

# Understanding the Sustainability of Online Software Development Groups in Bangladesh

by

S. M. Nafiur Rahman  
17241001

A thesis submitted to the Department of Computer Science and Engineering  
in partial fulfillment of the requirements for the degree of  
B.Sc. in Computer Science

Department of Computer Science and Engineering  
Brac University  
April 2020

© 2020. Brac University  
All rights reserved.

## Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

**Student's Full Name & Signature:**



---

S. M. Nafiur Rahman  
17241001

# Approval

The thesis titled “Understanding the Sustainability of Online Software Development Groups in Bangladesh” submitted by

1. S. M. Nafiur Rahman (17241001)

Of Spring, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on April 9, 2020.

## Examining Committee:

Supervisor:  
(Member)

Hossain Arif

---

Hossain Arif  
Assistant Professor  
Department of Computer Science and Engineering  
Brac University

Program Coordinator:  
(Member)

*Rabiul*

---

Md. Golam Rabiul Alam  
Associate Professor  
Department of Computer Science and Engineering  
Brac University

Head of Department:  
(Chair)

**Prof. Mahbub Majumdar**  
**Chairperson**  
**Dept. of Computer Science & Engineering**  
**Brac University**

---

Mahbubul Alam Majumdar  
Professor  
Department of Computer Science and Engineering  
Brac University

## Abstract

The field of software development currently is one of the most rapidly expanding sectors in terms of research as well as establishing a career. An immense amount of youths are entering it as new opportunities are being discovered and developments are being made on a daily basis. However, they are not receiving the best from the communities, because many of them are slowly stagnating. In this work, an endeavor was undertaken to determine the sustainability of online software development groups in Bangladesh. Loosely based on the OSS newcomer barrier model, a sustainability factor model was constructed. Six Facebook groups were selected and post data were collected to quantitatively represent member interaction by assigning reaction units. Based on the amount of posts against that of the reaction units, correlation was assessed, which yielded a Pearson coefficient value of 0.4753 with a P-Value of .062795 which is significant at  $p < .10$ , and a Spearman coefficient value of 0.471 with a 2-tailed p value of 0.06582 which is statistically non-significant. The coefficient values were then checked using R. The values implied a weak yet positive correlation between the amount of posts and reactions. Combined with the model, which revealed that 59.24% are support-related posts (among which 93.69% are being answered), the coefficient values enabled us to reach the decision that online software developer groups of Bangladesh are, as a whole, currently sustainable.

**Keywords:** Developer community; Sustainability; Support group; Correlation coefficient; Social media

# Acknowledgement

All praise to Allah for whom this work has been successful.

Thanks to my thesis supervisor, Hossain Arif sir, for his continuous support. And to Dilruba Showkat ma'am, who was the reason this work started.

# Table of Contents

|   |           |
|---|-----------|
| Declaration   | i         |
| Approval  | ii        |
| Ethics Statement  | iii       |
| Abstract  | iii       |
| Dedication  | iv        |
| Acknowledgment  | iv        |
| Table of Contents                                       | v         |
| List of Figures   | vii       |
| List of Tables  | viii      |
| Nomenclature  | ix        |
| <b>1 Introduction</b>                                   | <b>1</b>  |
| 1.1 Overview . . . . .                                  | 1         |
| 1.2 Objective . . . . .                                 | 2         |
| <b>2 Related Work</b>                                   | <b>3</b>  |
| <b>3 Methodology</b>                                    | <b>6</b>  |
| 3.1 Target Group Selection . . . . .                    | 6         |
| 3.2 Sustainability Factor Model . . . . .               | 8         |
| 3.3 Correlation Coefficient . . . . .                   | 9         |
| 3.3.1 Pearson’s Correlation Coefficient . . . . .       | 9         |
| 3.3.2 Spearman’s Rank Correlation Coefficient . . . . . | 9         |
| <b>4 Implementation</b>                                 | <b>11</b> |
| 4.1 Data Collection . . . . .                           | 11        |
| 4.1.1 Graph API . . . . .                               | 11        |
| 4.1.2 Headless Browser . . . . .                        | 12        |
| 4.1.3 Dataset . . . . .                                 | 14        |
| 4.2 Model Construction . . . . .                        | 15        |
| 4.3 Coefficient Calculation . . . . .                   | 19        |
| 4.3.1 Pearson’s Correlation Coefficient . . . . .       | 19        |

|          |   |           |
|----------|---|-----------|
| 4.3.2    | Spearman's Rank Correlation Coefficient . . . . . | 21        |
| 4.3.3    | Cross-Checking with R . . . . .                   | 21        |
| <b>5</b> | <b>Result Analysis</b>                            | <b>24</b> |
| <b>6</b> | <b>Conclusion</b>                                 | <b>26</b> |
| 6.1      | Limitations . . . . .                             | 26        |
| 6.2      | Future Work . . . . .                             | 27        |
|          | <b>Bibliography</b>                               | <b>29</b> |
|          | <b>Appendix</b>                                   | <b>30</b> |

# List of Figures

|      |  |    |
|------|--|----|
| 3.1  | Screenshots of selected groups . . . . .                                     | 7  |
| 4.1  | cURL command format for Graph API . . . . .                                  | 11 |
| 4.2  | Sample JSON response from Graph API . . . . .                                | 12 |
| 4.3  | NickJS code snippet for scraping group post data . . . . .                   | 13 |
| 4.4  | Various scraping features of Phantombuster service . . . . .                 | 13 |
| 4.5  | Collected data in JSON format . . . . .                                      | 14 |
| 4.6  | CSV file opened as a spreadsheet . . . . .                                   | 14 |
| 4.7  | Flowchart of working plan . . . . .  | 15 |
| 4.8  | A model showing the proportions of different post types . . . . .            | 18 |
| 4.9  | Comparison of amount of help wanted and received . . . . .                   | 19 |
| 4.10 | Calculated values for Pearson's correlation coefficient . . . . .            | 20 |
| 4.11 | Calculated values for Spearman's rank correlation coefficient . . . . .      | 22 |
| 4.12 | R language code to generate scatter plot and calculate coefficient . . . . . | 22 |
| 4.13 | QQplot of variable x . . . . .   | 22 |
| 4.14 | QQplot of variable y . . . . .   | 23 |
| 4.15 | Scatter plot of number of posts and reactions . . . . .                      | 23 |



# List of Tables

|     |  |    |
|-----|--|----|
| 4.1 | Labeled categories and Subcategories . . . . .                             | 16 |
| 4.2 | Subdivisions of subcategory "Technical (wanted)" . . . . .                 | 16 |
| 4.3 | Subdivisions of subcategory "Resource" . . . . .                           | 16 |
| 4.4 | Subdivisions of subcategory "Suggestion" . . . . .                         | 16 |
| 4.5 | Amount of total posts and reactions in different yearly quarters . . . . . | 17 |
| 4.6 | Amount of posts for different factors . . . . .                            | 19 |
| 4.7 | x and y values for coefficient calculation (from table 4.5) . . . . .      | 20 |

# Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

*API* Application programming interface

*GIF* Graphics Interchange format

*OSS* Open-source software

*SLR* Systematic literature review

# Chapter 1

## Introduction

### 1.1 Overview

The ever-growing usage of software in every sector of humans life these days, more and more people are being enticed by the mesmerizing concept of programming languages, the lucrative power of being able to create amazing software that everybody is so keen about. Since our eager step to integrate silicon chips to the mundane everyday objects such as a door or a bicycle, we have come quite a way in an astoundingly short amount of time. This has opened up innumerable opportunities – so many possibilities that even after people flocking to the fields, new ones are turning up frequently.

Developers, too, have found online groups and forums to be a lot better and easier way to interact with their peers. Support groups have been formed, and it has also attracted a remarkable number of beginners in the related fields. Interactions among the online community members have gone higher than ever before. But even such a great facility does not come without a few of its own problems either. Communities full of potential often can not maintain the inertia with which they start off. Furthermore, obstacles faced by newcomers in the sector sometimes turns into cause for frustration and result into the abandonment of further attempts by them. Thus, gradual lack of interaction makes way for decadence and the sustainability of once-thriving communities suffer and, slowly, they start to die out.

