaborative contributions in a GitHub.

We believe that this research enhances the understanding of the effects of human aspects - personality, race and ethnicity with respect to group dynamics in open source environments.

1.2 Definitions

We define the most used terms in this research as follows.

Collaborations - *The interactions between developers in the pull requests/issues in the projects.*

Collaborative groups - *The groups that contains members who have collaborated with all other members in the group.*

Personality - *The Big Five Ocean Personality percentile scores that are inferred using the IBM Watson Personality Insights from the member's written communication.*

Diversity in Personality - *The variance in the members' personality scores in a collaborative group with respect to each dimension of the Big Five Ocean Personality trait.*

Race and Ethnicity - *The Most likely estimated race and ethnicity of the developer by the Name Prism tool across the U.S. Census Bureau's race and ethnicity categories using the member's name.*

Diversity in Race and Ethnicity - *The Blau Index measures of a collaborative group with respect to their members' race and ethnicity.*

Chapter 2

Background and Related Work

First we present the theories related to personality. Then we present how diversity can affect group dynamics, followed by the related work with respect to personality and race and ethnicity with group dynamics. Finally, we describe the relation of our work with the previous work done in similar fields.

2.1 Theories of Personality

Personality research theories have long attempted to identify exactly the personality traits of individuals in order to explain their differences in behaviour and to predict how these individuals would behave in a given situation. In the past, some theories have included a myriad of possible personality traits such as the list of 4,000 personality traits from Allport's theory [4] or the list of 16 personality traits from Cattell's theory [22]. The Jungian personality type received a lot of attention and was measured by Myers-Briggs Type Indicator (MBTI) questionnare [83], but the validity of MBTI have had many criticisms [76, 96]. However, in the recent years, researchers believe that there are five core personality traits and two well developed personality models have dominated the personality research in academia [9]. These models are the Five Factor Model (FFM) [29] and the Big Five Model [45, 46].

Both models agree that there are "Big Five" broad categories of personality traits and they are often seen as one model because their respective five factors are highly correlated [47]. However, they slightly disagree on their theoretical bases. For this study, we have used the Big Five Model as it relies on that personality traits are coded in natural language, and differences in personality may become apparent through language use [76]. Moreover, the Big Five Model has been empirically well validated, and it has shown good reliability [123], validity [75], and consistency across cultures [37, 14].

From the 10 aspects of the Big Five personality [35] and the previous personality research in GitHub [58], We adopt the following definitions of the Big Five personality traits to be used in through this paper:

- **Openness to Experience**: this trait describes an individual's degree of intellectual curiosity, creativity, and preference for experiencing a variety of activities. Some facets are imagination, artistic interests, adventurousness, intellect, and liberalism. While people who score high in Openness tend to be more adventurous and creative, people who score low are often more traditional and conventional, and may have difficulties with abstract thinking. Openness to Experience will be referred to as openness for the rest of the research.

- **Conscientiousness**: this trait describes an individual's tendency to show self-discipline and act in an organized or thoughtful way. Some facets are self-efficacy, orderliness, dutifulness, achievement-striving, and cautiousness. While high scores in Conscientiousness mean that people are careful and diligent, low scores mean that people are impulsive and disorganized.

- **Extraversion**: this trait describes an individual's tendency to seek stimulation in the company of others. Some facets are friendliness, assertiveness, activity level, excitement-seeking, cheerfulness. People who score high in this trait are outgoing and social, but people who score low tend to be reserved.

- **Agreeableness**: this trait describes an individual's tendency to be compassionate and cooperative toward others. Some facets are trust, cooperation, altruism, sympathy, modesty, and moralism. A person who score high in Agreeableness is friendly and optimistic while a person who score low is critical and can be seen as aggressive.

- **Neuroticism**: this trait describes an individual's degree of vulnerability to unpleasant emotions like anger, anxiety, or depression. Some facets are anxiety, anger, depression, self-consciousness, and vulnerability. People who score high in Neuroticism tend to be perfectionist during their everyday activities, and experience stress or negative emotions. People who score low in Neuroticism are easier to remain calm and are less affected by stressful events.

2.2 Big Five Personality Traits in Software Engineering

The Big Five personality traits can be derived from textual interactions [128]. There are text analytic programs that link personality traits with individuals' linguistic style [128]. Thus, researchers have used advancements in language processing to study the effects of personality traits in Software Engineering [100, 12, 98, 18, 58].

Previous studies have used the Linguistic Inquiry and Word Count (LIWC) dictionary [93] to infer the personality traits of software developers. The LIWC tool uses a psychometrically-based dictionary that has been validated by individual judges and used in various experiments by Pennebaker and others [1] to count the functional words that correlate with the personality traits. Martinez *et al.* [73] empirical studied different Software Engineering roles and found that architects and the testers had high Extraversion scores, while developers had low Conscientiousness scores. Rigby and Hassan [100] studied the personality of developers in the Apache httpd server mailing list and their findings suggest that the two top developers in the project had similar personalities between them, but these personalities were different from the rest of developers in the project. Bazelli *et al.* [12] performed a similar study in Stack Overflow that used developer-related data. They found that the top reputed users in Stack Overflow score higher in Extraversion when compared to medium and low reputed users. Licorish *et al.* [71] analyzed the personality traits of developers contributing to the IBM Jazz repository and found that top contributors score high on Openness, coders score high on Neuroticism, and practitioners score high on Extraversion. Kanij *et al.* [62] studied the personalities of software testers and found that testers scored high on Conscientiousness in comparison to other members in the team. Soomro *et al.* [111] investigated the association between a software professional's personality with team climate and team performance. This study is an identification of 35 primary studies of personality based on MBTI and FFM models of personalities, and the authors state that most of the studies used undergraduate students as subjects and as surrogates of software professionals. However, the authors points out that extraversion seems to be one of the key personalities in team formation in most of the studies. Moreover, Rastogi and Nagappan [98] studied the relationship between the personality traits of about 400 active GitHub developers and their contributions. Their results showed that developers with different levels of contributions have different personality traits. The top contributors to GitHub projects were significantly more neurotic than other contributors. Moreover, they found that the personality traits of the most active GitHub developers evolved as

[1] http://homepage.psy.utexas.edu/homepage/

more conscientious, more extravert, and less agreeable.

Recently, language processing tools use machine learning to outperform rule-based methods that use the LIWC dictionary. A recent work by Paruma *et al.* [91] used IBM Watson Personality Insights[2] to retrieve personality traits for clustering developers together. They found a relationship between the personality traits of developers and their social and technical activities on a project. Calefato *et al.* [18] also used IBM Watson Personality Insights to analyze developers' personality in Apache projects. They observed that developers became more conscientious, agreeable, and neurotic over time. Moreover, they also state that developers' personality traits do not vary with their role, membership, and extent of contribution to the projects. Iyer *et al.* [58] also used IBM Watson Personality Insights to study the influence of personality traits on the pull request evaluation process in GitHub. They found that pull requests from more open, more conscientious, and less extravert developers have a higher chance of approval. They also found that pull requests closed by more conscientious, more extravert, and more neurotic developers have a higher likelihood of acceptance.

2.3 Diversity Theory

The diversity of individual members in a group and the group outcomes has been largely studied in different areas of knowledge [55, 112]. Organization diversity has also been studied from different lenses [125]. Most of these studies have focused on off-line working groups and have reported contradictory results with respect to the relationship between group members diversity and group outcome. While some studies showed significant positive correlations between work group diversity and group's performance [57, 103, 39, 24], other studies reported that group diversity negatively impacts the group's outcomes [55, 124]. However, these results are understandable since there may be different group outcomes for different groups mainly depending on their priorities.

A possible explanation to these contradictory results lies in the following three theories: similarity-attraction (SA) theory [17], social identity and social categorization (SISC) theory [113], and information-processing (IP) theory [105]. SA and SISC theories may explain the negative effects of group diversity in the group's outcomes as these theories postulate that people tend to categorize other people into groups based on similarities such as culture, beliefs, or values. These theories explain that some people may prefer to work with others similar to them, which creates more homogeneous and less diverse work groups. IP

[2] www.ibm.com/watson/services/personality-insights

theory may explain the positive effects of work group diversity in work group's outcomes as it postulates that work groups formed by individuals from different cultural/educational backgrounds and ideas can provide the work group with access to broader information and enhanced problem solving skills [59]. IP theory helps create more heterogeneous and diverse work groups.

2.4 Personality in Group Dynamics

Previous studies have investigated the relationship between group personality traits and how they are related to the group's outcome under varying fields and environments. Kichuk and Weisner [66] studied the effects of Big Five personalities in product design teams among students and found that successful teams have high scores of Extraversion, Agreeableness, and Neuroticism. Neuman *et al.* [87] studied a team of retail assistants and showed evidences that team personality diversity has a positive effect on team performance. Halfhill *et al.* [50] performed an integrated review of empirical researches involving group personality composition across different fields and among the 31 studies that the author considered, Conscientiousness and Openness were the top most personalities that emerged as task predictors. Acuna *et al.* [1] analysed 35 student teams and found that teams with high job satisfaction tend to have members who score highly on Agreeableness and Conscientiousness. Gilley *et al.* [44] theorized that the diversity of personality within a team helps to produce a better team outcome. Xia *et al.* [127] also studied the influence of project manager in another dimension of personality based on the DISC model (Dominant, Influence, Steadiness, and Compliance) and found that projects teams with dominant managers, along with those with more influential members and less dominant members, have higher success scores in IT companies. In Summary, the researches with respective to personality and group outcomes are varied along different dimensions like fields of study, professionals/student groups etc, and most of the research have different personality relationship with respect to the outcomes under considerations.

2.5 Race and Ethnicities in Group Dynamics

Previous studies on racial and ethnic diversity in work groups yield mixed results on the relationship between racial and ethnic diversity, and group outcomes. Jackson *et al.* found poor team's performance in teams comprised by people with different race and ethnicities when compared to homogeneous teams [60]. Pieterse and van Eekelen show that racial diversity in student software development teams does not have an impact on how well the team members cooperated to develop software products [95]. Contrary to these results other studies indicated that racial and ethnically diverse groups lead to more creativity and innovation due to complementary and learning opportunities [2, 69, 90]. Gupta *et. al* found that organizations which are racial and ethnically diverse are likely to have a positive impact on sales, productivity, market share, and innovation [49]. These results show varied empirical results obtained in different environments and settings and most of them are based on off-line working groups.

The intention of OSS communities is to behave as a meritocracy where the quality of the contributions are the only factor to accept contributions [42]. In these OSS communities, demographic characteristics such as race and ethnicity would be less salient [101]. However, recent studies started to raise awareness about demographic diversity in OSS communities. Vasilescu *et al.* [122] showed that increasing gender and tenure diversity in the software project is associated with greater productivity. Catolino *et al.* [21] found that the presence of women in a project team generally reduces the amount of community smells. Daniel *et al.* [34] analyzed the relationship between diversity on community engagement and market success in of 357 projects hosted on SourceForge. They found a positive influence of diversity based on developers' role on market success and community engagement, but a negative influence of diversity of spoken language and nationality on community engagement. Ortu *et al.* [90] found that racial and gender diversity was associated with increased sales revenue and more customers. In Github, an initial analysis by Nadri et al [84] lead them to believe that there may exist an unconscious bias against individual developers with ethnicity perceptible as Non-White and their reinforce the need for further studies on ethnic diversity in software engineering to foster a healthier OSS community.

2.6 Our Work in Relation to Previous Studies

An important difference between our work and previous related work is that we study groups of developers that have collaborated with each other in open source software projects. Our groups are formed of only collaborative developers in a project and we measure their successful collaborative contributions.

Furthermore, the part one of our research does not study the personalities of individual developers of a project but the relationship between (i) collaborative group member's personality, (ii) the diversity of personalities within a collaborative group and (iii) majority of developers with a specific personality trait in a collaborative group, to the group members' contributions as a whole.

Also with regard to the racial and ethnic diversity in part two of this research, this study analyses collaborative group diversity from a racial and ethnic point of view in GitHub. Furthermore, this study looks at the (i) distribution of collaborative group members race and ethnicity (ii) the difference in the distribution of collaborative contributions from homogeneous and heterogeneous collaborative groups (iii) relationship between the collaborative group members' race and ethnicity , and their collaborative contributions.

Chapter 3

Data Gathering and Model Construction

In this chapter, we explain the procedures used for gathering data including GitHub project selection, collaborative groups identification from the projects, data filtration, inferring human aspects for the developers in the collaborative groups. We also present the details related to the model used in the research and also the common variables used in the model construction for the research.

3.1 GitHub Project Selection

We choose GitHub because it is the largest online platform for collaborative software development. Moreover, previous studies have modeled GitHub as a social network to discover different communities of developers in OSS [114] and to discover how the different communication channels within GitHub capture and share knowledge when developing software [115].

We select the dataset used to study the personality effects in GitHub projects by Iyer *et al.* [58]. We choose this dataset because, this would help us compare our personality results with the previous results and also ensures that personality can be sufficiently inferred from the texts of the developers as the authors have already inferred personality for the developers in their research. This dataset was curated as a combination of the $11,000$ projects considered by Tsay *et al.* [118] and $15,000$ additional projects from the RepoReaper website.[1] The RepoReaper is a public dataset that contains around one million non-trivial projects curated by Munaiah *et al.* [82]. From the 26,000 projects, Iyer *et al.* [58] only filtered those projects that were classified as non trivial, which had proper maintenance, were not forks, had at least more than three contributors, and with at least 250 closed/merged pull request. After this filtering, Iyer *et al.*'s final dataset contains 1,860 projects.

At the time of our study, some of the 1,860 projects from Iyer *et al.*'s final dataset were made private or were inaccessible through the GitHub API.[2] After removing these projects, our final dataset contains 1,804 projects for which we were able to successfully get the pull request details from the start of each project to the end of December 2018 using the GitHub API.

3.2 Collaborative Groups Identification

Github being a popular Open Source Software(OSS) collaborative environment allows contributions from different developers including volunteers as well [20, 118]. However, not all the developers within a project will collaborate with all other developers [40]. In this research, we analyze only developers who have tightly collaborated with each other. We form groups based on the collaborative efforts of the developers in a project using a three step process. First - we form a collaborative network graph based on the collaboration in pull request/issue comments to find collaborative developers in a project, second - we find tightly coupled collaborations from the collaborative network graph to ensure that developers are collaborative to each other as well and finally, third - we find completely connected components from these tightly coupled collaborations to form groups of developers that have collaborated with each other in a project. A detailed explanation of these steps and their reasoning is as follows.

Step 1 – Creation of the collaborative network graph: To identify different developers who have collaborated in a GitHub project we used the idea of *collaborative network graph based on pull request/issue comments* that are used in previous research [90, 40]. Network graphs are graphs that constitutes the developer information with respect to other developers depending on the task involved. In Prior works, Ortu et al [90] investigates gender and national diversity by constructing an issue colloboration graph based on issue comments. Similarly, Mezour et al [40] investigates the different team structures in social coding in Github Projects by building a pull-based network. Since, in our

[1] https://reporeapers.github.io/results/1.html

[2] https://developer.github.com/v3/

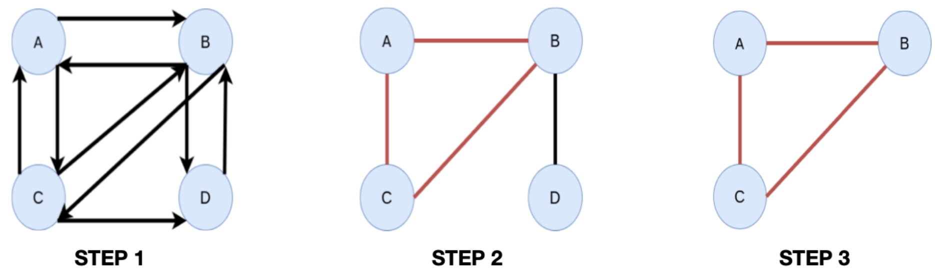


Figure 3.1: Steps to identify collaborative groups formed by developers A, B, and C inside the Project.

research we are interested in collaborative groups of developers, we build our collaborative network graph based on the requester and commenter of pull request/issue comments. For each pull request and issue report, we identified who requested the pull request/issue and who commented in it. With this information we built the collaborative network graph. Step 1 in Figure 3.1 illustrates the collaborative network graph of developers A, B, C, and D in a project. Each developer has an edge pointing to another developer if the developer has commented on a pull request/issue requested by the other developer. Step 1 in Figure 3.1 shows that developer C has collaborated with developer D, but developer D has not collaborated with developer C because there is no directed edge from developer D to developer C (e.g., developer C has commented on a pull request/issue by developer D but, developer D has not commented on a pull request/issue opened by developer C). This enables us to identify developers that have collaborated with other developers to be considered as a member for our collaborative groups.

Step 2 – Identification of tightly coupled collaborations: From our step 1, we have identified developers who have collaborated with other developers. However, these developers also include requesters who have only requested an artifact and have not commented on the other developer's request (For eg: developer D in step 1 of fig 3.1). Hence, in order to find developers who have strictly collaborated with each other, we identify the tightly and weakly coupled collaborations within the network graph. A tightly coupled collaboration exits only if the collaboration between two developers are both ways, i.e when both developers have collaborated with each other. This ensures that the developers can be included in a collaborated group since they have collaborated with each other. When there is only one collaboration (one directed edge) between two developers, it states that only one developer has collaborated with the other developer and not vice versa. Hence, we call it a weakly coupled interaction. In this research we donot consider the weakly coupled interaction because they donot provide sufficient evidence to prove that both the developers have collaborated with each other - that is interacted on each other entities. Step 2 in Figure 3.1 shows that developers A and B, B and C, and A and C are tightly collaborated couples. But, developers C and D are a weakly collaborated couple because only developer C has collaborated with developer D and, hence, we do not consider it for the next step. Filtering tightly interactive couples from this step ensures that the collaboration is bidirectional i.e both developers have collaborated with each other. Hence, we can safely consider them as a member for including in our collaborative group.

Step 3 – Identification of Collaborative groups: From step 2, we have filtered developers having tightly coupled collaborations to ensure that we include developers who have strictly collaborated with each other. In this step, we extract the members of the collaborative groups within a project by identifying the completely connected components or cliques of three or more developers. We did not consider a couple (two developers) as a group because, we are concerned about the collaboration of a member with multiple members and not a single member alone as done in previous research [58, 84]. A completely connected component or clique as they are generally called, are subgraphs of a graph where all the nodes are connected with each other. A clique from our collaborative network graph of tightly coupled collaborations would provide us with a group of developers who not only have collaborated with each other but as well as collaborated with all other members in the clique. This step ensures that we did not consider any sporadic developers for our groups since all members have collaborated with each other. We used the python module *networkx* by Daniel *et al.* [8] to identify the cliques as groups from the collaborative network graph of a project. Step 3 in Figure 3.1 illustrates that there is only one group in the project. This group is formed by developers A, B, and C because all of them have tightly collaborated with each others. From the 1804 projects in our dataset, following the three step process we were able to extract $30,950$ tightly coupled interactive groups having $26,811$ developers from $1,413$ projects. Note that, we consider developers within a project only and not across projects. One of the possible reasons for projects not having collaborative groups maybe because they used some other channels for communication like slack, discord or other collaboration channels other than GitHub. Figure 3.2 shows one of the real time collaborative group formed in a project 18F/tock[3] which is a time management tool.

[3] https://github.com/18F/tock

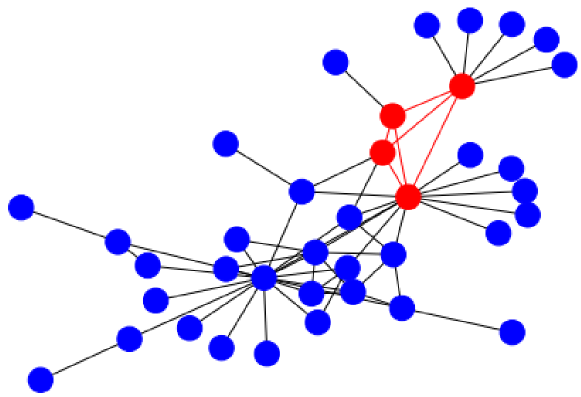


Figure 3.2: Example of a real group formed in project 18F/tock. Nodes represent the developers in the project and edges represent their tightly coupled interactions. Red nodes and edges indicate the developers in the group with their interactions. There may be other group in the graph which are not highlighted for better reference.

3.3 Data Filtration

From the 1,413 projects in our dataset, following the three step process we were able to extract 30,950 tightly coupled interactive groups having 26,811 developers. Previous research have stated that diversity in national origin and language may cause occasional confusion in communication over the use of idioms and misinterpreted emotion [121]. Inorder to facilitate sufficient communication, from the 26,811 developers we filtered 20,683 developers having atleast 1000 words in the comments for a project. This supports that we are only looking at developers who have actively communicated in the project and also provides sufficient texts for extracting their personality as explained in section 3.5.1. We formed 22,517 collaborative groups from 991 projects of the initial 30,950 groups having 18,580 developers from the 20,683 developers. Since, Github being an online collaborative platform, not all developers will be active through out the project and for a healthy project some rate of turnover is in fact desirable, as new developers brings in new abilities and ideas [122, 79]. Therefore, From the 22,517 groups, we gathered 7,905 collaborative groups having 5,487 unique developers that were active in the year 2018. We considered a group active only if the median number of pull request that is successfully merged by the members of the group is greater than 0. This ensures that the majority of the group members have successfully contributed to the project in 2018 by merging pull requests at least once. Also note that we only consider pull request if a member of the group has requested it and another member from the group has commented in it since we are measuring only collaborative contributions. Moreover, GitHub as a platform is evolving and changing continuously. Even the number of developers using GitHub has evolved from 24 million in 2017 to more than 37 million in 2018 [89]. Inorder for the time frame to be the same for all the groups, we considered the contribution in a time interval of the recent most year 2018 in our dataset. This also helps to compare our results with future studies. For conducting the personality analysis, our dataset had $7,905$ collaborative groups with $5,487$ unique members from 463 projects. The minimum number of members in group were 3, median number of members were 6 and maximum number of members in the group were 22.

For conducting our race and ethnicity analysis, We need to infer the race and ethnicity of the developers. From the $5,487$ developers using the Stanford Named Entity Recognizer and the Name-Prism tool as explained in section 3.5.2 we were able to infer the race and ethnicity of $4,174$ developers. Following Vasilescu *et al.* [122], we only considered those collaborative groups for which we were able to infer race and ethnicity for 75% of the group members. Therefore, our final dataset has $4,570$ collaborative groups made up of $4,137$ unique members from 377 projects. The minimum number of members in group were 3, median number of members were 6 and maximum number of members in the group were 22.

3.4 Data Insights

To provide further insights for our dataset, We also tried to infer country of the members from the location information provided in the developer's GitHub profile. Since the location is a structureless string, we extracted the country using the *geotext* tool [25]. Out of the $5,487$ unique developers in our personality analysis, we were able to infer the countries of $3,056$ developers. These developers were distributed among 76 countries with each country having a median 7 number of developers and a standard deviation of 129.417 with the top 5 countries have the most population being (US = 1114, Germany = 356, UK = 178, France = 113, Canada = 105). Also, Out of the $4,137$ unique developers from the project in our race and ethnicity analysis we were able to infer the countries of $2,486$ developers. These developers were distributed among 82 countries with each country having a median 8 number of developers and a standard deviation of 111.38 with the top 5 countries having the most population being (US = 928, Germany = 286, UK = 154, Canada = 85, France = 80). Moreover, the projects considered for personality analysis varied in 39 different languages with the most popular languages being Python (106 projects), C++ (57 projects), Java (55

Table 3.1: Personality related project details. Total = 463

Attributes	Minimum	Median	Maximum	Standard deviation
stars	3	1,202	79,348	9413.741
forks	2	442	28,844	2,975
project age	532	2413	3933	570
unique developers	3	7	321	20.95
groups	1	4	2742	130.87

Table 3.2: Race and Ethnicity related project details. Total = 377

Attributes	Minimum	Median	Maximum	Standard deviation
stars	5	1,442	79,348	10,020.238
forks	8	489	28,844	3,228.085
project age	532	2398	3933	565.217
unique developers	3	7	245	18.087
groups	1	4	1041	61.010

projects) and Ruby (51 projects). Also, the projects considered for the race and ethnicity analysis varied in 37 different languages with the most popular languages being Python (93 projects), C++ (47 projects), Java (43 projects) and Ruby (36 projects).

Table 3.1 provides the insights of the 463 projects that are used in the personality research and Table 3.2 provides the insights of the 377 projects that were used in the race and ethnicity analysis with their different statistics - star count, forks count, project age (in days from the creation date of the project to the point of our data collection), unique developers (the number of unique developers that are considered as part of a group in that project) and groups as the total number of collaborative groups formed.

3.5 Inferring Human Aspects

3.5.1 Inferring Personalities

We used the IBM Watson Personality Insights [38] to gather the personality information of the developers in the collaborative groups. The IBM Watson Personality Insights is an API based service where we can infer the personalities of the developers as percentiles based on raw texts. The service first tokenizes the input text to develop a representation in an n-dimensional space. The service uses an open-source word-embedding technique - Glove [94] to obtain a vector representation for the words in the input text. It then feeds this representation to a machine-learning algorithm that infers a personality profile with Big Five characteristics. To train the algorithm, the service uses scores from surveys that were conducted among thousands of users along with data from their Twitter feeds [38]. This method outperforms the rule based LIWC approach used in previous research [98, 93]. Furthermore, The IBM Watson Personality Insights has been recently used to predict personalities in a varied amount of work [61, 26, 67, 58]. We used the GitHub API to gather the comments posted by developers of a collaborative group from a project. We then used the same preprocessing technique as Iyer *et al.* [58] before inferring the personalities from the IBM Watson Personality API. These preprocessing techniques are the following:

- We only considered comments from developers whose language was English. We used the polyglot library [4] to infer the dominant language in the raw text. Although, the service provides personality in different languages, we wanted our scores to be consistent and hence only chose profiles whose communication was in English.
- We converted the markdown format comments to HTML code and removed the code contents by using their code tag.
- We lower cased all characters in the comments.
- We removed all the special characters except the punctuation marks from the comments.

IBM Watson Personality Insights recommends that the raw text should at least have 600 words for the results to be reliable. We only considered developers who had more than $1,000$ words across their comments for increased reliability. Hence, from the $26,811$ developers from our initial $30,950$ groups, we filtered $20,683$ developers having atleast 1000 words forming $22,517$ collaborative groups.

3.5.2 Inferring Race and Ethnicity

We used the given name of GitHub developers to infer their race and ethnicity. We first used the Stanford Named Entity Recognizer [43] with English language to classify GitHub

[4] https://pypi.org/project/polyglot/

developers' names as either person, organization, or location. We then selected only the names that were classified as a person's name and used the Name-Prism tool [129] to infer the race and ethnicity of that GitHub developers' names. The Name-Prism tool introduces name-embedding and utilizes the concept of homophily to create a name-based race and ethnicity classification tool. Name-embedding converts each name to a vector and recognizes contexts and similarity of names in the same context. Name-Prism is trained over 74 Million labeled names from 118 countries samples and classifies a given name into six different races and ethnicities with an F1 score of 0.795 [129]. Based on U.S. Census Bureau, Name-Prism uses six racial and ethnic groups: White, Black, Hispanic, API (Asian, Pacific Islander), AIAN (American Indian and Alaska Native), and 2PRACE (Mixed Race) to build the race and ethnicity classifier. Name-Prism produces a confidence rate between 0 and 1 for each racial and ethnic group. Finally, we assigned a unique race and ethnicity to each developer when the confidence rate obtained from Name-Prism tool was equal or higher than 0.6 for API developers, 0.7 for Hispanic developers, 0.8 for Black developers, and 0.8 for White developers. We chose different confidence intervals for each ethnic group after doing a manual sensitivity analysis of the Name-Prism tool. We randomly selected 40 GitHub developers from our dataset whose dominant race and ethnicity group was higher than 50%. For each developer, we analyzed all the information available in their GitHub profile (e.g., personal web, picture ...) to manually infer the developer's race and ethnicity with high confidence. We finally compare our manual results with the Name-Prism results and found that the tool correctly identifies the dominant race and ethnicity of each developer with different thresholds. When Name-Prism could not predict a race and ethnicity with a confidence higher than 0.6 for API developers, 0.7 for Hispanic developers, and 0.8 for Black and White developers, we classified this developer's race and ethnicity as "Unknown". Furthermore, the "Unknown" category also has developers without full names or with empty names in their GitHub profiles and developers for which the Stanford NER tool failed. The distributions of the race and ethnicity of the members are provided in Table 3.3.