Attracted by the titillating charm of the software development sector, many newcomers have soon discovered a different nature of the undertaking, facing a lot of troubles penetrating this novel field [20]. A study by Steinmacher et al.[10] showed the many barriers newcomers usually face, by applying a combination of a systematic literature review (SLR) and feedback from test subjects selected from the concerned population. Their research has indeed been aimed at open-source software projects, which is a very specific area. People joining in commercial communities also face problems common among all the developers in general.

## 1.2 Objective

The purpose of this work is to analyze the interaction among the online community members in order to gauge from their communication the state of the sustainability of the community they make up together – to find out exactly how helpful are they being to each other, how much support they newcomers are getting from the veterans and how useful that actually is being.

# Chapter 2

## Related Work

The open-source software (OSS) barrier model, however, despite being exclusively constructed with the purpose of identifying the obstacles faced in OSS developers, has provided us a very well-structured framework that may be followed and easily improved upon in order to perform research on any related topic, as discussed by Balali et al.[23]. The categorization of the subjects of the research, for example, is essential in all cases, since the feedback from one group will invariably be influenced by the unique characteristics of that group[5]. A study by Burke[4] showed that matters such as gender diversity may affect the amount and type of support an individual may or may not be getting from the community, even in as diversified a sector as open-source software development. Interaction among the community members, possibly the most influential among the factors that decide the status of the support an individual receives, as Tewksbury[7] shows, now heavily relies on the medium of interaction. With the growing popularity of social media platforms, people are no longer dependent on physical gatherings to seek guidance or share their problems. Social media are playing a vital role in how the interactions between two individuals are being performed, evident from a study by Farzan[17]. And it is no longer being limited within the same two people in question, since support groups or forums naturally make it possible for other people to join in, since all support groups generally maintain all discussions open to all members[2]. This causes additional factors to enter the calculation, as the support that is being given may, at any time, go any of these two ways: (1) a lot of positive interaction may strengthen it, according to Millen[1], resulting in the receiver being more eager to interact with the members of the group in the future, or (2) negative feedback may force the person seeking support to discard all connections to that group, and thus getting discouraged to seek support from any similar group or forum in the future.

Online groups, especially social media, introduce a lot of factors that can easily cause people's behavior to deviate significantly from the normal. Such groups and forums, for example, contain documented content of previous interactions among members. Some even have dedicated space for such content or documents. Because the groups are spaces open to anyone for asking about something and there exists a lot of resources piled over time, it is too often observed that someone asking for help is ignored for a time being or continuously. This type of situation is created because of different expectations among members of different seniority – which, in its core, amounts to only a lack of proper communication. Incidents like this eventually leads

to a person getting discouraged and leaving the community altogether. Similar to the incident of resources, as the works of Steinmacher[10] and Showkat[20] have shown, other factors such as lack of proper official documentation also creates a negative attitude in people towards first the community, and then the work itself, resulting in dropouts.

A major lacking in case of such online groups, especially the “public” ones, originates from a blessing of their own nature. Groups and forums that usually are eager targets of users naturally build their popularity and user-base based on the content they provide. But, since the continuation of content generation also depends on the amount and nature of the interaction among the group members, any fluctuation in the group dynamic can easily influence the status of the content to go up on the group for a significant period thereafter. This is verily depends on not only the veteran members, but also the newcomers[16]. Although their contribution may not be markedly great in the now, many possess the potential to become an invested member of the group in the future[3]. Because of that, newcomer barriers often, however indirectly, end up holding a visible sway in the sustainability of a community in the long run.

The few ways a community can be directed to a more sustainable path include the inevitable task of making members more interested in participating more[3]. A number of works have found a somewhat contradictory yet unsurprising fact that user interaction can go up wit the help of offline interaction among the community members[21]. Out in the open, faced with each other, very obviously, the members will associate more substantive value to other members, ending up, in simpler terms, bonding better[11][21][12]. In developer communities, the idea is seen in the form of regular meet-ups and occasional events. So, it becomes clearly evident that physical interactions also, beside online presence, can go a long way in ensuring the sustainability of an online community.

The fact that interaction encourages participation is a pivotal reason the sustainability of such a group hinges on the proportion of effective interaction among members. And since, in modern times, online platforms provide almost ridiculously easy access to everyone from peers to potential collaborators, researchers flock to such platforms for the purpose of data collection, especially to various social media. A study undertaken by Chaki[15] solidifies the role of social media in crucial real-life decisions. Aided by the platforms themselves, by making use of a variety of helpful application programming interfaces (API), it is now surprisingly easy to collect a rather gigantic amount of data from different social media. Tanuja[22] talked about statistical analysis on data collected from social networks, discussing how a variety of analytic methods could be performed such data. Content classification, sentiment analysis, trend detection, criminal intelligence, prediction and recommendation systems are only some of the possibilities[22][6][18].

Among other statistical methods, studying the correlation between two variables helps reveal the relationship between the said variables in a remarkable way that is at the same time swift and yet effective. The concept of correlation revolves around the idea of simple relationships between variable that show their dependency in various

scenarios. Although most commonly associated to linear relationships, correlation can be useful in a number of situations[8].

In case of correlation calculation, social media data provides a more robust set of values to work with because of the simply immense amount of data that can be collected. Overwhelmingly vast in number and rich in nature with so many features of human behavior and interaction, social media data presents a very detailed canvas with which any work on correlation calculation gets a remarkable amount of leeway to function[9]. The possibility of shortcomings with smaller amount of data are eliminated, which makes the process able to paint a much clearer picture in the course of the study. The relationships of the bivariate values emerge with effective identification.

Because of the nature of social media data, a variety of studies have been performed using them. Due to the increasing number of online learning platforms, many has already moved to integrate social media as well. Besides boosting the scope of the reach, this has also ensured more presence from the potential learner base[8]. Social media being an open world for meeting between newcomers and veterans, the beginners have found invaluable opportunities to consult with the experienced about the future choices they would have otherwise made based on much less wisdom. Opportunities of higher studies have opened up, as have ways for consulting with others already in the know about the matters in question, behind which the role of social media is almost singular.

Similarly, business analysis has found its way to social media data, as observed by Mukherjee[14], exploiting the subtle relationship between the behavior of users of a particular brand before and after they have been exposed to stimuli in the form of social participation. From the works of Straton[19], Yadav[8], and Farzan[17], it is evident that other sectors are being equally competent, regardless of whether they have commercial intentions or academic[13].

# Chapter 3

## Methodology

Two methods was chosen, in this case, to determine the sustainability of online groups:

1. Constructing sustainability factor model, and
2. Calculating correlation coefficients.

After superfluously deciding the crucial factors (but *not precisely categorizing*), both methods would be pursued separately, so as no to influence the result of either. After a significant amount was data is collected, a final model would emerge from the improvements, and coefficients would be calculated. Then, a decision on the sustainability would be reached based on the results of both methods.

### 3.1 Target Group Selection

Six Facebook groups were selected for our purpose. Careful attention was paid in the selection process so that the chosen groups did not end up representing only a specialized group of people or a certain sector of the developer community. No particular programming language was held in consideration beforehand but, while selecting the groups, it was generally planned that a few representatives from the most popular languages now-a-days ought to be present. Equal consideration was paid to include languages designed for both front-end and back-end management, as was to the notion of different platforms.