3.5.3 Inferring Gender

We used gender diversity as a control variable since it has been shown as a positive predictor of productivity by Vasilescu *et al.* [122]. To infer the gender of GitHub developers we followed Vasilescu *et al.* [122] and used the gendercomputer tool[5]. This tool benefits if the country of the member is also provided. Since some names can refer to men in one

[5] https://github.com/tue-mdse/genderComputer

Table 3.3: Race and Ethnicity Distribution

	Population	Proportion(%)
White	3097	74.86
API	331	8.00
Hispanic	146	3.53
Black	4	0.10
AIAN	0	0
2PRACE	0	0
Unknown	559	13.51
Total	4137	100

Table 3.4: Gender Distribution

	Population	Proportion(%)
Man	3289	79.50
Woman	260	6.28
Gender-neutral	70	1.70
Unknown	518	12.52
Total	4137	100

country and women in another country, we gathered the country of the members from the location information provided in the developer's GitHub profile. Since the location is a structureless string, we extracted the country using the *geotext* tool [25]. If the country of a developer was successfully extracted, we used the country else we depend on the first name of the member to infer the gender. For developers whom we were not able to gather names, we included those names in the "Unknown" category. We only consider the gender if the Stanford NER recognized the string as the person's name as well. The distributions of the gender of the GitHub developer are provided in Table 3.4.

3.6 Model Construction

We build mixed-effects regression models using the generalized linear mixed-effects model function (glmer) available in the R package lme [11] as similar studies have done previously [58, 118]. Mixed-effects regression models are an extension of simple linear models to allow both fixed and random effects, and are particularly used when there is non independence in the data, such as arises from a hierarchical structure. Since we form numerous groups from the same projects, we produce a hierarchical structure where the projects include groups. Hence we use the project id as the random effects and the features mentioned in Table 4.1 as fixed effects. We consider the estimate of the fixed effects to be considered important if they were statistically significant ($\rho < 0.05$). Effect size is a measure of the strength of the relationship between variables [63]. We report the effect sizes of the variables obtained using the ANOVA statistical test [32]. Moreover, we also report the marginal and the conditional R-Squared values for the mixed effects model used. Marginal R-squared value represents the variation explained by the fixed-factors alone, while conditional R-Squared represents the variation explained by both fixed and random effects in the model [86]. We used the MuMIn R package to compute the marginal and the conditional RSquared values[6].

3.7 Model Variables

In this section, we describe the common variables that are used in the model constructions. Since independent variables are specific to each research question, their explanation is provided in their corresponding chapters along with the insights regarding the distribution of the variables.

3.7.1 Control Variables

We identified our control variables from previous studies related to pull request acceptance [118, 58, 122]. These can be grouped into group factors (i.e, gender diversity of a group, the group size, and the group age) and project factors (i.e., the project size and the project age). We measured the gender diversity of the members of a collaborative group using the Blau index [15] $Blau\_index = 1 - \sum_{i=1}^{R} {p_i}^2$. In this case, R represents three gender names categories (i.e, man, woman, gender-neutral) in the groups and p_i indicates the proportion of the members in that category to the total members in the collaborative group. Following Vasilescu *et al.* [122], for the group with members for whom we were not able to infer the gender, the measure only considers the fraction of the group for which we could infer the gender. We measured the collaborative group size as the number of

[6] https://CRAN.R-project.org/package=MuMIn

developers in a group. We measured the collaborative group age as the number of years each group has been active since they were formed until December 2018. We considered a collaborative group active if the median number of collaborative pull requests merged by the members of the group is greater than 0. This ensures that the group members were not dormant during a year. We include this feature because some of the developers might have prior interactions with each other which might increase their contributions [20]. We measured the project size as the size in Kilobytes of the project to which the group contributed during 2018. We measured the project age as the difference in days between December 31, 2018 and the first time the project was created. We also measured the popularity of the projects as the project star count that the project had at the point of our data collection. Tables [4.1,5.1] provides further insights regarding the distribution of the variables used in the models.

3.7.2 Dependent Variable

For our Dependent variable, we measured the contributions per collaborative group as the median number of collaborative pull requests merged by the developers in the group during 2018. Previous studies have considered pull requests as the primary source of contribution to a project [118, 81]. Middleton *et al.* [81] show that pull request is the most important project interaction that makes an outside developer to be accepted as a team member. Tsay *et al.* [118] study the influence of social and technical factors when contributing to OSS projects using the pull requests as the primary means of contribution. We used the GitHub API[7] to collect the pull request details. Note that we only considered pull requests requested by the members of the collaborative group if another member from the group have collaborated on it. This ensures that we measured the collaborative contributions of the developers and not the individual contributions. Previous studies related to group composition have used the mean, minimum, maximum and median as group measures [92, 13, 54, 72, 119]. Using the mean, minimum or maximum for a skewed distribution may not be an optimal solution, since the result would be skewed by the highest value considering we are measuring group contributions with a varied number of members in the group. Hence, we used the median measure to avoid skewed distributions in the number of pull requests merged by the developers in the collaborative groups.

[7] https://developer.github.com/v3/

Chapter 4

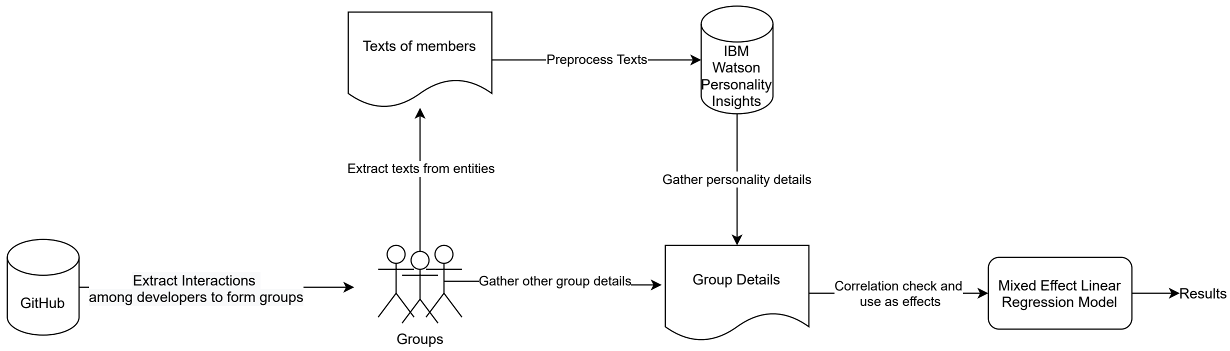
Personality Analysis

In this chapter we explain the methodology, results, discussions and threats of the part one of this research regarding the analysis of relationship between collaborative group members' personalities and the frequency of their collaborative contributions.

4.1 Data Modelling

We start this part of the research by extracting the collaboration groups from the projects [3.2], then we extract the texts from members of the groups to infer the personality traits using the IBM Watson Personality Insights [3.5.1]. After inferring the personalities we also collect other group details [3.7] and then finally we use mixed effects regression model to analyse each of our research questions. This overall process is explained in detail in chapter 3 and the figure 4.1 reminds the readers of the process used in this part of the research.

For the independent variables used in this part of the research, it is necessary to convert the individual personality scores of the developers in the group into group composition scores. For each member in the group, we infer the personality score along the five dimensions of personality (OCEAN) ranging from 0 to 1. Similar to our dependent variable [3.7.2], using the mean for a skewed distribution may not be an optimal solution, since the result would be skewed by the highest value. For example, if we consider a group with only one high neurotic member and all other low neurotic members, then the result from the mean would pose this group as being highly neurotic. Similarly, using the minimum or maximum value may result in the same case. Hence, to mitigate this effect we use the median personality scores of the members in a group to calculate the group's scores along the



.

Figure 4.1: Overview of Methodology used in the research analysis

five dimensions of personality (OCEAN). For example, to calculate the Openness score of a group we compute the median from the Openness scores of the group members. Moreover, the median has been used to represent group composition in previous personality related studies as well [54, 72, 119]. The group's scores along the five dimensions of personality is our primary independent variable in RQ1.

In a collaborative group, members with respect to personality differ from one another quantitatively and not qualitatively, that is, on a discrete attribute - the personality percentile score ranging from 0 to 1. In this research, we wanted to analyse how spread apart are the personality of members with respect to each of the five dimensions of the Big Five Ocean Personality traits. Variance and standard deviation are measurements of the spread between numbers in a dataset. Variance is measured as the average of the sum of the squared differences from the mean and standard deviation is measured as the square root of the variance. Although both variance and standard deviation can measure the spread of the population, variance provides a diversity score that emphasizes on the differences in the spread of the population than standard deviation. Also Variance has been used in previous studies to measure the variability of a group as well [10, 88, 106]. In this research, we use variance as a measure of diversity in personalities of the developers in a group for RQ2.

For RQ3, to measure the proportion of developers belonging to a specific personality trait, we categorize developers as being high if their personality score in that specific personality trait is greater than 0.75. Please note that the personality trait is a percentile score ranging from 0 to 1 that is calculated by the IBM Watson Personality Insights considering around 1 million twitter users [38] . Then we analyze collaborative groups having exactly one developer scoring high in a specific trait, collaborative groups having majority number of developers (more than half the developers but not all) scoring high in a specific trait, collaborative groups having minority number of developers (more than one but less than half developers) scoring high in a specific trait, and collaborative groups having all developers scoring high in a specific personality trait. Hence for RQ3, the primary independent variables are the collaborative groups categorized according to the number of developers having high scores in a specific personality trait. Table 4.1 provides insights on the distribution of the variables used in the model construction.

We examined the distribution of the data and found that the dependent variable (median number of merged collaborative pull requests by the group) was skewed. Thus, we normalized the dependent variable using the log transformation. Since some of our independent features (see Table 4.1) differ in range drastically, we used the default minmax scale() function provided by R [117] to scale them to the same range. We used the Spearman correlation test from variable clustering on the features to remove those ones that were highly correlated ($\rho = 0.7$). For the projects, we found that the star_count, watchers_count and forks_count were highly correlated, and hence we included only the star_count in our analysis. Any of the three features could have been removed and it wont impact the anlaysis. In RQ3, we found that groups having majority developers as highly open to be highly correlated with the group size, and hence we removed group size from our analysis. Also, in RQ3 we found that groups having exactly one developer with high Conscientiousness to be highly correlated with groups not even having one developer with high Conscientiousness and hence we removed the groups not even having one developer with high Conscientiousness from our analysis. We also used Variance Inflation Factor (VIF) to check for multicollinearity in the data. The VIF values start from 1 indicating higher values for higher multicollinearity. We did not find any values greater than 2 in our dataset [27] and hence we did not find any variables with multicollinearity.

4.2 Results

4.2.1 RQ1 : Does the personality of collaborative group members have a relationship with the frequency of their collaborative contributions?

Motivation: Some of the previous studies have looked upon the effects of personality of individual developers with respect to their roles in pull request acceptance [58]. Although individual member's personality effect is important, the personality traits of the collabo-

Table 4.1: Features used in RQs for model constructions. Notice that each observation on our dataset is a group

Feature	Used in	Description	Min	Median	Max	SD
		Independent Features				
		Developers' personality Features				
Median_Openness	RQ1	It represents the median of the developers openness in a group.	0.022	0.93	0.99	0.091
Median_Conscientiousness	RQ1	It represents the median of the developers conscientiousness in a group.	0.043	0.250	0.852	0.112
Median_Extraversion	RQ1	It represents the median of the developers extraversion in a group.	0.001	0.025	0.415	0.044
Median_Agreeableness	RQ1	It represents the median of the developers agreeableness in a group.	0.000	0.001	0.563	0.019
Median_neuroticism	RQ1	It represents the median of the developers neuroticism in a group.	0.002	0.402	0.988	0.129
Variance_Openness	RQ2	It represents the variance of the developers Openness in a group.	0	0.012	0.248	0.026
Variance_Conscientiousness	RQ2	It represents the variance of the developers conscientiousness in a group.	0	0.009	0.167	0.011
Variance_Extraversion	RQ2	It represents the variance of the developers extraversion in a group.	0	0.001	0.131	0.005
Variance_Agreeableness	RQ2	It represents the variance of the developers agreeableness in a group.	0	0	0.310	0.004
Variance_Neuroticism	RQ2	It represents the variance of the developers Neuroticism in a group.	0	0.022	0.286	0.022
All_high	RQ3	It represents the groups having all developers with high values for a specific personality trait.	-	-	-	-
Exactly_one	RQ3	It represents the groups having exactly one high developer for a specific personality trait.	-	-	-	-
Minority	RQ3	It represents the groups having minority of the developers having high scores in the specific personality trait.	-	-	-	-
Majority	RQ3	It represents the groups having majority of the developers having high scores in the specific personality trait.	-	-	-	-
		Group Features				
Group_size	All	It represents the number of members in the group.	3	7	22	5.6
Group_age	All	It represents the number of years the group were active.	1	3	8	1.275
		Project Features				
Project_size	All	It represents the size of the project at the point of data collection.	1686	235269	2330298	239937.6
Project_age	All	It represents the number of days from the project created date to the point of data collection.	532	2818	3933	581.925
Project_Star_Count	All	It represents the popularity the project.	3	17822	79348	18253.78
		Dependent Features				
Median pull request merged	All	Median number of collaborative pull request merged by the members of the group	0.5	11	106	17.99

Table 4.2: Mixed effects linear regression model : median of group's personalities Vs the median number of collaborative pull request merged. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Variables	Estimate	Std. Error	ρ sigf
Intercept	1.048	0.234	***
Median_Openness	−0.002	0.176	
Median_Conscientiousness	0.518	0.157	**
Median_Extraversion	−1.872	0.496	***
Median_Agreeableness	0.617	1.119	
Median_Neuroticism	0.144	0.125	
Group_Size	2.420	0.053	***
Group_Age	−0.540	0.059	***
Project_Star_Count	0.822	0.307	**
Project_Size	0.872	0.350	*
Project_Age	−0.394	0.232	
Marginal R2	0.413		
Conditional R2	0.697		
AIC	17991.95		
BIC	18082.63		

rative group members as a whole is important as well. Mezour et al [40] analysed different team structures that are formed in GitHub projects and states that there exists a number of developers who maintain repeated interactions in projects. We believe that understanding the collaborative group members personality would provide insights into fostering a healthier collaborative environment.

Approach: We used a mixed effects linear regression model similar to the mixed effects logistic regression model used in previous studies [58, 122]. The independent variable is the personality of group members which is computed using the median personality scores of the developers in the group. We calculated the median for each of the Big Five personality traits (i.e, Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) and used them as fixed effects. Other variables used in the model are provided in table 4.1.

Results: Table 4.2 presents the results of the mixed effects linear regression model. We found that while group member's median Conscientiousness has a statistically significant positive relationship (ρ $value$ < 0.001) to group's collaborative contributions with an estimate of 0.518, group member's median Extraversion has a statistically significant negative relationship (ρ *value* < 0.001) with an estimate of -1.872.

The effect sizes of the significant personalities are relatively small. Median Conscientiousness explains 0.441% and median Extraversion explains 0.781% of the data variance in the dependent variable. These results indicate that *group members' personality have a relationship with their collaborative contributions.*

> While Median Conscientiousness is positively related to collaborative group members' contributions, Median Extraversion is negatively related to collaborative group members' contributions.

4.2.2 RQ2: Does diversity in personality of collaborative group members have a relationship with the frequency of their collaborative contributions?

Motivation: There may be implicit variations among the developers' personality traits within a group that could have bigger effects on group's collaborative contributions. A group may have developers with different personalities that might complement each other and hence increase contributions or the personalities could actually hinder each other and decrease the collaborative contributions [17, 113, 105, 44]. In this RQ, we analyze the diversity of the developers' personality traits within a group to find whether they are related to the group members' collaborative contributions.

Approach: We used a mixed effects linear regression model as same as in RQ1. To measure the diversity of the personalities in a group, we used their variance. We calculated the variance of the group members personality for each of the Big Five personality traits (OCEAN) and used them as our independent variable. Other variables used in the model are provided in table 4.1.

Result: The results of the mixed effects linear regression model are provided in Table 4.3. Among the five personality dimensions, we found only three to be statistically significant. We found that diversity with respect to Conscientiousness is negatively related (ρ *value* < 0.001) to the group members' collaborative contributions with an estimate of -4.757. Moreover, we found that diversity with respect to Neuroticism is negatively related (ρ *value* < 0.05) to the group's collaborative contributions with an estimate of -0.994. Finally, variance in agreeableness is positively related ((ρ *value* < 0.05) to the group members' collaborative contributions with an estimate of 4.546.

Table 4.3: Mixed effects linear regression model : diversity of group personalities Vs the median number of collaborative pull request merged. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Variables	Estimate	Std. Error	ρ sigf
Intercept	1.258	0.133	***
Variance_Openness	0.453	0.431	
Variance_Conscientiousness	-4.757	0.940	***
Variance_Extraversion	-0.854	1.913	
Variance_Agreeableness	4.546	2.295	*
Variance_Neuroticism	-0.994	0.461	*
Group_Size	2.433	0.052	***
Group_Age	-0.567	0.058	***
Project_Size	1.029	0.352	**
Project_Age	-0.434	0.232	.
Marginal R2	0.416		
Conditional R2	0.703		
AIC	17970.53		
BIC	18061.21		

The effect sizes are very small for the diversity in personalities. Variance in Conscientiousness, Neuroticism and Agreeablenss has an effect size of 1.123%, 0.203% and 0.172% respectively.

These results show that *diversity in personality of collaborative group members does have a relationship with their collaborative contributions in GitHub.*

> Diversity with respect to Conscientiousness and Neuroticism is negatively related to group members' collaborative contributions and Diversity with respect to Agreeableness is positively related to group members' collaborative contributions.

Table 4.4: Distribution of groups in proportion of developers having high personality scores

Personality	Exactly one dev high	Minority	Majority	All high	Other groups
Openness	96	426	3744	3574	65
Conscientiousness	165	21	18	12	7689
Extraversion	7	0	0	0	7898
Agreeableness	2	0	0	0	7903
Neuroticism	1834	1272	61	62	4676

4.2.3 RQ3: Does proportion of developers with a specific personality trait in a group have a relationship with the frequency of their collaborative contributions?

Motivation: Some personality traits may be better suited for working with others in comparison with other personality traits [35]. For example, developers who score high in Extraversion tend to be more social and they could collaborate well in a group. We believe analyzing the relationship between the proportion of developers with a specific personality trait in a group and their collaborative contributions can provide useful insights into knowing whether having more or fewer members of a personality trait is beneficial.

Approach: We constructed five mixed-effects regression models for each of the Big Five personality traits. We found groups having exactly one developer, groups where majority of the developers, groups where minority of the developers and groups where all developers belong to one of the Big Five Personality traits. Table 4.4 shows the distribution of the majority of developers in a specific trait. From Table 4.4, it is interesting to note that 23% of the groups have exactly one developer with a high neurotic score. Moreover, there seems to be very few groups with high scores in Extraversion and Agreeableness. Other variables used in the model are provided in table 4.1.

Results: Out of all the personality traits Openness, Conscientiousness, and Neuroticism show statistically significant results. From Table 4.5 we can observe that groups having exactly one highly open developer and not even one highly open developer have a statistically significant negative relationship with group members' collaborative contributions. Also, groups having majority of the group members as highly open developers have a statistically significant positive relationship with group members' collaborative contri-

Table 4.5: Mixed effects linear regression model: Composition of proportion of groups with respect to Openness Vs the median number of collaborative pull request merged. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Variables	Estimate	Std. Error	P sigf
Intercept	1.181	0.141	***
Exactly_one_dev_high	-0.588	0.095	***
Minority	-0.084	0.049	.
Majority	0.031	0.026	***
Other_Groups	-0.711	0.170	***
Group_Age	0.157	0.061	*
Project_Size	1.083	0.373	**
Project_Age	-0.542	0.247	*
Project_Star_Count	1.073	0.328	**
Marginal R2	0.086		
Conditional R2	0.521		
AIC	19494.24		
BIC	19570.97		

butions with an estimate of 0.031. Similarly, from the model results in Table 4.6, we can observe that groups having majority of the developers as highly Conscientious have a positive relationship with groups' collaborative contribution with an estimate of 1.003. Finally, from the model results in Table 4.7, we can observe that groups having the majority of the developers as highly neurotic have a statistically significant positive relationship with groups' contributions with an estimate of 0.793.

With respect to Openness, the effect size of the groups with majority of people having high scores explains 93.10% of the variance in the dependent variable. With respect to Conscientiousness, the effect size of the groups with majority of people having high scores explains 57.96% of the variance in the dependent variable. With respect to Neuroticism, the effect size of the groups having majority of developers with high scores explains 6.19% of the variance in the dependent variable.

> Collaborative groups having majority of developers being highly open, conscientious, or neurotic have a positive relationship with group members' collaborative contributions.

Table 4.6: Mixed effects linear regression model: Composition of proportion of groups with respect to Conscientiousness Vs the median number of collaborative pull request merged. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Variables	Estimate	Std. Error	P sigf
Intercept	-0.7939	0.553	***
Exactly_one_dev_high	0.532	0.355	
Minority	0.423	0.367	
Majority	1.003	0.314	**
Group_Age	-0.073	0.55	
Project_Size	1.296	1.189	
Project_Age	1.461	0.838	*
Project_Star_Count	1.36	0.712	.
Marginal R2	0.098		
Conditional R2	0.581		
AIC	611.470		
BIC	641.848		

4.3 Discussions

We present our discussion with respect to each personality trait and their relationship with the collaborative contributions of the group.

Openness is a trait related to curiosity, creativity and preference for experiencing a variety of activities. Previous studies show that students and managers who score high on Openness tend to resolve conflicts through integration style which is a win-win situation [6]. The Github developers have a higher chance of getting their contributions accepted [58]; and they are between the top contributors in GitHub [98]. Not surprisingly, our results also indicate that groups having a majority of highly open developers are more likely to merge a higher number of collaborative pull requests. In RQ3 we found that having only one highly open developer in the group has a statistically negative relationship with the group members' collaborative contributions. But, this relationship is statistically positive when groups have majority developers as highly open developers. This may mean that highly open developers tend to work effectively with other highly open developers. We think that contributions from highly open GitHub developers are more likely to have more creative solution due to their personality. These solutions can be easily accepted or successfully discussed until integration.

Table 4.7: Mixed effects linear regression model: Composition of proportion of group with respect to Neuroticism Vs the median number of collaborative pull request merged. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Variables	Estimate	Std. Error	P sigf
Intercept	0.981	0.225	***
Exactly_one_dev_high	0.333	0.190	.
Minority	0.370	0.190	.
Majority	0.793	0.205	***
Other_Groups	0.127	0.191	
Group_Size	2.219	0.059	***
Group_Age	-0.482	0.058	***
Project_Size	0.886	0.351	*
Project_Age	-0.384	0.232	.
Project_Star_Count	0.837	0.309	**
Marginal R2	0.415		
Conditional R2	0.703		
AIC	17927.45		
BIC	18011.16		

Table 4.4 shows the population of groups having a majority of highly open developers. About 47% of the groups have a majority of highly open developers and 45% of the collaborative groups have all developers scoring high in Openness. This may indicate that a good proportion of collaborative groups in GitHub have a majority of highly open developers. We think that the reason for that is the relation between Openness and curiosity. GitHub developers can be motivated because they can contribute to millions of different projects and can learn different technologies.

Our results indicate that a good proportion of collaborative groups in GitHub have a majority of highly open developers. These groups have a statistically positive relationship with the frequency of group members' collaborative contributions.

Conscientiousness is a trait related to the desire to do a task in an efficient and organized way. Previous studies have found that software development student teams with high job satisfaction tend to have members with high Conscientiousness scores [1]. In GitHub, requesters and closers having high Conscientiousness have a higher chance of pull request acceptance [58]. Also, the top GitHub contributors scored high on Conscientiousness and contributors who participated frequently in the GitHub evolved as more conscientious over years [98]. Our results are consistent with previous results as we found that: the median Conscientiousness of a collaborative group has a statistically significant positive relationship with group members' collaborative contributions (RQ1) and collaborative groups having a majority of highly Conscientious developers have a statistically positive relationship with group members' collaborative contributions (RQ3). Since software development can be viewed as a combination of intricate components that demands reusability of code and modularity, it may demand a certain amount of orderliness and discipline. We think that contributions from conscientious GitHub developers are more likely to be more efficient and organized.

Although it is theorized that diversity in a group increases the group's effectiveness [44] and previous results states that the absolute difference between a closer and requester Conscientiousness of a pull request affects the pull request acceptance positively [58], our results from RQ2 indicate that less diverse group with respect to Conscientiousness have more group members' collaborative contributions. This may be due to the difference in the roles considered as we measure the collaborative contribution of developers as opposed to being a closer/requester of a pull request. We can only postulate that highly conscientious members may tend to be more achievement oriented and may not work well with relaxed people.

It is also interesting to note that Conscientiousness is the only personality among the

Big Five personality trait that has statistically significant results in all our research questions. This may indicate that Conscientiousness play a major role in the group members collaborative' contributions. The distribution of groups with highly conscientious members on GitHub from Table 4.4 indicates that among the 7,905 groups only 3% (216) have atleast one highly conscientious developer.

Our results indicate that Conscientiousness has a statistically positive relationship with the frequency of group members' collaborative contributions. However, only 3% of GitHub's groups have at least one highly conscientious developer.

Extraversion is a trait related to be outgoing, talkative, and energetic. Previous studies found both statistically significant positive and negative relationship with developer's contributions. Licorish *et al.* [71] found that among the developers contributing to IBM Jazz repository, practitioners scored high on Extraversion. Rastogi and Nagappan [98] found that top contributors scored high on Extraversion. Iyer *et al.* [58] stated that Extraversion has a statistically significant negative effect for a requester and a statistically significant positive effect for a closer of a pull request. Cullen *et al.* [33] studied 27 consultation projects in specific timelines and states that Extraversion has a statistically significant positive effect when there is no task conflict and has a statistically significant negative effect when there is a task conflict.

Surprisingly, our results showed that median Extraversion of a group has statistically significant negative relationship with group members' contributions (RQ1). We think that these results can be explained by previous studies in psychology that showed a correlation between Extraversion and dominance [78] and outlined it as an indication of being overly assertive as well [74]. In online collaborative environments, it may not be clear how facets of Extraversion manifests. Different from office settings, we speculate that highly extraverted GitHub developers can exhibit a behaviour that is not understood without additional clues. Recently, Kern *et al.* [65] found that top developers with high productivity score have noticeably lower Extraversion than average professionals. This finding provides evidence that the Extraversion trait may be different across software development context.

Even though Extraversion relates to being more social, from Table 4.4, we are able to see that among the 7,905 groups, there were almost no groups having more than one highly extravert developer.

Our results indicate that Extraversion has a statistically negative relationship with the frequency of group members' collaborative contributions.

Agreeableness is a trait that describes a person as compassionate and cooperative towards others. Some facets are trust, cooperation, altruism, sympathy, modesty, and moralism. Previous studies have found (1) a positive correlation for the propensity to trust with the number of pull request merged in a distributed software environment [19]; and (2) that top contributors scored low on Agreeableness [98]. However, other studies did not find any statistically significance between the Agreeableness and the acceptance of a pull request [58]; and they did not find any correlation between Agreeableness and team outcome [66]. Since Agreeableness is a positive trait we expected that it will have a positive influence on the group's contributions. However, we did not find any significant results for group members' median Agreeableness to be positively correlated with the group members' collaborative contributions.