With these considerations in mind, the following six Facebook groups were chosen:

1. NodeJS Bangladesh
2. Bangladesh Java Developers
3. Python Bangladesh
4. Dart/Flutter Programmers in Bangladesh
5. Programming Bangladesh
6. PHP Programmers BD



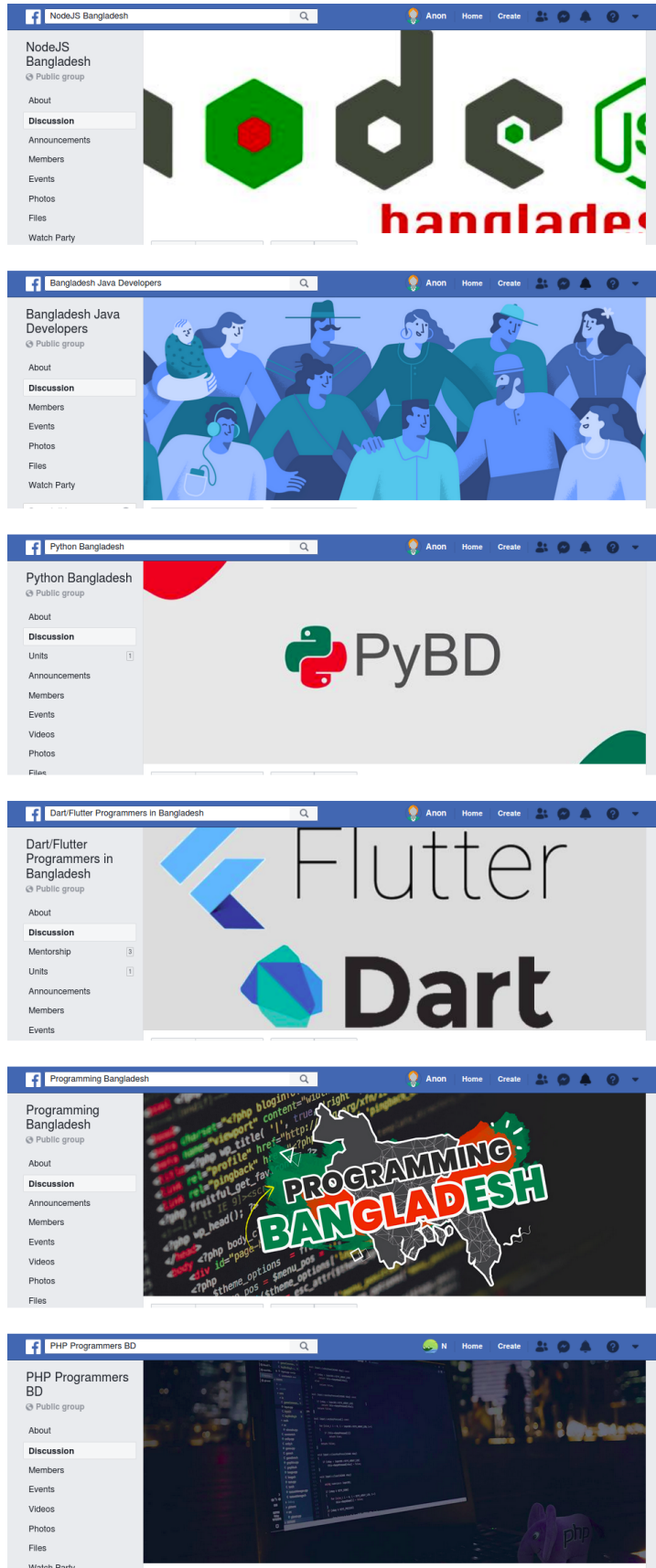


Figure 3.1: Screenshots of selected groups

Because of the widespread popularity, NodeJS was an expected choice. It won over others such as AngularJS and VueJS due to its staggeringly vast amount of packages. PHP is one of the most commonly used language for back-end-related tasks, and a lot of software, and thus developers, in our country have been built based on PHP, which implies that a remarkable number of veteran developers are familiar with it. Since NodeJS provides more than a sufficient amount of frameworks, Flutter earned its place among the selected. Java and Python were both considered as more of a general-purpose language. Java also covered the Android platform, Kotlin being relatively new in our country still. And Python is arguably, but only so, considered the most popular all-purpose language these days.

## 3.2 Sustainability Factor Model

The pivotal factor of sustainability measurement shares close connections to the exchange of assistance in a support group, online or otherwise. People join these groups from a very practical need of either getting the possible help that generally is readily available there or providing others with assistance based on their own knowledge and experience, and thus becoming something of a mentor figure in the community. As these two groups of members reciprocate in this way, people with other primary goals start joining the budding community, with variable purposes such as promoting themselves or looking for potential career opportunities. Different roles start forming among members, and, gradually, a full-fledged community is created. With the course of time, as the exchange between the different types of members persists, the community retains its sustainability.

Therefore, the condition of the sustainability of a support group could effectively be gauged by measuring the amount of help being demanded and being provided. For our purpose, we would design a preliminary model, comprising of the initial sustainability factors, which would be improved over time with the help of interaction data collected from the target groups. Throughout the revisions, the model might even discard the initial factors altogether (which is unlikely, however) but it would definitely acquire newer ones that subtly influence the state of sustainability in a group.

For this endeavor, post data of the Facebook groups would be collected, including post content, their replies, and the amount of interaction they received from the group members. The post content would be categorized based on the factors of sustainability, which might feature multi-level hierarchies. Upon division into the categories, a comparative model would be constructed that would, at a glance, be able to show how the different categories relate to each other, and which category precedes which one by how much. After a sufficient amount of data collection is done, the relationship among the categories would enable us to draw a conclusion about the state of the sustainability in question.

### 3.3 Correlation Coefficient

Since a model created from the mere proportion of the amounts of posts can hardly yield a decisive statement, a quantitative measure must be performed as well. Studying the correlation between value sets is a comparatively reliable method of determining the relationship between said values. For our purpose, we would make use of Pearson's product-moment correlation coefficient and Spearman's rank correlation coefficient.

#### 3.3.1 Pearson's Correlation Coefficient

Also known as the bivariate correlation, the Pearson coefficient is a splendid way to study the correlative relationship between two variables in a linear capacity. In terms of Cartesian plots, the correlation from the cluster created by the placement of the  $x$  and  $y$  values is translated to a more comprehensible, and precise, quantitative value, within the range of -1 and +1.

The Pearson coefficient can be calculated using the following equation:

$$r = \frac{\sum((x - M_x)(y - M_y))}{\sqrt{\sum(x - M_x)^2} \sqrt{\sum(y - M_y)^2}} \quad (3.1)$$

where,

$M_x$  is the mean of  $x$  values, and

$M_y$  is the mean of  $y$  values.

The sign of the value of the coefficient states whether the two variables are related or not; a positive value implies positive correlation, and a negative value refers to the opposite. The amount of the positive value, again, expresses how strong a relation the two variable shares. Considering  $r$  the correlation coefficient,

$$+ 0.5 < r < +1 \quad (3.2)$$

implies a strong relationship, whereas

$$0 < r < +0.5 \quad (3.3)$$

implies a weak one.

#### 3.3.2 Spearman's Rank Correlation Coefficient

Unlike the Pearson coefficient, the Spearman coefficient determines the correlation by studying the rankings of two variables. In the process of calculating this, the values are sorted and then assigned ranks, which then reveals the type of dependency among the values. It is different from the Pearson correlation by being nonparametric in nature, which gives it some advantage over the Pearson coefficient by being able to detect correlation correctly even if the points create significantly uneven scatter clusters on a Cartesian plot.

The Spearman coefficient can be calculated by using following equation:

$$r_s = \frac{\text{covariance}}{x_r \text{st.dev.} * y_r \text{st.dev.}} \quad (3.4)$$

# Chapter 4

## Implementation

### 4.1 Data Collection

The dataset construction process is comprised of data collection from the target Facebook groups via two methods and then reformation the obtained data in a structured way to be used for the calculation. The initial method to collect the group data had been the usage of Facebook’s Graph API. However, due to a change in the usage policy in 2019, the method was abandoned in preference of a headless browser (basically, crawling the web pages in question).

#### 4.1.1 Graph API

In order for easy interaction with various Facebook features, a number of services are open to developers so that they can communicate with the platform to create applications and services using data obtained from the social network. Among several of its APIs are Graph API. And the collection of a few of this API’s endpoints used to interact with Facebook groups is called the Groups API.

The Groups API works by assuming that an application is being installed for the target group, with the groups owners’ permission. Since in order to simply collect data, in practice, no application necessarily needs to be developed, only an application instance is created on Facebook for Developers. The ”developer” (the person who wants to collect the data) is required to create an app instance, and generate temporary API tokens to continue using the API, within the limitations set by Facebook.

Using the Groups API to collect posts for a target group requires running a cURL command following the format shown in figure 4.1, as provided on the Facebook for Developers official documentation page.

```
curl -i -X GET \  
"https://graph.facebook.com/554611227955614/feed?limit=5&access_token=EAACEdEos0..."
```

Figure 4.1: cURL command format for Graph API

Data returned by the Groups API, variable based on the parameters passed corresponding to required data, are obtained in the format displayed in 4.2.

```
{
  "data": [
    {
      "from": {
        "name": "Himanshu Sharma",
        "id": "512130107215"
      },
      "message": "I love this group!",
      "updated_time": "2018-06-28T23:31:49+0000",
      "id": "554611227955614_1708186035931455"
    }
  ]
}
```

Figure 4.2: Sample JSON response from Graph API

Data for each post appear as an object item in the "data" array in the JSON response.