From our results of RQ2, we found that groups that are more diverse in Agreeableness tend to have more successful contributions. However, considering that the collaborative groups in our dataset had very low Agreeableness scores (median = 0.001, sd = 0.019), we are not able to further research this result. We believe that most of the collaboration in GitHub were made by developers having low agreeable scores. This supports the previous results of Rastogi et al [98] where the author stated that top contributors scored low on Agreeableness. However, further research is warranted to conclude this statement.

> *Our results indicate that diversity with respect to Agreeableness is positively correlated with the frequency of group members' collaborative contributions.*

Neuroticism is a trait related to be perfectionist, and to experience stress or negative emotions. From table 4.4, we can see that 23% of the groups have only one neurotic person. It is also the highest in terms of groups having exactly one developer scoring high in a personality trait. Regarding the effects of Neuroticism, Iyer *et al.* [58] found a positive impact of closer's Neuroticism with pull request acceptance. Licorish *et al.* [71] studied the personality traits in the IBM Jazz repository and found that coders score high on Neuroticism. Rastogi and Nagappan [98] studied the personalities of 400 active GitHub contributors and found that the top contributors were statistically significantly more neurotic than other contributors. Our results are consistent with previous studies and show that groups having a majority of highly neurotic people have a statistically significant positive relationship with group members' collaborative contributions (RQ3). We also show that that diversity among the developers with respect to Neuroticism have a negative relationship with group members' collaborative contributions (RQ2). Shoss *et al.* [109] in their research related to others' oriented perfectionism, states that a perfectionist improve others' work by setting high standards for their work. We believe that this may be a reason that a group having highly neurotic developers performs better.

Our results indicate collaborative groups diverse in Neuroticism have a negative relationship with the frequency of group members' collaborative contribution. Also, groups having a majority of developers as highly neurotic people have a statistically significant positive relationship with the frequency of group members' collaborative contributions.

In summary, the personality traits of a collaborative group are related to the group members' collaborative contributions. The diversity and majority of developers for a specific personality trait in a group also matters when the outcome is to have a higher number of successful collaborative contributions. From our findings, we can observe that while the group members' median Conscientiousness has a positive relationship to the group members' collaborative contributions, the group members' median Extraversion has a negative relationship. These findings also indicate that it is more likely to have more collaborative contributions when groups are less diverse with respect to Conscientiousness and Neuroticism. Finally, with relation to the majority of developers for a personality trait in the group, we observe that groups formed by highly open developers, highly conscientious developers, and highly neurotic developers in GitHub groups are correlated to a higher number of collaborative contributions.

We also bring to notice that there may be other patterns of personality composition that could improve the successful contributions of a collaborative group. Our work provides sufficient evidence to suggest that personality plays a role in group dynamics where collaboration is concerned and personality needs to considered not only for individual developers but also for developers as a collaborative group. This work is an initial exploration of how the personality traits of online developers in a collaborative group are related to the group members' collaborative contributions.

4.4 Threats to Validity

We present our validity threats in terms of the four main threats in empirical software engineering research [126].

Construct Validity Other communication properties such as the strength of the relationship measured by the frequency of comments between two developers, or other social properties would have been useful as well. However, we primarily focused on features that represent a collaborative group than an individual, since our study is based on finding the best group composition with respect to personalities.

We only considered collaboration in the writtern comments between the developers. Thus, other types of interactions outside GitHub such as mailing lists or IRC channels are not considered in our study. The primary problem being unable to relate the GitHub account with the accounts on mailing lists or IRC channels, and of course collecting the data from these channels.

It is worth noting that there is no strict definition or simple metric that can measure the contributions of a group. We selected the frequency of collaborative pull request merged as previous studies have used that metric as a proxy to measure the contributions [118, 122] and also since we are analyzing the collaborative efforts of the developers. However, contributions can be different for different groups of developers.

Internal Validity It is possible that the personality traits we obtained from IBM Watson Personality Insights may not actually represent the true personality of the GitHub developers. To validate the service IBM collected survey responses and twitter feeds from 1500 to 2000 participants for different characteristics and languages [1]. To report their results IBM used mean absolute error - a metric that should be low as possible used to measure the difference between actual and predicted values , and correlation - a metric that should be high as possible is used to calculate the interdependence of two variables. In a scale of 0 to 1, the MAE was 0.12 and the correlation in a range of -1 to 1 was 0.31 for the actual and predicted values. These results are suggested as the benchmark results in some of the previous personality studies [108, 102, 80]. In spite of this, we are only concerned with the digital footprint or the perceived personality - that is how a developers personality is perceived by another developer. Moreover, self reported personalities scores also shows strong correlation with observers ratings as well [30]. Therefore, we do not claim this as an issue.

Another concern revolves around the fact that the actual content of GitHub comments is both technical and software engineering specific. We have mitigated this issue by removing the contents of the $< code >$ html tag from the comments and focusing on natural language only. Eventhough, the IBM Watson Personality Insights suggests that 600 words is sufficient to predict the personality, we have considered only developers having more than 1000 words for our study. Hence, we claim that the personality traits that we extracted are reasonably true given the wealth of GitHub comments used.

External Validity Although we analyzed projects hosted on the most popular social coding platform (GitHub), our findings cannot be directly generalized to other social coding

[1]https://cloud.ibm.com/docs/personality-insights?topic=personality-insights-science

platforms (e.g., GitLab). Another concern would be that the data we analysed may not be representative of the true population. However, our dataset comes from a combination of projects used by Tsay *et al.* [118] and Iyer *et al.* [58]. Moreover, we have provided a replication package to advance and gather further evidence in the field of personality traits in SE.

Conclusion Validity

In RQ3 , we consider developers having personality score over 0.75 as developers having high scores in a personality trait for better clarity and categorization. We used 0.75 because it represents the fourth quadrant in the personality scores provided by IBM Watson Personality Insights ranging from 0 to 1 that is compared among one million twitter users. Some others may feel that this threshold may not be the optimal solution. However, quadrants is a concept that is used way back from boxplots and we feel that it makes it easier to replicate and better understand the results.

We believe that researcher should actively look for more social features affecting the collaborative group's contributions on GitHub. Our approach uses only written comments to form groups within a project. Nonetheless, they allow us to infer the personality of developers.

Chapter 5

Race and Ethnicity Analysis

In this chapter we explain the methodology, results, discussions and threats of the part two of this research regarding the analysis of relationship between diversity of collaborative group members' race and ethnicity, and the frequency of their collaborative contributions in Github.

5.1 Data Modelling

The goal of this study is to assess to what extent racial and ethnic diversity of collaborative group members are related to their collaborative contributions in GitHub. We first selected projects from GitHub for our analysis, then we formed collaborative based on the developers interaction with each other 3.2. Following that, we inferred the race and ethnicity of the developers in the group using the Name-Prism tool and Stanford Named Entity Recognizer 3.5.2. Finally, we analyzed the relationship between the racial and ethnic diversity of collaborative group members and their collaborative contributions using the mixed effects regression model. This overall process is explained in detail in chapter 3 and the figure 5.1 reminds the readers of the process used in this part of the research.

Our main independent variable is the diversity in race and ethnicity of the group members in each collaborative group. In a collaborative group, members differ from one another qualitatively and not quantitatively, that is, on a categorical attribute - race and ethnic category. In such experiments, the diversity index should be a measure that indicates the variety of the population under consideration [52]. The Blau Index or Gini–Simpson Index [15] and Teachman's (entropy) index [116] are similar indexes that measures the variety

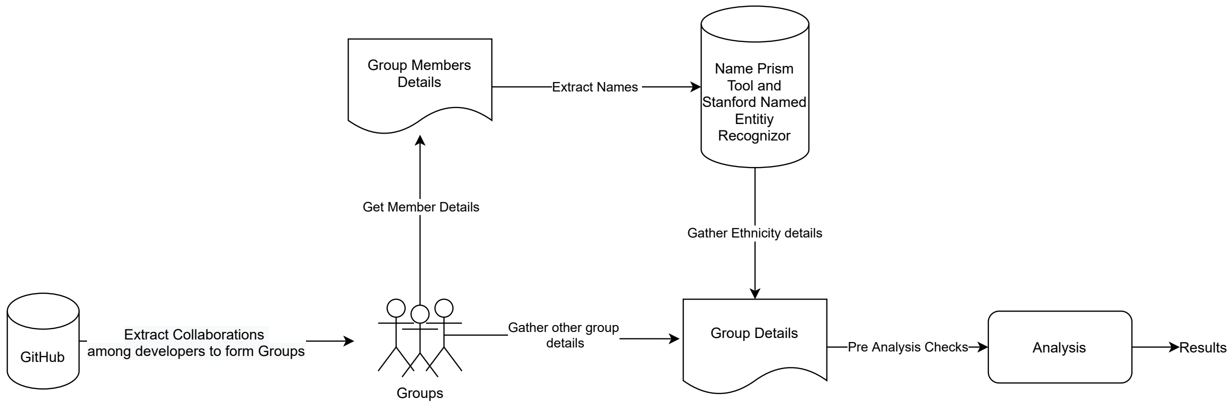


Figure 5.1: Overview of methodology used in the research analysis

of the population arising out of categorical variables. Both measures are highly correlated and the only difference is the ranges in which the diversity measures are inferred [52]. The Blau Index has a tidier range and is used in several diversity research concerning human characteristics as well. (e.g., [59, 3, 24, 122]). In this research, we use the Blau index to measure the diversity of the collaborative groups members with respect to their race and ethnicity. We calculated the racial and ethnic diversity of members of a collaborative group as $Blau\_index = 1 - \sum_{i=1}^{R} {p_i}^2$ where R represents the six races and ethnicities in the collaborative group and p_i indicates the proportion of the members in that category to the total members in the collaborative group. The higher the Blau index value, the more diverse are the member of the collaborative group and vice versa. Following Vasilescu *et al.* [122], for the collaborative groups with members for whom we were not able to infer the race and ethnicity, the measure only considers the fraction of the group for which we could infer the race and ethnicity. Table 4.1 provides insights on the distribution of the variables used in the model construction.

5.2 Results

5.2.1 RQ1: What is the distribution of collaborative group members races and ethnicities in GitHub ?

Motivation: Racial and ethnic diversity is a goal for fairness in software organizations. Currently, big tech companies have increased their efforts to increase racial and ethnic

Table 5.1: Features used in RQ3 for model constructions analysing race and ethnic diversity. Notice that each observation on our dataset is a group

Feature	Used in	Description	Min	Median	Max	SD
Independent Features						
Developers' Racial and Ehtnic Diversity Features						
Racial_and_Ethnic_Diversity	RQ3	It represents the racial and ethnic diversity of the group.	0	0	0.667	0.178
Group Features						
Group_gender_diversity	RQ3	It represents the gender diversity of the group.	0	0	0.667	0.174
Group_size	RQ3	It represents the number of members in the group.	3	6	22	4.86
Group_age	RQ3	It represents the number of years the group were active.	1	3	8	1.285
Project Features						
Project_size	RQ3	It represents the size of the project at the point of data collection.	1871	181012	2330298	230489.4
Project_age	RQ3	It represents the number of days from the project created date to the point of data collection.	532	2587	3933	615.6846
Project_Star_Count	RQ3	It represents the popularity the project.	5	12486	79348	17412.26
Dependent Features						
Median pull request merged	RQ3	Median number of collaborative pull request merged by the members of the group	0.5	9	75	13.61

diversity in their companies. For example, Microsoft[1] is committed to improve representation, especially for Black and African American, and Hispanic and Latin employees, across senior leadership roles. Google[2] expanded their recruiting efforts during 2019 to hire people from 15 Historically Black College & Universities, 39 Hispanic-Serving Institutions, and 9 women's colleges in the U.S. Linux Foundation [3] has committed to advance diversity and inclusion in their organization. Answering this research question, we aim to provide numbers and knowledge that help researchers and practitioners understand the racial and ethnic diversity of collaborative members in Github.

Approach: Although there are different ways to determine the racial and ethnic distribution of members in a group, we decided to calculate which races and ethnicities have the higher and lower number of members in a collaborative group. That way we can understand which racial and ethnicities are not proportionally represented in the collaborative groups.

[1]https://blogs.microsoft.com/blog/2020/10/21/microsofts-2020-diversity-inclusion-report-a-commitment-to-accelerate-progress-amidst-global-change/

[2]https://diversity.google/annual-report/

[3]https://www.linuxfoundation.org/press-release/2020/10/linux-foundation-focuses-on-science-and-research-to-advance-diversity-and-inclusion-in-software-engineering/

Table 5.2: Race and Ethnicity Composition of members in GitHub collaborative groups. Total groups : 4,570. API - Asian, Pacific Islander; AIAN - American Indian and Alaska Native, 2PRACE - Mixed Race

Race and Ethnicity Group	All Members with same Race and Ethnicity (All)	Majority of the members with same Race and Ethnicity (Majority)	Minority of the members with the same Race and Ethnicity (Minority)
White	1,239	2,992	286
API	6	31	1,397
Hispanic	23	24	687
Black	0	0	14
AIAN	0	0	0
2PRACE	0	0	0
Total	1,268	3,047	2,384

Hence, we classified the collaborative groups based on the number of members belonging to a particular racial and ethnic group. We classified the collaborative groups into three categories: 1) collaborative groups having all members with the same race and ethnicity; 2) collaborative groups having majority (more than half but not all the members) of the members with the same race and ethnicity; and 3) collaborative groups having minority (less than or equal to half the members) of the members with the same race and ethnicity. Note that the third category does not have mutually exclusive groups, since the minority of members of a group may be from different race and ethnicities.

Results: Table 5.2 shows that among the $4,570$ collaborative groups, 27.74% (1,268/4,570) of the collaborative groups are comprised of the developers with the same race and ethnicity and 66.67% (3,047/4,570) of the collaborative groups have a majority of developers belonging to the same race and ethnicity. This in itself shows the differences in the racial and ethnic composition among different collaborative groups.

Considering the individual races and ethnicities, from Table 5.2 we can infer that the majority of population is formed by White developers. 97.71% (1,239/1,268) of the collaborative groups where all members have the same race and ethnicity are formed by White developers. Also, among the $3,047$ collaborative groups with a majority population of developers with same race and ethnicity, White developers are the majority population of developers in 98.19% (2,992/3,047) of the collaborative groups. This clearly indicates that White developers are the most visible race and ethnicity within a collaborative groups when compared to other racial and ethnic groups.

We did not find any developer as AIAN or 2PRACE within the $4,570$ collaborative groups. Thus, no conclusion regarding AIAN or 2PRACE distribution in collaborative groups could be made. From the $1,268$ collaborative groups in which all members have the same race and ethnicity, only 6 collaborative groups (0.47%) have all members as API developers. Among the $4,570$ collaborative groups, the API race and ethnicity has the highest number of collaborative groups 30.56% (1,397) in which API developers are a minority.

From the $1,268$ where all members belong to the same race and ethnicity, 1.81% (23/1,268) of the collaborative groups are formed by Hispanic developers. Moreover, 0.78% (24/3,047) of the collaborative groups with the majority of members from the same race and ethnicity were formed by a majority of Hispanic developers. Considering the population of Black developers, only 0.3% (14/4,570) of the total collaborative groups have Black developers. All these 14 collaborative groups have a minority population of Black developers.

> Collaborative groups have different racial and ethnic distribution of members from one another. But, White developers form larger part of the population among collaborative groups.

5.2.2 RQ2: Do homogeneous and heterogeneous collaborative groups with respect to race and ethnicity have a different distribution of their collaborative contributions ?

Motivation: As previously stated in Section 2.3, the similarity-attraction theory and the social identity and social categorization theory explain that some people may prefer to work with others similar to them, which creates more homogeneous and less diverse groups. IP theory may explain the positive effects of work group diversity in group's outcomes as it postulates that groups formed by individuals from different cultural/educational backgrounds and ideas can provide the groups with access to broader information and enhanced problem solving. Although diversity can provide group with access to broader information and enhanced problem solving skills [59], they could also lead to difficulties in communication due to the difference in languages [121], among other problems. In this research question we try to ascertain whether homogeneous and heterogeneous collaborative groups with respect to race and ethnicity of their group members have statistically significant differences in the distribution of their collaborative contributions.

Approach: For this research question, we classify collaborative groups with all members belonging to the same race and ethnicity as homogeneous groups and collaborative groups having at least one member from a different racial and ethnic background as heterogeneous groups. Then, we performed a Mann-Whitney U-test (two sided), also known as Wilcoxon rank sum test [77], to find whether homogeneous and heterogeneous groups with respect to race and ethnicity have different distribution of collaborative contributions. The Mann-Whitney U-test is a widely used [56, 70, 107, 7] non parametric test for testing whether two samples are likely to be derived from the same population. We define our null hypothesis as follows.

- H0 - Members from homogeneous and heterogeneous groups with respect to race and ethnicity have similar contributions.

The Mann-Whitney U-test rejects the null hypothesis (H0) if ($\rho < 0.05$). If we can reject H0, it means that there is a statistically significant difference in the collaborative contributions among members from homogeneous and heterogeneous group.

Results: From the Mann-Whitney U-test , we were able to infer that (z = -23.741, ρ = $< 2.2e-16$) rejecting our null hypothesis (H0). The difference in location shift was 6.00, with a confidence interval of [5.5, 6.999]. Moreover, considering the distribution between the homogeneous (median = 4, sd = 7.395) and heterogeneous groups (median = 11.5, sd = 14.507) from Figure 5.2, heterogeneous groups have a higher median number of collaborative contributions. Hence, we conclude that homogeneous and heterogeneous groups, with respect to race and ethnicity, have different distribution of collaborative group members' contributions with heterogeneous groups having higher median number of group contributions than homogeneous groups.

> The difference between the collaborative contributions from members of homogeneous groups and the contributions from members of heterogeneous groups is statistically significant. Heterogeneous groups have higher median of collaborative contributions than homogeneous groups.

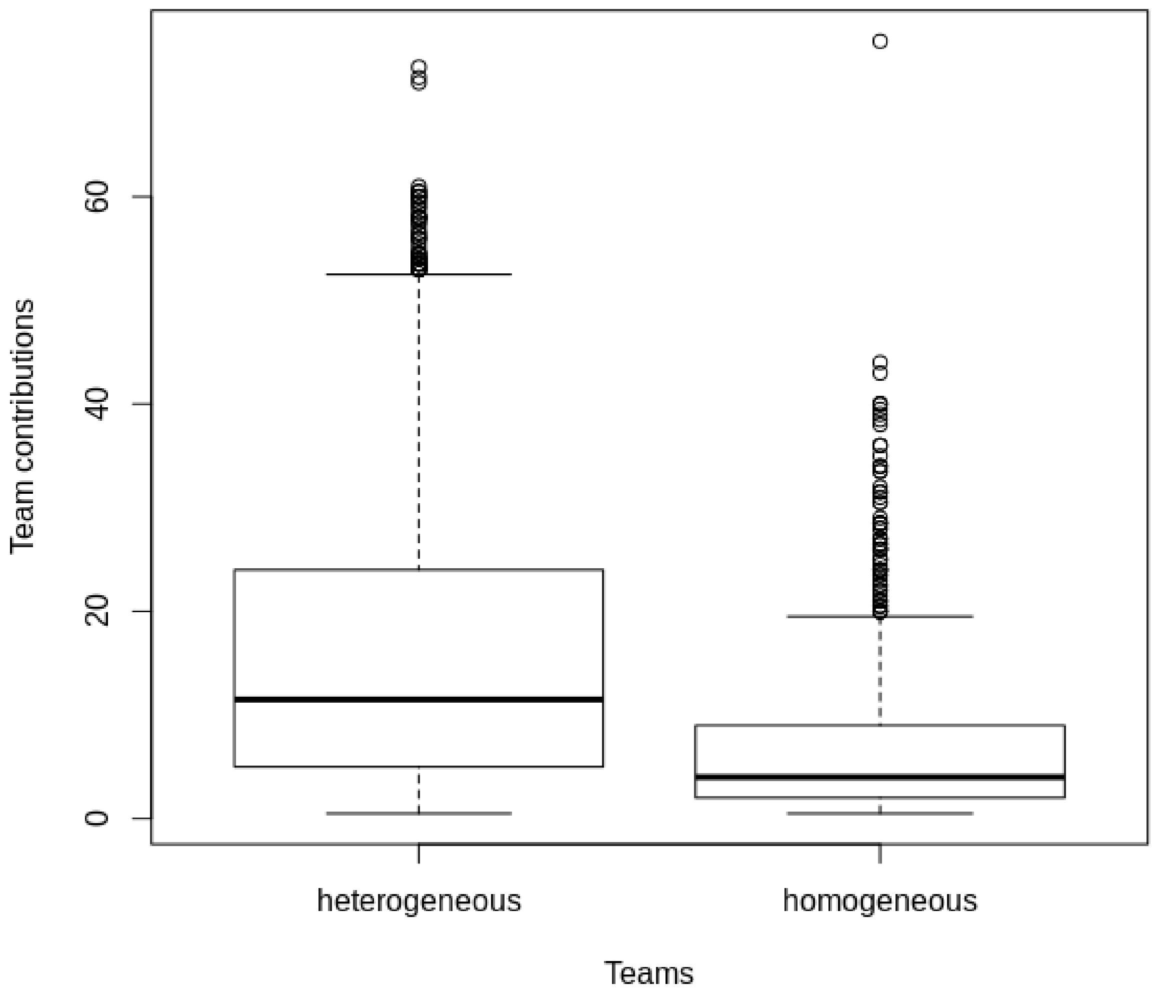


Figure 5.2: Homogeneous and heterogeneous groups with respect to race and ethnicity with their collaborative contributions.

5.2.3 RQ3 Does racial and ethnic diversity of collaborative group members have a relationship with their collaborative contributions?

Motivation: Some of the previous studies have showed that organizations that are racial and ethnically diverse have poor performance. [60]. Some other studies have showed that racial and ethnically diversity lead to more creativity and innovation due to complementary and learning opportunities [2, 69]. With this research question we aim to investigate whether the racial and ethnic diversity of the collaborative group members has a positive relationship with their collaborative contributions in GitHub. We believe that the answer to this question may provide beneficial insights for fostering a healthier collaborative environment with respect to race and ethnicity.

Approach: In order to find the relationship between racial and ethnic diversity, and group member's collaborative contributions, we used a mixed effects linear regression model as explained in Section 3.6. We used the racial and ethnic diversity of a collaborative group as the independent variable in the model. The other variables used in the model are provided in table 5.1.

Results: Table 5.3 indicate that racial and ethnic diversity in a collaborative group has a statistically significant positive relationship ($\rho\ value < 0.001$) with the group members collaborative contributions. It has an estimate of 0.224 indicating that for an increase in racial ethnic diversity, the group members' collaborative contribution increase by a factor of 0.224. The racial and ethnic diversity had a small effect size of 0.656%.

> Racial and ethnic diversity among the collaborative group members has a statistically positive relationship with the group members' collaborative contributions in GitHub.

5.3 Discussions

Collaborative groups have different racial and ethnic distribution of members from one another. But, White developers form larger part of the population among collaborative groups.

Table 5.3: Mixed effects linear regression model : racial and ethnic diversity of collaborative group members Vs the median number of collaborative pull request merged. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Variables	Estimate	Std. Error	ρ sigf
Intercept	1.183	0.152	***
Racial_and_Ethnic_diversity	0.224	0.076	**
Gender_diversity	0.231	0.082	**
Project_Size	1.185	0.412	**
Project_Age	-0.495	0.261	.
Project_Star_Count	0.740	0.333	*
Group_Size	2.517	0.070	***
Group_Age	-0.510	0.075	***
Marginal R2	0.360		
Conditional R2	0.684		
AIC	10,499.13		
BIC	10,563.4		

Previous studies regarding the relationship between the geographical location and developers contributions by Rastogi *et al.* [99] have shown that among the top seventeen countries contributing to GitHub, twelve are from North America or Europe where the majority population is white. Our results support Rastogi *et al.*'s results as we show that among the developers in the collaborative groups, a large population is formed by developers as White.

This in itself throws light upon a more serious issue that developers of other racial and ethnic groups may not be proportionally represented in GitHub communities. Especially, considering that we form collaborative groups based on the collaborative efforts of the developers. Table 3.3 indicates that only 11.63% of the developer population is made up of Black developers, Hispanic developers, and API developers. Moreover, we were not able to infer the race and ethnicity of 13.51% of the developers. This may be due to two reasons: 1) developers intentionally did not want to reveal their names; and 2) the tools that we used were not able to correctly identify their races and ethnicities. However, including the fact that no tool can be perfect, we are of the idea that if a tool cannot infer the race and ethnicity of the developers, then the developers as a global community will have trouble doing so as well. This result may indicate that developers do not feel confident about revealing their identities. Hence, we urge the GitHub community to work towards forming a more diverse and safe community with respect to racial and ethnic diversity.

The difference between the collaborative contributions from members of homogeneous groups and the contributions from members of heterogeneous groups is statistically significant. Heterogeneous groups have higher median of collaborative contributions than homogeneous groups.

Theoretically, racial and ethnic heterogeneous groups can have access to a broader range of experiences, perspectives, and knowledge information when compared to homogeneous groups. Lazear [68] stated that multicultural teams can create more communication conflicts and less cohesion due to language barriers. However, the multicultural background of the members of a team can contribute to improve the organization's innovation performance [68]. Van Knippenberg and Schippers [120] argued that heterogeneity also can bring more perspectives and more information, ideally spurring problem-solving and creativity [120].

Our results indicate that collaborative groups formed by members from more than one race and ethnicity (heterogeneous groups) have a statistically significant higher median of collaborative contributions than groups formed by members from the same race and ethnicity (homogeneous groups). We believe that this finding is of high importance as (1) it might indicate that the initial cultural barriers between members of heterogeneous groups in online platforms do not prevent them from having a high number of collaborative contributions; and (2) projects with a majority of heterogeneous groups are likely to have a higher frequency of collaborative contributions than projects with a majority of homogeneous groups. Therefore, it would be of high priority that GitHub communities can increase their efforts to support the creation of more heterogeneous work groups.

Racial and ethnic diversity among the collaborative group members has a statistically positive relationship with the group members' collaborative contributions in GitHub. Work group diversity can be desirable in OSS development because diversity can result in varied backgrounds and ideas, which provide the group with access to broader information, and enhanced problem solving skills [59]. On the other hand, due to greater perceived differences in values, norms, and communication styles, members in more diverse groups become more likely to engage in stereotyping, cliquishness, and conflict [55]. Results from Vasilescu *et al.* [121] show that differences in national origin and language may lead to difficulties in communication or "occasional confusion over the use of idioms and misinterpreted emotion" , causing "a lot of stress for the rest of the team". Brixy *et al.* [16] indicate that among startups, not all combinations of national origins matter for innovation, but only those that are associated with differences in cognitive approaches and knowledge. The results from RQ3 indicate that racial and ethnic diversity has a statistically positive relationship with group members' collaborative contributions. In accordance with the information-processing (IP) theory [105], we believe that SE demands varied problem solving skills and collaborative groups formed by individuals from different cultural/educational backgrounds can provide the group with access to broader information and enhanced problem solving skills [59]. Moreover, Hankerson *et al.* argues that technology artifacts could take up the characteristics of the creators and could potentially show race [51]. From our research, we also have evidences that racial and ethnically diversity in collaborative gorups are related to the frequency of group members' collaborative contributions. Thus, fostering racial and ethnic diversity in work groups is not only ethical but it would remove racial bias in the technology developed and might increase the number of successful contributions.