### 4.1.2 Headless Browser

Data scraping from a website requires a user agent from the browser (or, "human") end that mimics the signatures of a human user to pose as a normal visitor to the target website. Scraping can easily done by combining a headless browser (that will pose as the human user) with a JavaScript library (that will do the actual extraction from the HTML page). For our purpose, we have used the open-source NickJS library briefly before switching to its web service Phantombuster.

**NickJS:** Making use of the Headless Chrome, which is a form of the Chromium browser only excluding any Graphical User Interface component, NickJS receives the HTML code of a target website, and then scrapes up the required elements as commanded, using JavaScript. Figure 4.3 shows a snippet of NickJS code written to get loaded post data from a group.

**Phantombuster:** Phantombuster is simply a web service developed by the aforementioned NickJS team. By removing the coding, this minimizes almost all the hassle of data collection by automating the actual scraping. There are a number of features of Phantombuster, not limited to Facebook only, that are presented in figure 4.4.

The free Phantombuster web service allows for 1-hour scraping per day for 14 days. And it requires the session cookies of a valid Facebook user account to scrape the data. From the perspective of Facebook, when the scraping runs, that user seems to be visiting the target web page(s). A simple cautionary measure is to be adopted while scraping, since the user account can be blocked if a human also logs into the account manually while it runs (It appears to Facebook as log-ins from two different locations).

```

const Nick = require("nickjs")
const nick = new Nick()

;(async () => {

  const tab = await nick.newTab()
  await tab.open("https://www.facebook.com/groups/300432120041516")

  await tab.untilVisible("#contentArea") // Make sure target content is loaded
  await tab.inject("http://code.jquery.com/jquery.min.js") // Use jQuery for scraping

  // Get group posts
  const posts = await tab.evaluate((arg, callback) => {
    const data = []
    $(".userContentWrapper").each((index, element) => {
      data.push({
        poster: $(element).find("_1dwg._1w_m").text(),
        content: $(element).find(".ldwg._1w_m").text(),
        timestamp: $(element).find(".timestampContent").text(),
        url: $(element).find(".5pcq").attr("href"),
        reaction: $(element).find("_68wo").text()
      })
    })
    callback(null, data)
  })

  // Print collected data in browser console
  console.log(JSON.stringify(posts, null, 2))

})();
.then(() => {
  console.log("Done")
  nick.exit()
})
.catch((err) => {
  console.log(`Error: ${err}`)
  nick.exit(1)
})
})

```

Figure 4.3: NickJS code snippet for scraping group post data

The image shows the Phantombuster service interface, which offers various scraping tools for Facebook. On the left, there is a search bar and a 'Sort by' dropdown menu set to 'Featured'. Below that, a 'Categories' list includes LinkedIn, Sales Navigator, Instagram, Twitter, Email, Facebook (highlighted), Google, Quora, Medium, Pinterest, Integrations, and Others. The main area displays six feature cards, each with a Facebook icon and a 'Use this Phantom' button:

- Facebook Ads Library Scraper**: Automatically scrape your competitors Facebook Ads.
- Facebook Auto Liker**: Automatically like posts of people you're close to.
- Facebook Event Guests Exporter**: Extract a list of all guests of any Facebook event.
- Facebook Group Extractor**: Scrape Facebook profiles, names & information from members of Facebook groups.
- Facebook Message Sender**: Automatically send messages to Facebook profiles.
- Facebook Page Reviews Collector**: Extract Reviews from a list of Facebook pages.

Figure 4.4: Various scraping features of Phantombuster service

### 4.1.3 Dataset

The collected data was obtained in JSON format (as shown in figure 4.5), which was then converted to CSV. This facilitated easy viewing and editing by a human by opening it as an Microsoft Excel file.

```
{
  "poster": "Dulal Hossain DH",
  "content": "Spring tool suite IDE keu ki use koren. Amar aktu help lage.",
  "timestamp": "January 28",
  "url": "https://www.facebook.com/groups/300432120041516/permalink/2142618089156234/",
  "comments": [
    {
      "commenter": "Dulal Hossain DH",
      "timestamp": "Monday, January 28, 2019 at 11:37 PM",
      "content": "Spring mvc ar jonno consol a error exception gula asteche na..."
    },
    {
      "commenter": "Dulal Hossain DH",
      "timestamp": "Monday, January 28, 2019 at 11:40 PM",
      "content": "Project run on server korle\nConsol a all time ai 4 line ase ....",
      "replies": [
        {
          "commenter": "Anirban Das",
          "timestamp": "Tuesday, January 29, 2019 at 6:23 AM",
          "content": "Dulal Hossain DH bro apnar screen shot deke [...]"
        },
        {
          "commenter": "Dulal Hossain DH",
          "timestamp": "Tuesday, January 29, 2019 at 7:44 AM",
          "content": "thanks @Anirban Das bro"
        }
      ]
    }
  ]
}
```

Figure 4.5: Collected data in JSON format

|    | A                | B  | N                    | O | P | Q                                      |
|----|------------------|--|----------------------|---|---|--|
| 1  | poster           | content  |                      |   |   | category v3                            |
| 2  | Abir D Bahen     | Dear boro vhai and bonera.ami professional java shikhte                        | [columns hidden ...] |   |   | asking for suggestion / recommendation |
| 3  | Samrat Al Shah   | Desktop Application বানানোর জন্য Java এর কোন Framework ব্যবহার                 | [columns hidden ...] |   |   | asking for suggestion / recommendation |
| 4  | Gagan Ajay Bah   | If an interviewer asks you can we change in main method                        | [columns hidden ...] |   |   | helping / sharing resource             |
| 5  | Dulal Hossain    | Spring tool suite IDE keu kj use koren. Amar aktu help                         | [columns hidden ...] |   |   | other question                         |
| 6  | Nahidujaman      | It is skipping gender, please solve the problem.\npackage                      | [columns hidden ...] |   |   | technical query                        |
| 7  | Navneet Rabadiya |  | [columns hidden ...] |   |   |  |
| 8  | আতিক হাসান শান্ত | আমি ফাইনাল ইন্টারভিউয়ের জন্য আজ নিয়ে পরামর্শ করতে চাই। কিন্তু আজ কোন আইডিয়া | [columns hidden ...] |   |   | asking for suggestion / recommendation |
| 9  | Dipen Adroja     | Java Variable Interview Questions for  | [columns hidden ...] |   |   | helping / sharing resource             |
| 10 | Azad Ahmed       | Web development এর জন্য Laptop এর config কোন হওয়া উচিত। বি:                   | [columns hidden ...] |   |   | other                                  |
| 11 | Azad Ahmed       | Web development এর জন্য Laptop এর config কোন হওয়া উচিত। বি:                   | [columns hidden ...] |   |   | other                                  |
| 12 | Hogue MD Zah     | Anyone know about the digital-ocean 5\$/ monthly VPS                           | [columns hidden ...] |   |   | other question                         |
| 13 | Adroja Manish    |  | [columns hidden ...] |   |   |  |
| 14 | ইমরাদ আদিন নিধর  |  | [columns hidden ...] |   |   |  |
| 15 | ইমরাদ আদিন নিধর  |  | [columns hidden ...] |   |   |  |
| 16 | ইমরাদ আদিন নিধর  |  | [columns hidden ...] |   |   |  |
| 17 | Dipen Adroja     | Interview Questions related to Types of Binary Tree.!!\n1.                     | [columns hidden ...] |   |   | helping / sharing resource             |
| 18 | ইমরাদ আদিন নিধর  |  | [columns hidden ...] |   |   |  |
| 19 | ইমরাদ আদিন নিধর  | আজ্ঞা হবে আজ পাঠ্যসময়ের সাথে... ছপে আসুন ০১ আপট হুপু ২৫০০ মিনিটে... \nইমরাদ   | [columns hidden ...] |   |   | administration to sustain              |
| 20 | Dinen Admia      | Second Introductory Article on Tree Data Structure                             | [columns hidden ...] |   |   | helping / sharing resource             |

Figure 4.6: CSV file opened as a spreadsheet



## 4.2 Model Construction

Considering the various factors of an online developer community, an initial model is designed that will refer to the sustainability of the community. The interactions of the community members is categorized and preliminary factors are coined. The required data is collected from the target groups on Facebook. As more data is obtained, the preliminary model needs to be adjusted by adding new factors, or by reforming the categories, or by modifying the hierarchies of the categorization. After a sufficiently useful amount of data is collected (this is signaled by either collecting all the post data of a group or the increasing absence of the introduction of new factors against time), the final model is constructed.