5.4 Threats to Validity

Construct Validity: We use tools like Name-Prism, Stanford Entity detector to infer the race and ethnicity. These tools could have false positives. However, as these tools have been used in previous research and we also take only the combined results from more than one tool, we do not consider it as an evident risk for this research. Moreover, we are interested in examining the race and ethnicity, and gender only in the same way one member could infer from a GitHub environment.

Internal validity: We have used the same models that were used in previous researches like the Blau index and the mixed effects regression models. Our dataset also comes directly from the GitHub API. Hence, we believe that we have provided sufficient evidence for our internal threats.

We identify the race and ethnicity of the developers based on their names using the Name Prism Tool [129]. This tool has been evaluated in previous studies [5] and has a F1 score of 0.795. Moreover, this tool has been trained over 74M labeled names from 118 countries samples and is also used in a variety of researches [36, 64, 97, 48].

Furthermore, there may be concerns regarding the validity of the names given by the developers in GitHub. However, we are only concerned about the perceived race and ethnicity, and gender of the developer, that is how one developer might perceive another developer's race and ethnicity, or gender and not the real race and ethnicity, or gender of the developer under question.

External Validity: There may be some concerns with the data not being representative of the true population. Our dataset comes from a previous research that was conducted to deduct personalities in GitHub environment [58]. The dataset is a mixture of two data sources from Tsay *et al.* [118] and a sample obtained from Repo Reaper [82]. We have analyzed a total of 4,570 collaborative groups filtered from 22518 groups for our results, which we think is a good representation of the population.

Conclusion Validity: We considered the racial and ethnicity primarily in the context of the U.S. Considering other races and ethnicities from the rest of the world may change our results. However, the races and ethnicities that we have considered are the most used races and ethnicities when referring to race and ethnic diversity in tech[4] [5] [6] and also in previous studies [84]. We hope that this research will nevertheless spark necessary conversations about race and ethnicity in SE beyond U.S. contexts.

[4] https://diversity.google/

[5] https://www.microsoft.com/en-us/diversity

[6] https://diversity.fb.com/read-report/

Chapter 6

Conclusion

We have conducted a large-scale empirical study on the relationship between collaborative group members' human aspects and the frequency of their collaborative contributions in GitHub. For analyzing the first human aspect - personality, collaborative group members personality traits were extracted from the textual comments of developers in GitHub and modeled on the Big Five personality using the IBM Personality Insights service. We started our study by identifying the median personalities of collaborative groups that may have a relationship with their collaborative contributions. We then looked into the diversity of these personalities that may be related to the the group members collaborative contributions. We finally took our research further by finding the relationship between the proportion of developers with respect to a specific personality trait in a collaborative group and the frequency of their collaborative contribution. We observe a statistically significant relationship between Openness, Conscientiousness, Extraversion, and Neuroticism of a collaborative group and their collaborative contributions. The detailed results are presented in Section 4.2 and discussed in Section 4.3.

For analyzing the second human aspect - race and ethnicity, We use the Name Prism tool and the Stanford Named Entity Recognizer to find the race and ethnicity of the group members. Then we use the mixed effects regression model to find the relationship between racial and ethnic diversity of the group members and the group members' collaborative contributions. Our study finds that 1) collaborative groups have different group members distribution with respect to race and ethnicity and a large part of the population are comprised by White developers; 2) Homogeneous and Heterogeneous collaborative groups with respect to race and ethnicity have different distribution of group members' contributions with heterogeneous groups having higher median number of collaborative contributions than homogeneous groups;. and 3) Racial and ethnic diversity among the group members have a statistically positive relationship with group members' collaborative contributions in GitHub.

We believe that our work enhances the understanding of the effects of human aspects - personality, race and ethnicity in open source environments by analyzing the relationship between a group of collaborative developers in open source environment and the frequency of their collaborative contributions [31]. However, there may be more than one group composition that could provide a better outcome. Hence, further research is required to find those compositions. For example, the influence of a member with one specific personality on another member with another specific personality on the Big Five personality traits can be studied. This could provide insights on the influence of group members more deeply. Furthermore, other communication channels like slack, discord can be seen as a source to collect collaborative data. Also with regard to our second part of research, further research could identify the barriers that prevent different racial and ethnic developers from participating in GitHub. Also, a qualitative survey could be done to further explain the reasons behind the unbalanced population of developers with respect to their race and ethnicity. This could lead to developing of tools that could allow us to develop and enrich a racial and ethnically diverse online community.

Replication package: We are sharing our dataset in the interest of encouraging others to replicate and build upon our work. For now, the data can be found here: https://bit.ly/3ejwocZ

References

[1] Silvia T Acuña, Marta Gómez, and Natalia Juristo. How do personality, team processes and task characteristics relate to job satisfaction and software quality? *Information and Software Technology*, 51(3):627–639, 2009.

[2] Alberto Alesina and Eliana La Ferrara. Ethnic diversity and economic performance. *Journal of economic literature*, 43(3):762–800, 2005.

[3] Muhammad Ali, Carol T Kulik, and Isabel Metz. The gender diversity–performance relationship in services and manufacturing organizations. *The International Journal of Human Resource Management*, 22(07):1464–1485, 2011.

[4] Gordon W Allport and Henry S Odbert. Trait-names: A psycho-lexical study. *Psychological monographs*, 47(1):1–171, 1936.

[5] Bedoor K AlShebli, Talal Rahwan, and Wei Lee Woon. The preeminence of ethnic diversity in scientific collaboration. *Nature communications*, 9(1):1–10, 2018.

[6] David Antonioni. Relationship between the big five personality factors and conflict management styles. *International journal of conflict management*, 9(4):336–355, 1998.

[7] A. Arcuri and L. Briand. A practical guide for using statistical tests to assess randomized algorithms in software engineering. In *2011 33rd International Conference on Software Engineering (ICSE)*, pages 1–10, 2011.

[8] Daniel A. Schult Aric A. Hagberg and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[9] Murray R Barrick, Michael K Mount, and Timothy A Judge. Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and assessment*, 9(1-2):9–30, 2001.

[10] Murray R Barrick, Greg L Stewart, Mitchell J Neubert, and Michael K Mount. Relating member ability and personality to work-team processes and team effectiveness. *Journal of applied psychology*, 83(3):377—-391, 1998.

[11] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

[12] Blerina Bazelli, Abram Hindle, and Eleni Stroulia. On the personality traits of stackoverflow users. In *2013 IEEE international conference on software maintenance*, pages 460–463, New York City, 2013. IEEE.

[13] Suzanne T Bell. Deep-level composition variables as predictors of team performance: a meta-analysis. *Journal of applied psychology*, 92(3):595—-615, 2007.

[14] Verónica Benet-Martínez and Oliver P John. Los cinco grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the big five in spanish and english. *Journal of personality and social psychology*, 75(3):729–750, 1998.

[15] Peter Michael Blau. *Inequality and heterogeneity: A primitive theory of social structure*, volume 7. Free Press New York, 1977.

[16] Udo Brixy, Stephan Brunow, and Anna D'Ambrosio. The unlikely encounter: Is ethnic diversity in start-ups associated with innovation? *Research Policy*, 49(4):103950, 2020.

[17] Donn Erwin Byrne. *The attraction paradigm*, volume 462. Academic Press, 1971.

[18] Fabio Calefato, Giuseppe Iaffaldano, Filippo Lanubile, and Bogdan Vasilescu. On developers' personality in large-scale distributed projects: the case of the apache ecosystem. In *2018 IEEE/ACM 13th International Conference on Global Software Engineering (ICGSE)*, pages 87–96, New York City, 2018. IEEE.

[19] Fabio Calefato, Filippo Lanubile, and Nicole Novielli. A preliminary analysis on the effects of propensity to trust in distributed software development. In *2017 IEEE 12th international conference on global software engineering (ICGSE)*, pages 56–60, New York City, 2017. IEEE.

[20] Casey Casalnuovo, Bogdan Vasilescu, Premkumar Devanbu, and Vladimir Filkov. Developer onboarding in github: the role of prior social links and language experience. In *Proceedings of the 2015 10th joint meeting on foundations of software engineering*, pages 817–828, 2015.

[21] Gemma Catolino, Fabio Palomba, Damian A Tamburri, Alexander Serebrenik, and Filomena Ferrucci. Gender diversity and women in software teams: How do they affect community smells? In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, pages 11–20. IEEE, 2019.

[22] Raymond B Cattell. The main personality factors in questionnaire, self-estimate material. *The Journal of Social Psychology*, 31(1):3–38, 1950.

[23] Raymond B Cattell. Personality: A systematic theoretical and factual study. 1950.

[24] Jilin Chen, Yuqing Ren, and John Riedl. The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 821–830, 2010.

[25] Xiaoyu Chen, Dongming Wang, and Ting Zhao. Geotext: an intelligent dynamic geometry textbook. *ACM Communications in Computer Algebra*, 46(3/4):171–175, 2013.

[26] S. Chowdhury, C. Walter, and R. Gamble. Toward increasing collaboration awareness in software engineering teams. In *2018 IEEE Frontiers in Education Conference (FIE)*, pages 1–9, New York City, 2018. IEEE.

[27] Jacob Cohen, Patricia Cohen, Stephen G West, and Leona S Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge, New York, 2013.

[28] Stephen Cornell and Douglas Hartmann. *Ethnicity and race: Making identities in a changing world*. Sage Publications, 2006.

[29] Paul T Costa and Robert R McCrae. *The NEO personality inventory*. Psychological Assessment Resources Odessa, 1985.

[30] Paul T Costa Jr and Robert R McCrae. *The Revised NEO Personality Inventory (NEO-PI-R)*. Sage Publications, Inc, 2008.

[31] Kevin Crowston, Qing Li, Kangning Wei, U Yeliz Eseryel, and James Howison. Self-organization of teams for free/libre open source software development. *Information and software technology*, 49(6):564–575, 2007.

[32] Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman. An anova test for functional data. *Computational statistics & data analysis*, 47(1):111–122, 2004.

[33] Kristin L Cullen-Lester, Hannes Leroy, Alexandra Gerbasi, and Lisa Nishii. Energy's role in the extraversion (dis) advantage: How energy ties and task conflict help clarify the relationship between extraversion and proactive performance. *Journal of Organizational Behavior*, 37(7):1003–1022, 2016.

[34] Sherae Daniel, Ritu Agarwal, and Katherine J Stewart. The effects of diversity in global, distributed collectives: A study of open source project success. *Information Systems Research*, 24(2):312–333, 2013.

[35] Colin G DeYoung, Lena C Quilty, and Jordan B Peterson. Between facets and domains: 10 aspects of the big five. *Journal of personality and social psychology*, 93(5):880, 2007.

[36] Rebecca Diamond, Tim McQuade, and Franklin Qian. The effects of rent control expansion on tenants, landlords, and inequality: Evidence from san francisco. *American Economic Review*, 109(9):3365–94, 2019.

[37] John M Digman. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440, 1990.

[38] IBM Cloud Docs. Ibm personality insights, 2020.

[39] Christopher P Earley and Elaine Mosakowski. Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management journal*, 43(1):26–49, 2000.

[40] Mariam El Mezouar, Feng Zhang, and Ying Zou. An empirical study on the teams structures in social coding using github projects. *Empirical Software Engineering*, 24(6):3790–3823, 2019.

[41] Robert Feldt, Lefteris Angelis, Richard Torkar, and Maria Samuelsson. Links between the personalities, views and attitudes of software engineers. *Information and Software Technology*, 52(6):611–624, 2010.

[42] Joseph Feller and Brian Fitzgerald. A framework analysis of the open source software development paradigm. In *ICIS 2000 proceedings of the twenty first international conference on information systems*, pages 58–69. Association for Information Systems (AIS), 2000.

[43] Jenny Rose Finkel, Trond Grenager, and Christopher D Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 363–370, 2005.

[44] Jerry W Gilley, M Lane Morris, Alina M Waite, Tabitha Coates, and Abigail Veliquette. Integrated theoretical model for building effective teams. *Advances in Developing Human Resources*, 12(1):7–28, 2010.

[45] Lewis R Goldberg. An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216, 1990.

[46] Lewis R. Goldberg. The structure of phenotypic personality traits. *American psychologist*, 48(1):26, 1993.

[47] Lewis R Goldberg, John A Johnson, Herbert W Eber, Robert Hogan, Michael C Ashton, C Robert Cloninger, and Harrison G Gough. The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality*, 40(1):84–96, 2006.

[48] Will Gornall and Ilya A Strebulaev. Gender, race, and entrepreneurship: A randomized field experiment on venture capitalists and angels. *Available at SSRN 3301982*, 2019.

[49] Richa Gupta. Workforce diversity and organizational performance. *International Journal of Business and Management Invention*, 2(6):36–41, 2013.

[50] Terry Halfhill, Eric Sundstrom, Jessica Lahner, Wilma Calderone, and Tjai M Nielsen. Group personality composition and group effectiveness: An integrative review of empirical research. *Small group research*, 36(1):83–105, 2005.

[51] David Hankerson, Andrea R Marshall, Jennifer Booker, Houda El Mimouni, Imani Walker, and Jennifer A Rode. Does technology have race? In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 473–486, 2016.

[52] David A Harrison and Katherine J Klein. What's the difference? diversity constructs as separation, variety, or disparity in organizations. *Academy of management review*, 32(4):1199–1228, 2007.

[53] Ong Choon Hee. Validity and reliability of the big five personality traits scale in malaysia. *International Journal of Innovation and Applied Studies*, 5(4):309–315, 2014.

[54] L Richard Hoffman. Similarity of personality: a basis for interpersonal attraction? *Sociometry*, 21(4):300–308, 1958.

[55] Sujin K Horwitz and Irwin B Horwitz. The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of management*, 33(6):987–1015, 2007.

[56] Wen-Hao David Huang, Denice Ward Hood, and Sun Joo Yoo. Gender divide and acceptance of collaborative web 2.0 applications for learning in higher education. *The Internet and Higher Education*, 16:57–65, 2013.

[57] Vivian Hunt, Dennis Layton, and Sara Prince. Diversity matters. *McKinsey & Company*, 1(1):15–29, 2015.

[58] Rahul N Iyer, S Alex Yun, Meiyappan Nagappan, and Jesse Hoey. Effects of personality traits on pull request acceptance. *IEEE Transactions on Software Engineering*, pages 1–12, 2019.

[59] Susan E Jackson and Aparna Joshi. Diversity in social context: a multi-attribute, multilevel analysis of team diversity and sales performance. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 25(6):675–702, 2004.

[60] Susan E Jackson, Aparna Joshi, and Niclas L Erhardt. Recent research on team and organizational diversity: Swot analysis and implications. *Journal of management*, 29(6):801–830, 2003.

[61] Madhura Jayaratne and Buddhi Jayatilleke. Predicting personality using answers to open-ended interview questions. *IEEE Access*, 8:115345–115355, 2020.

[62] Tanjila Kanij, Robert Merkel, and John Grundy. An empirical investigation of personality traits of software testers. In *2015 IEEE/ACM 8th International Workshop on Cooperative and Human Aspects of Software Engineering*, pages 1–7, New York City, 2015. IEEE.

[63] Ken Kelley and Kristopher J Preacher. On effect size. *Psychological methods*, 17(2):137, 2012.

[64] Elisabeth Kempf and Margarita Tsoutsoura. Partisan professionals: Evidence from credit rating analysts. Technical report, National Bureau of Economic Research, 2018.

[65] Margaret L Kern, Paul X McCarthy, Deepanjan Chakrabarty, and Marian-Andrei Rizoiu. Social media-predicted personality traits and values can help match people to their ideal jobs. *Proceedings of the National Academy of Sciences*, 116(52):26459–26464, 2019.

[66] Susan L Kichuk and Willi H Wiesner. The big five personality factors and team performance: implications for selecting successful product design teams. *Journal of Engineering and Technology management*, 14(3-4):195–221, 1997.

[67] Madan Krishnamurthy, Khalid Mahmood, and Pawel Marcinek. A hybrid statistical and semantic model for identification of mental health and behavioral disorders using social network analysis. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1019–1026, New York City, 2016. IEEE.

[68] Edward P Lazear. Globalisation and the market for team-mates. *The Economic Journal*, 109(454):15–40, 1999.

[69] Neil Lee and Max Nathan. Does cultural diversity help innovation in cities: evidence from london firms. Technical report, London School of Economics and Political Science, LSE Library, 2011.

[70] Nale Lehmann-Willenbrock, Joseph A Allen, and Annika L Meinecke. Observing culture: Differences in us-american and german team meeting behaviors. *Group Processes & Intergroup Relations*, 17(2):252–271, 2014.

[71] Sherlock A Licorish and Stephen G MacDonell. Personality profiles of global software developers. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, pages 1–10, New York, NY, 2014. Association for Computing Machinery.

[72] Gardner Lindzey and James A Urdan. Personality and social choice. *Sociometry*, 17(1):47–63, 1954.

[73] Luis G Martínez, Antonio Rodríguez-Díaz, Guillermo Licea, and Juan R Castro. Big five patterns for software engineering roles using an anfis learning approach with ramset. In *Mexican International Conference on Artificial Intelligence*, pages 428–439, New York City, 2010. Springer.

[74] Robert R McCrae and Paul T Costa. Updating norman's" adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of personality and social psychology*, 49(3):710–721, 1985.

[75] Robert R McCrae and Paul T Costa. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, 52(1):81–90, 1987.

[76] Robert R McCrae and Paul T Costa Jr. Reinterpreting the myers-briggs type indicator from the perspective of the five-factor model of personality. *Journal of personality*, 57(1):17–40, 1989.

[77] Patrick E McKnight and Julius Najab. Mann-whitney u test. *The Corsini encyclopedia of psychology*, pages 1–1, 2010.

[78] Albert Mehrabian. Analysis of the big-five personality factors in terms of the pad temperament model. *Australian journal of Psychology*, 48(2):86–92, 1996.

[79] Claudia de O Melo, Daniela S Cruzes, Fabio Kon, and Reidar Conradi. Interpretative case studies on agile team productivity and management. *Information and Software Technology*, 55(2):412–427, 2013.

[80] Gregory J Meyer, Stephen E Finn, Lorraine D Eyde, Gary G Kay, Kevin L Moreland, Robert R Dies, Elena J Eisman, Tom W Kubiszyn, and Geoffrey M Reed. Psychological testing and psychological assessment: A review of evidence and issues. *American psychologist*, 56(2):128, 2001.

[81] Justin Middleton, Emerson Murphy-Hill, Demetrius Green, Adam Meade, Roger Mayer, David White, and Steve McDonald. Which contributions predict whether developers are accepted into github teams. In *2018 IEEE/ACM 15th International Conference on Mining Software Repositories (MSR)*, pages 403–413, New York City, 2018. IEEE.

[82] Nuthan Munaiah, Steven Kroh, Craig Cabrey, and Meiyappan Nagappan. Curating github for engineered software projects. *Empirical Software Engineering*, 22(6):3219–3253, 2017.

[83] Isabel Briggs Myers, Mary H McCaulley, and Robert Most. *Manual, a guide to the development and use of the Myers-Briggs type indicator.* consulting psychologists press, 1985.

[84] Reza Nadri. On the relationship between the developer's perceptible ethnicity and the evaluation of contributions in github, 2020.

[85] Reza Nadri, Gema Rodriguez Perez, and Meiyappan Nagappan. Insights into non-merged pull requests in github: Is there evidence of bias based on perceptible race? *IEEE Software*, 2020.

[86] Shinichi Nakagawa and Holger Schielzeth. A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in ecology and evolution*, 4(2):133–142, 2013.

[87] George A Neuman, Stephen H Wagner, and Neil D Christiansen. The relationship between work-team personality composition and the job performance of teams. *Group & Organization Management*, 24(1):28–45, 1999.

[88] George A Neuman and Julie Wright. Team effectiveness: beyond skills and cognitive ability. *Journal of Applied psychology*, 84(3):376—-389, 1999.

[89] octoverse. Github population details, 2018.

[90] Marco Ortu, Giuseppe Destefanis, Steve Counsell, Stephen Swift, Roberto Tonelli, and Michele Marchesi. How diverse is your team? investigating gender and nationality diversity in github teams. *Journal of Software Engineering Research and Development*, 5(1):1–18, 2017.

[91] Oscar Hernán Paruma Pabón, Fabio A González, Jairo Aponte, Jorge E Camargo, and Felipe Restrepo-Calle. Finding relationships between socio-technical aspects and personality traits by mining developer e-mails. In *2016 IEEE/ACM Cooperative and Human Aspects of Software Engineering (CHASE)*, pages 8–14, New York City, 2016. IEEE.

[92] Miranda AG Peeters, Harrie FJM Van Tuijl, Christel G Rutte, and Isabelle MMJ Reymen. Personality and team performance: a meta-analysis. *European Journal of Personality: Published for the European Association of Personality Psychology*, 20(5):377–396, 2006.

[93] James W Pennebaker, Ryan Boyd, Kayla Jordan, and Kate Blackburn. The development and psychometric properties of liwc2015, 2015.

[94] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, 2014. Association for Computational Linguistics.

[95] V Pieterse and MCJD van Eekelen. Cultural diversity and the performance of student software engineering teams. In *Kabanda, S.(ed.), SACLA 2018: 47th Annual Conference of the Southern African Computing Lecturers' Association Krystal Beach Hotel, Gordon's Bay Western Cape, South Africa, June 18-20, 2018*, pages 120–134. Cape Town: Southern African Computer Lecturers' Association, 2018.

[96] David J Pittenger. Cautionary comments regarding the myers-briggs type indicator. *Consulting Psychology Journal: Practice and Research*, 57(3):210, 2005.

[97] Daniel Preoţiuc-Pietro and Lyle Ungar. User-level race and ethnicity predictors from twitter text. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1534–1545, 2018.

[98] Ayushi Rastogi and Nachiappan Nagappan. On the personality traits of github contributors. In *2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE)*, pages 77–86, New York City, 2016. IEEE.

[99] Ayushi Rastogi, Nachiappan Nagappan, Georgios Gousios, and André van der Hoek. Relationship between geographical location and evaluation of developer contributions in github. In *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, pages 1–8, 2018.

[100] Peter C Rigby and Ahmed E Hassan. What can OSS mailing lists tell us? a preliminary psychometric text analysis of the apache developer mailing list. In *Fourth International Workshop on Mining Software Repositories (MSR'07: ICSE Workshops 2007)*, pages 23–23, New York City, 2007. IEEE.

[101] Christine M Riordan and Lynn McFarlane Shore. Demographic diversity and employee attitudes: An empirical examination of relational demography within work units. *Journal of applied psychology*, 82(3):342, 1997.

[102] Brent W Roberts, Nathan R Kuncel, Rebecca Shiner, Avshalom Caspi, and Lewis R Goldberg. The power of personality: The comparative validity of personality traits,

socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological science*, 2(4):313–345, 2007.

[103] Joel Ross, Lilly Irani, M Six Silberman, Andrew Zaldivar, and Bill Tomlinson. Who are the crowdworkers? shifting demographics in mechanical turk. In *CHI'10 extended abstracts on Human factors in computing systems*, pages 2863–2872. Association for Computing Machinery, New York, NY, USA, 2010.

[104] Wendy D Roth. The multiple dimensions of race. *Ethnic and Racial Studies*, 39(8):1310–1338, 2016.

[105] Gerald R Salancik and Jeffrey Pfeffer. A social information processing approach to job attitudes and task design. *Administrative science quarterly*, pages 224–253, 1978.

[106] RD Savage. Ii.—personality factors and academic performance. *British Journal of Educational Psychology*, 32(P3):251–253, 1962.

[107] Christian Schulte, Temo Barwari, Abhishek Joshi, Konstantinos Theofilatos, Anna Zampetaki, Javier Barallobre-Barreiro, Bhawana Singh, Nils A Sörensen, Johannes T Neumann, Tanja Zeller, et al. Comparative analysis of circulating noncoding rnas versus protein biomarkers in the detection of myocardial injury. *Circulation research*, 125(3):328–340, 2019.

[108] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, 8(9):1–16, 2013.

[109] Mindy K Shoss, Kori Callison, and LA Witt. The effects of other-oriented perfectionism and conscientiousness on helping at work. *Applied Psychology*, 64(1):233–251, 2015.

[110] Deborah R Smith and William E Snell Jr. Goldberg's bipolar measure of the big-five personality dimensions: reliability and validity. *European Journal of Personality*, 10(4):283–299, 1996.

[111] Arjumand Bano Soomro, Norsaremah Salleh, Emilia Mendes, John Grundy, Giles Burch, and Azlin Nordin. The effect of software engineers' personality traits on team climate and performance: A systematic literature review. *Information and Software Technology*, 73:52–65, 2016.

[112] Günter K Stahl, Martha L Maznevski, Andreas Voigt, and Karsten Jonsen. Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of international business studies*, 41(4):690–709, 2010.

[113] Henri Tajfel. Social psychology of intergroup relations. *Annual review of psychology*, 33(1):1–39, 1982.

[114] Damian A Tamburri, Fabio Palomba, Alexander Serebrenik, and Andy Zaidman. Discovering community patterns in open-source: a systematic approach and its evaluation. *Empirical Software Engineering*, 24(3):1369–1417, 2019.

[115] Jirateep Tantisuwankul, Yusuf Sulistyo Nugroho, Raula Gaikovina Kula, Hideaki Hata, Arnon Rungsawang, Pattara Leelaprute, and Kenichi Matsumoto. A topological analysis of communication channels for knowledge sharing in contemporary github projects. *Journal of Systems and Software*, 158:110416–11028, 2019.

[116] Jay D Teachman. Analysis of population diversity: Measures of qualitative variation. *Sociological Methods & Research*, 8(3):341–362, 1980.

[117] R Core Team et al. R: A language and environment for statistical computing, 2013.

[118] Jason Tsay, Laura Dabbish, and James Herbsleb. Influence of social and technical factors for evaluating contribution in github. In *Proceedings of the 36th international conference on Software engineering*, pages 356–366, New York City, 2014. Association for Computing Machinery.

[119] Bruce W Tuckman and Dominick S Orefice. Personality structure, instructional outcomes, and instructional preferences. *Interchange*, 4(1):43–48, 1973.

[120] Daan Van Knippenberg and Michaela C Schippers. Work group diversity. *Annu. Rev. Psychol.*, 58:515–541, 2007.

[121] Bogdan Vasilescu, Vladimir Filkov, and Alexander Serebrenik. Perceptions of diversity on git hub: A user survey. In *2015 IEEE/ACM 8th International Workshop on Cooperative and Human Aspects of Software Engineering*, pages 50–56. IEEE, 2015.

[122] Bogdan Vasilescu, Daryl Posnett, Baishakhi Ray, Mark GJ van den Brand, Alexander Serebrenik, Premkumar Devanbu, and Vladimir Filkov. Gender and tenure diversity in github teams. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 3789–3798, New York City, 2015. Association for Computing Machinery.

[123] Chockalingam Viswesvaran and Deniz S Ones. Measurement error in "big five factors" personality assessment: Reliability generalization across studies and measures. *Educational and Psychological Measurement*, 60(2):224–235, 2000.

[124] Warren E Watson, Kamalesh Kumar, and Larry K Michaelsen. Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of management journal*, 36(3):590–602, 1993.

[125] O'REILLYCA WILLIAMSKY. Demographyand diversityinorganizations: Are-viewof40yearsof research. *ResearchinOrganizationalBehavior*, 20(3):77–140, 1998.

[126] Claes Wohlin, Per Runeson, Martin Höst, Magnus C Ohlsson, Björn Regnell, and Anders Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, Berlin, 2012.

[127] Xin Xia, David Lo, Lingfeng Bao, Abhishek Sharma, and Shanping Li. Personality and project success: Insights from a large-scale study with professionals. In *2017 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pages 318–328, New York City, 2017. IEEE.

[128] Tal Yarkoni. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3):363–373, 2010.

[129] Junting Ye, Shuchu Han, Yifan Hu, Baris Coskun, Meizhu Liu, Hong Qin, and Steven Skiena. Nationality classification using name embeddings. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1897–1906, 2017.