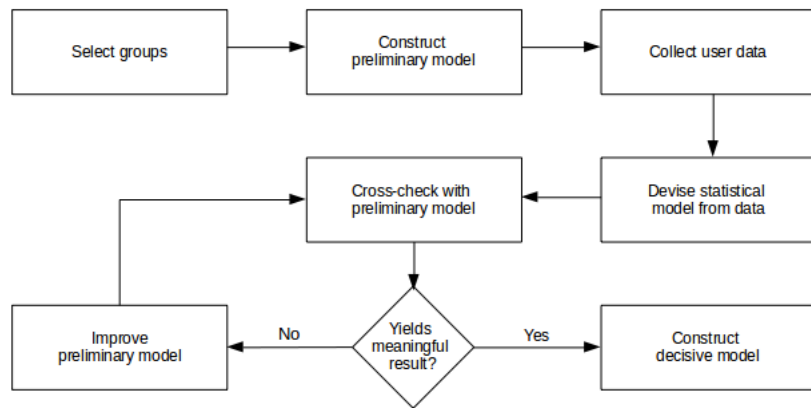


Figure 4.7: Flowchart of working plan

**Factor categorization:** The initial factors considered in our case had revolved around the idea of the proportion between the amount of support demanded in the groups and the amount received. With these two as the primary points of interest, other factors were then added that expressed the amount of opportunities—which were not explicitly related to these two—that appeared in the group. The categories and subcategories that had been developed over the iterations comprising of the improvement and cross-checking process are displayed in table 4.1. For ease of modification, the factors (belonging to all hierarchies) had been assigned labels.

Some subcategories formed their own hierarchies, as shown in tables 4.2, 4.3, and 4.4.

| Category      | Subcategory | Label | Detail                                     |
|---------------|-------------|-------|--|
| Help wanted   | Technical   | d01   | Error description/screenshot provided      |
|               | Resource    | d02   | Books, tutorials, etc.                     |
|               | Suggestion  | d03   | How to start, which is better, etc.        |
|               | Company     | d04   | Inquiries about companies (for career)     |
|               | Coaching    | d05   | Direct teaching via inbox, Skype, etc.     |
|               | Other       | d06   | Unspecified/vague                          |
| Help received | Technical   | r01   | (All same as previous)                     |
|               | Resource    | r02   |  |
|               | Suggestion  | r03   |  |
|               | Company     | r04   |  |
|               | Coaching    | r05   |  |
|               | Other       | r06   |  |
| Promotion     | Self        | p01   | Own blog, channel, series, etc.            |
|               | Company     | p02   | Specific companies, start-ups, etc.        |
|               | Other       | p03   | Relation with subject unidentified         |
| Opportunity   | Job         | o01   | Part- or full- time for a mentioned period |
|               | Freelance   | o02   | Individual projects/tasks                  |
|               | Collab.     | o03   | Multiple people discuss to set conditions  |
|               | Volunteer   | o04   | Unpaid contribution                        |
| Community     | Admin.      | c01   | Keeping group environment friendly         |
|               | Online      | c02   | In-group discussions, competitions, etc.   |
|               | Offline     | c03   | Real-life meetups, events, etc.            |

Table 4.1: Labeled categories and Subcategories

| Subcategory        | Sub-level | Label | Detail                                |
|--------------------|-----------|-------|---------------------------------------|
| Technical (wanted) | Setup     | d0101 | Installation, environment setup, etc. |
|                    | Error     | d0102 | Error in code                         |
|                    | Tool      | d0103 | Libraries, frameworks, APIs, etc.     |
|                    | Hopeless  | d0104 | "Solve it for me!"                    |
|                    | Hint      | d0105 | Hints on how to solve                 |

Table 4.2: Subdivisions of subcategory "Technical (wanted)"

| Subcategory | Sub-level   | Label | Detail   |
|-------------|-------------|-------|--|
| Resource    | Doc         | d0201 | Official documentation                           |
|             | Self (free) | d0202 | Free self-learn content (e.g., blogs, tutorials) |
|             | Self (paid) | d0203 | Paid self-learn content (e.g., courses)          |
|             | Contest     | d0204 | Programming contests                             |

Table 4.3: Subdivisions of subcategory "Resource"

| Subcategory | Sub-level  | Label | Detail                             |
|-------------|------------|-------|------------------------------------|
| Suggestion  | Config     | d0301 | Hardware or software configuration |
|             | Comparison | d0302 | Which one is better/best           |
|             | Idea       | d0303 | Asking for project idea            |
|             | Start      | d0304 | How/where to begin                 |

Table 4.4: Subdivisions of subcategory "Suggestion"

**Posts and reactions:** A total of 14,177 posts had been collected for our purpose, among which 1,797 had been discarded because of being of non-helpful nature (irrelevant spam post, deleted external content, etc.), resulting in 12,380 posts that could be worked with.

Reaction units (hereafter to be referred simply as "reactions", but not to be confused with Facebook's "react" feature for posts) were calculated for posts by considering member interactions with them. In case of posts, comments, and their replies, each "like" or "react" was considered worth 1 reaction unit, and each comment or reply was considered worth 2 reactions. Reactions from replies of nested comments were counted as part of the original post.

| Quarter | Total posts | Discarded | Usable posts | Reactions |
|---------|-------------|-----------|--------------|-----------|
| 2016 Q1 | 750         | 95        | 655          | 2742      |
| 2016 Q2 | 694         | 98        | 596          | 2434      |
| 2016 Q3 | 875         | 112       | 763          | 1979      |
| 2016 Q4 | 843         | 84        | 759          | 2693      |
| 2017 Q1 | 687         | 102       | 585          | 2981      |
| 2017 Q2 | 784         | 104       | 680          | 2377      |
| 2017 Q3 | 981         | 137       | 844          | 3129      |
| 2017 Q4 | 652         | 113       | 539          | 2778      |
| 2018 Q1 | 695         | 95        | 600          | 3528      |
| 2018 Q2 | 914         | 101       | 813          | 3148      |
| 2018 Q3 | 1009        | 127       | 882          | 2973      |
| 2018 Q4 | 1184        | 115       | 1069         | 3380      |
| 2019 Q1 | 863         | 112       | 751          | 3418      |
| 2019 Q2 | 976         | 121       | 855          | 3185      |
| 2019 Q3 | 1227        | 139       | 1088         | 3611      |
| 2019 Q4 | 1043        | 142       | 901          | 3226      |

Table 4.5: Amount of total posts and reactions in different yearly quarters

Assigning the numbers obtained from the collected data to appropriate category and subcategory labels produce a comprehensive model of the nature of different types of posts seen in the groups. Figure 4.8 displays the comparative domains of the factors listed in table 4.1, except for the "help received" category the subcategories of which overlaps those of the "help wanted" category.

The *amount* of posts (as opposed to the *percentage*) corresponding to different labels is given in table 4.6.

**Relation between "wanted" and "received":** Since the categorization of the factors could not allow for separation of the "help wanted" type posts from the "help received" type posts, these two factors shared their posts due to overlapping. We identified "received" posts based on whether the assistance requested in a post was provided or not, where the original request in the post also made it to be marked as a "wanted" post. Because of this, the "received" posts actually form a subset of the

”wanted” posts, the exclusions being those in which no discernible assistance was provided.

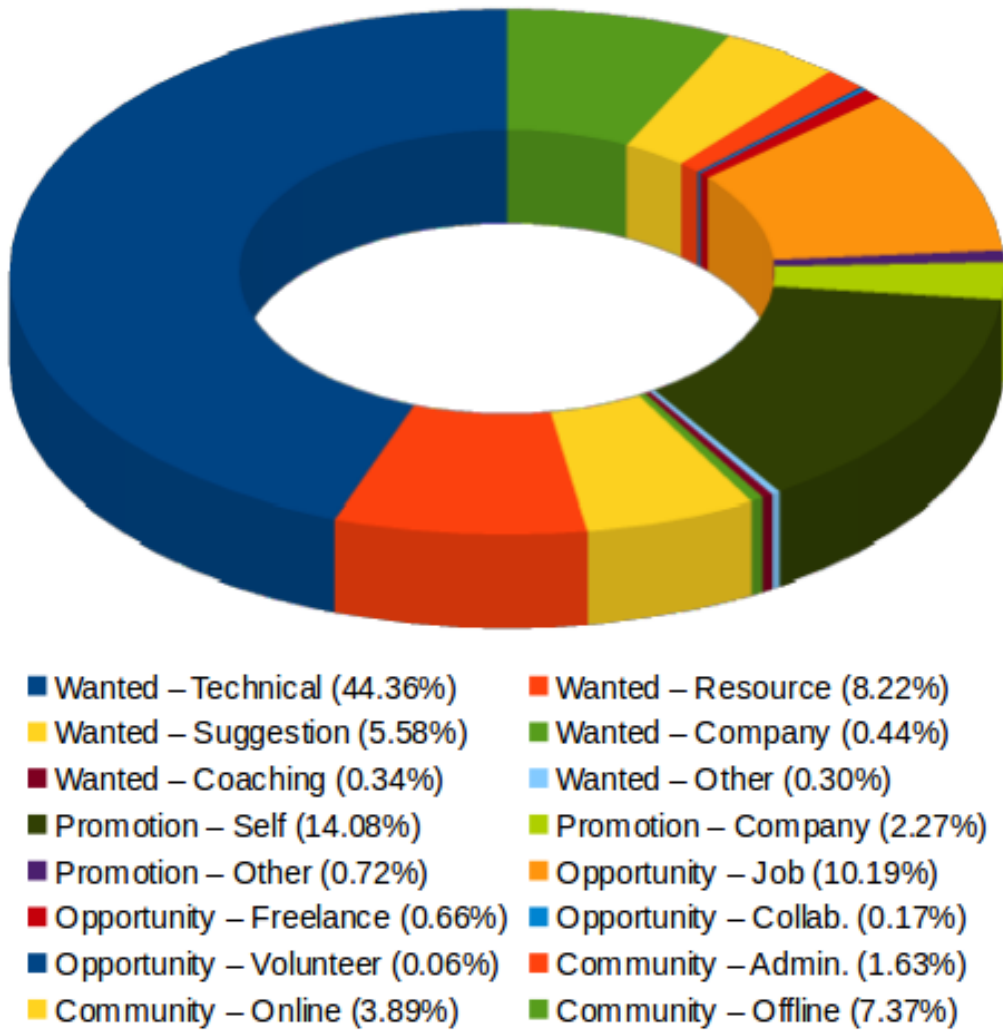


Figure 4.8: A model showing the proportions of different post types

The relation between these two factor categories are presented in figure 4.9 in the form of easily-comparable values.

| Factor         | Amount of posts | Factor        | Amount of posts |
|----------------|-----------------|---------------|-----------------|
| (W) Technical  | 5492            | (P) Self      | 1743            |
| (W) Resource   | 1018            | (P) Company   | 281             |
| (W) Suggestion | 691             | (P) Other     | 89              |
| (W) Company    | 54              | (O) Job       | 1261            |
| (W) Coaching   | 42              | (O) Freelance | 82              |
| (W) Other      | 37              | (O) Collab.   | 21              |
| (R) Technical  | 5129            | (O) Volunteer | 8               |
| (R) Resource   | 1003            | (C) Admin.    | 168             |
| (R) Suggestion | 974             | (C) Online    | 481             |
| (R) Company    | 52              | (C) Offline   | 912             |
| (R) Coaching   | 9               |               |                 |
| (R) Other      | 4               |               |                 |

Table 4.6: Amount of posts for different factors

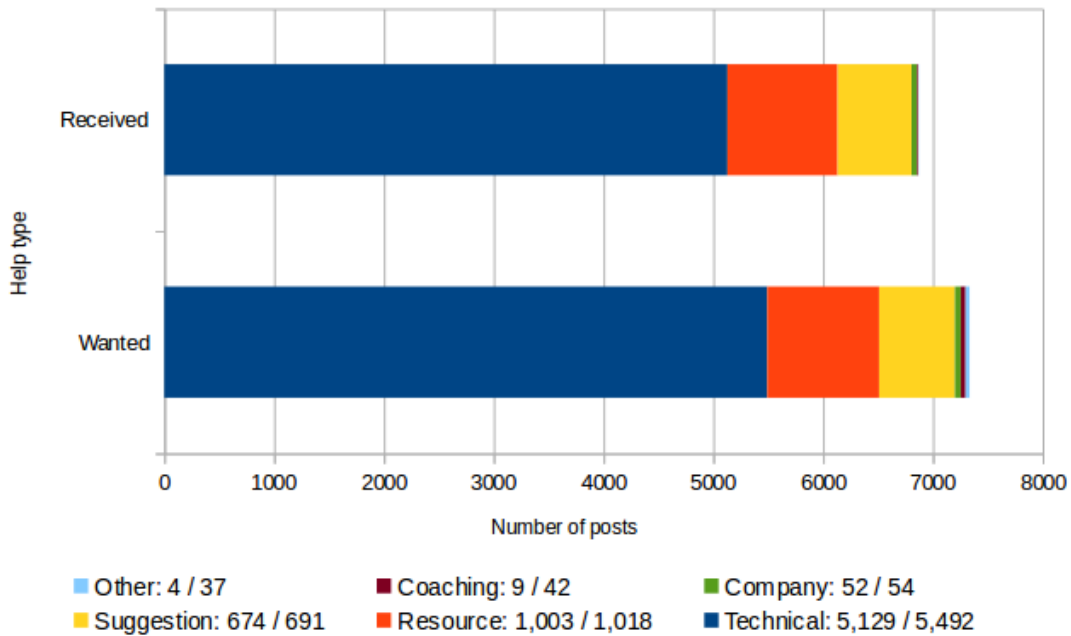


Figure 4.9: Comparison of amount of help wanted and received

### 4.3 Coefficient Calculation

The number of posts from each quarter comprise the  $x$  values, whereas the number of reactions make up the  $y$  values for the correlation calculation.

#### 4.3.1 Pearson's Correlation Coefficient

From the collected data, in the process of coefficient calculation, the following values are derived:

Summation of  $x$  values,  $\sum x = 14177$

| x    | y    |
|------|------|
| 750  | 2742 |
| 694  | 2434 |
| 875  | 1979 |
| 843  | 2693 |
| 687  | 2981 |
| 784  | 2377 |
| 981  | 3129 |
| 652  | 2778 |
| 695  | 3528 |
| 914  | 3148 |
| 1009 | 2973 |
| 1184 | 3380 |
| 863  | 3418 |
| 976  | 3185 |
| 1227 | 3611 |
| 1043 | 3226 |

Table 4.7: x and y values for coefficient calculation (from table 4.5)

Mean of  $x$  values,  $M_x = 886.062$   
Summation of  $y$  values,  $\sum y = 47582$   
Mean of  $y$  values,  $M_y = 2973.875$   
Number of sets,  $N = 16$   
 $\sum (x - M_x)^2 = 461872.938$   
 $\sum (y - M_y)^2 = 3046267.75$   
 $\sum ((x - M_x)(y - M_y)) = 563832.125$

| $x - M_x$ | $y - M_y$ | $(x - M_x)^2$ | $(y - M_y)^2$ | $(x - M_x)(y - M_y)$ |
|-----------|-----------|---------------|---------------|----------------------|
| -136.062  | -231.875  | 18513.004     | 53766.016     | 31549.492            |
| -192.062  | -539.875  | 36888.004     | 291465.016    | 103689.742           |
| -11.062   | -994.875  | 122.379       | 989776.266    | 11005.805            |
| -43.062   | -280.875  | 1854.379      | 78890.766     | 12095.180            |
| -199.062  | 7.125     | 39625.879     | 50.766        | -1418.320            |
| -102.062  | -596.875  | 10416.754     | 356259.766    | 60918.555            |
| 94.938    | 155.125   | 9013.129      | 24063.766     | 14727.180            |
| -234.062  | -195.875  | 54785.254     | 38367.016     | 45846.992            |
| -191.062  | 554.125   | 36504.879     | 307054.516    | -105872.508          |
| 27.938    | 174.125   | 780.504       | 30319.516     | 4864.617             |
| 122.938   | -0.875    | 15113.629     | 0.766         | -107.570             |
| 297.938   | 406.125   | 88766.754     | 164937.516    | 120999.867           |
| -23.062   | 444.125   | 531.879       | 197247.016    | -10242.633           |
| 89.938    | 211.125   | 8088.754      | 44573.766     | 18988.055            |
| 340.938   | 637.125   | 116238.379    | 405928.266    | 217219.805           |
| 156.938   | 252.125   | 24629.379     | 63567.016     | 39567.867            |

Figure 4.10: Calculated values for Pearson's correlation coefficient

From the values calculated, Pearson's correlation coefficient can be figured out using equation 4.1.

$$r = \frac{\sum((x - M_x)(y - M_y))}{\sqrt{\sum(x - M_x)^2} \sqrt{\sum(y - M_y)^2}} \quad (4.1)$$

Calculated Pearson's correlation coefficient,  $r = 0.4753$

For the 16 pairs of values, a significance level of 0.10 yields a P-Value of .062795, which is statistically significant at  $p < .10$ .

### 4.3.2 Spearman's Rank Correlation Coefficient

The Spearman coefficient can be calculated using equation 4.2.

$$r_s = \frac{\text{covariance}}{x_r \text{st.dev.} \times y_r \text{st.dev.}} \quad (4.2)$$

The following values can be obtained based on the collected data:

Mean of  $x$  ranks = 8.5  
Standard deviation of  $x$  ranks = 4.76  
Mean of  $y$  ranks = 8.5  
Standard deviation of  $y$  ranks = 4.76  
Covariance = 10.67

Using equation 4.2:

Calculated Spearman's ranks correlation coefficient,  $r_s = 0.471$ .

For the calculated  $r_s$  value, the 2-tailed p value is 0.06582, which is statistically non-significant.

### 4.3.3 Cross-Checking with R

By feeding the values of the number of posts and the reactions to the R command-line, the coefficient calculation was performed to validate the results obtained above.

The scatter plot in figure 4.15 shows the relationship between the two values  $x$  and  $y$ , in this case, the amount of posts and the amount of reactions. The point placement in the plot implies a positive tendency, albeit somewhat weak in nature.

The coefficient value of 0.47 calculated via the R command-line also verifies the result obtained from the aforementioned calculation.

| x    | y    | $x_r$ | $x_r - M_x$ | $y_r$ | $y_r - M_y$ | $(x_r - M_x) * (y_r - M_y)$ |
|------|------|-------|-------------|-------|-------------|-----------------------------|
| 750  | 2742 | 5.00  | -3.50       | 5.00  | -3.50       | 12.25                       |
| 694  | 2434 | 3.00  | -5.50       | 3.00  | -5.50       | 30.25                       |
| 875  | 1979 | 9.00  | 0.50        | 1.00  | -7.50       | -3.75                       |
| 843  | 2693 | 7.00  | -1.50       | 4.00  | -4.50       | 6.75                        |
| 687  | 2981 | 2.00  | -6.50       | 8.00  | -0.50       | 3.25                        |
| 784  | 2377 | 6.00  | -2.50       | 2.00  | -6.50       | 16.25                       |
| 981  | 3129 | 12.00 | 3.50        | 9.00  | 0.50        | 1.75                        |
| 652  | 2778 | 1.00  | -7.50       | 6.00  | -2.50       | 18.75                       |
| 695  | 3528 | 4.00  | -4.50       | 15.00 | 6.50        | -29.25                      |
| 914  | 3148 | 10.00 | 1.50        | 10.00 | 1.50        | 2.25                        |
| 1009 | 2973 | 13.00 | 4.50        | 7.00  | -1.50       | -6.75                       |
| 1184 | 3380 | 15.00 | 6.50        | 13.00 | 4.50        | 29.25                       |
| 863  | 3418 | 8.00  | -0.50       | 14.00 | 5.50        | -2.75                       |
| 976  | 3185 | 11.00 | 2.50        | 11.00 | 2.50        | 6.25                        |
| 1227 | 3611 | 16.00 | 7.50        | 16.00 | 7.50        | 56.25                       |
| 1043 | 3226 | 14.00 | 5.50        | 12.00 | 3.50        | 19.25                       |

Figure 4.11: Calculated values for Spearman's rank correlation coefficient

```
install.packages("ggplot2")
library(ggplot2)

# Import data
data <- read.table("data.dat", header=FALSE)
# Set column names
names(data) <- c("post", "react")

# Plot points
scatter_plot <- ggplot(data, aes(post, react))
scatter_plot + geom_point() + labs(x = "Total posts", y = "Total reactions")

# Calculate correlation
cor.test(data$post, data$react, method = "pearson", conf.level = 0.1)
```

Figure 4.12: R language code to generate scatter plot and calculate coefficient

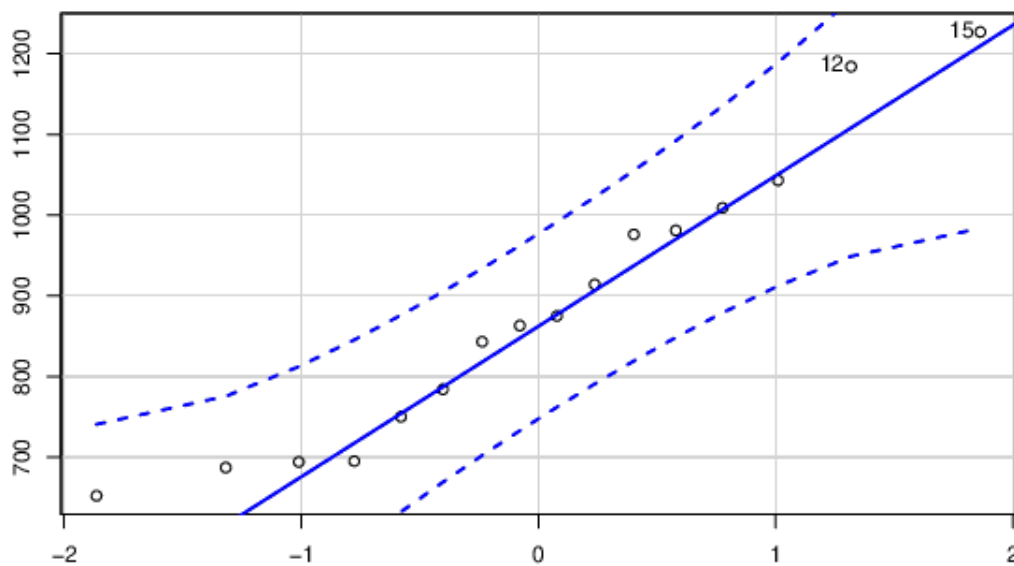


Figure 4.13: QQplot of variable x



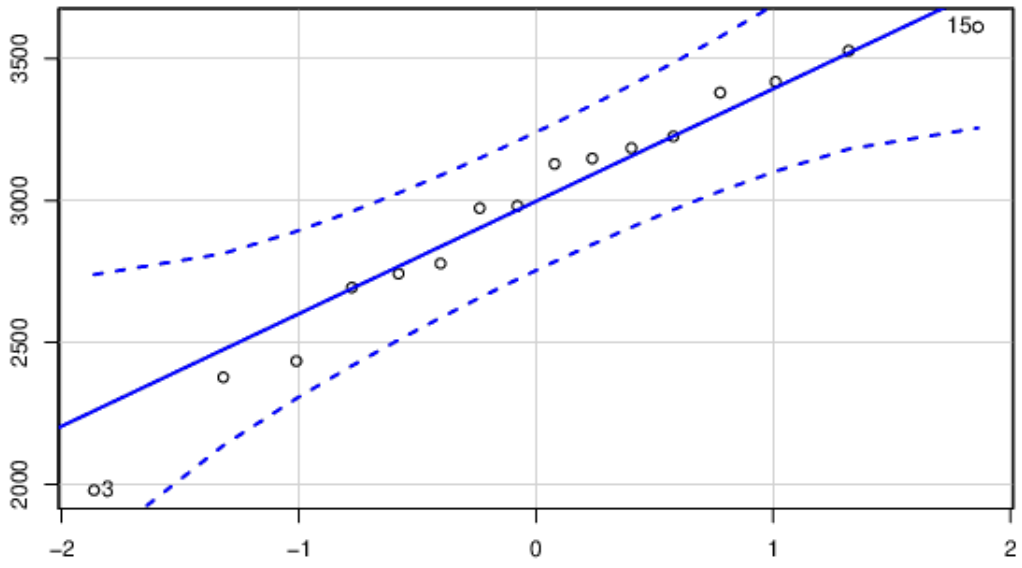


Figure 4.14: QQplot of variable y

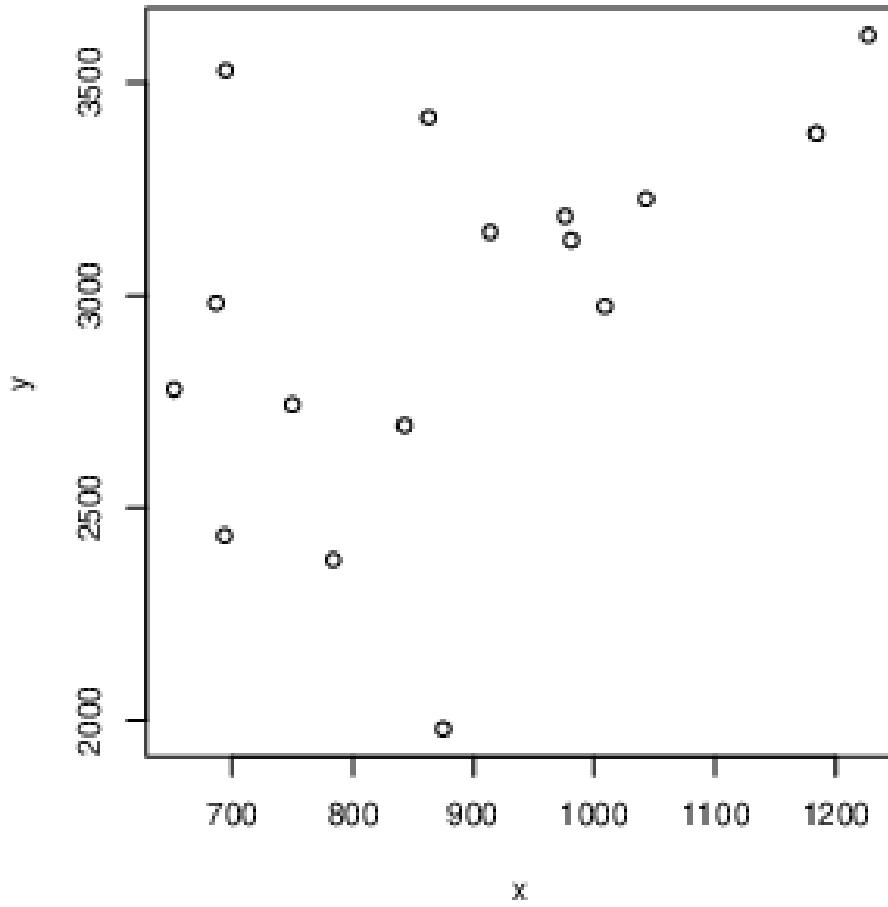


Figure 4.15: Scatter plot of number of posts and reactions

# Chapter 5

## Result Analysis

**Sustainability factor model:** From the model shown in figure 4.8, it is apparent that most of the posts appearing in the target Facebook groups are related to assistance and support of various kinds. 44.36% among the total number of 12,380 posts are help-related; 14.08% are about self-promotion of individual members; 10.19% provide job opportunities with mentioning a preset occupational period; 8.22% directly or indirectly ask for a variety of resources; 7.37% posts are associated with community engagement in the real-world (as opposed to online participation); and 5.58% posts refer to suggestions and recommendations from other members of the groups.

This implies that a significant percentage of the interaction of the group is at least centered around asking for support from other members—which, actually, is the primary purpose for the creation of support groups of this kind. The values in figure 4.9 show a healthy proportion of these demands for assistance being fulfilled, as the ratio between the amount of help wanted and received is 1.07:1, revealing that an impressive 93.39% of the pleas for help are getting answered. Although lack of further decisive feedback from the original posters keep us from deciding that all these answers are sufficiently meaningful, the sole action of replying with a possible solution keeps the interaction in the groups going.

Moreover, the percentage of technical help-related posts (44.36%) is significant because of the characteristic explanatory nature of such posts, since one of the conditions of being identified as a help type post had been the addition of relevant code, error log, or system configuration pertaining to the problem. Providing a way to go through the steps that the original poster themselves had helps tremendously in the debugging process, which boosts the possibility of other members engaging with the post in question. As a result, technical help posts, even if indirectly, contribute to increasing the interaction in a group.

**Correlation coefficients:** The nature of correlation coefficients is to present the relation between two sets of values and provide a quantitative dimension so that the significance of the relation can be judged. Within the range of -1 and +1, the nearer the coefficient value advances towards +1 from +0.5, the stronger the relation between the  $x$  and  $y$  values is considered; and the closer it approaches to 0 from +0.5, the weaker.

From our calculations, the Person coefficient value turned out to be 0.4753, and the Spearman coefficient 0.471. Both the values are positive, which indicates that there is a relation between the number of posts and the amount of reactions received. However, since the values are less than +0.5, the relation is a weak one. This implies that the growth of posting in the groups is somewhat influenced by the amount of reaction it receives from the group members, although not strongly enough to be picked up by a correlation as a decisive factor.

Combining the insights obtained from the sustainability factor model and the values generated via coefficient calculation, it becomes evident that the target Facebook groups currently are, overall, in a sustainable condition. Interactions of the community members are keeping them active as support is being constantly requested and provided. The amount of help wanted that is related to resources and recommendations are comparatively a lot lower than explicit technical assistance. Engagement will increase further if, instead of direct problem-solving, more attention is paid to self-improvement resources.

Various companies can officially start recruiting from the groups themselves rather than relying on individuals posting occasionally, removing the possibility of good opportunities not being lost due to lack of appropriate attention. This will cause professional interest to grow in members, making them more eager to participate in group activities, ensuring future sustainability. Companies will be able to find reliable employees, vouched by veteran members of the groups. This will encourage them to expend resources for the betterment of the members, and thus the community itself.

As sustainability is hardly an idea that can be completely expressed via the relationship of a couple of factors, there still is much to study about the complex relationship among various subcategories listed in Chapter 4. Research works can be performed on how exactly a factor is affecting the others and how they may be influenced. For example, if the amount of collaborative opportunities can be increased, the members will be able to form strong professional networks, which will both open doors to higher studies for those who are still studying and attract more scholars to form partnerships with potential youth. Through this type of works, young developers from Bangladesh will be able to get exposure to the global environment, which will not only make them more experienced but will also strengthen the developer community as a whole.

# Chapter 6

## Conclusion

Communities, both online and offline, thrive as long as its members, through their reciprocal relationships, preserve the continuity of a healthy environment. As long as the members engage with each other, the community will be considered to be in a sustainable state. The target groups in our work had turned out to have retained their sustainability. However, further study devoid of the effect of some of the limitations in our endeavor definitely has the potential to reveal something more intricate about the state of the developer community.

### 6.1 Limitations

**Limited number of groups:** Developers are as diverse a community as programming languages or frameworks are. Each sector also gradually instills in the individuals its unique flavor of said industry. We had tried to maintain variety, by paying attention to diversity in platforms, type, and capabilities. Still, "developer community" is quite vast to cover effectively.

**Active and passive interaction:** Interaction, being of qualitative in nature, is hard to put into numbers. Member interaction in a group can demonstrate manifold aspects which are difficult to translate into quantitative measurements. For example, commenting on a post is more engaging than simply "liking" it; and the content of a comment can be both partially helpful and yet hurtful. The reaction unit assignment (1 for "reacts" and 2 for comments) took care of the former issue, but the latter could not be satisfactorily resolved.

**Image posts:** Our work focused on textual content of posts and comments only. Textual posts with background images were stripped down to the texts only. And comments with nothing but images, GIFs, and Facebook stickers were simply counted as a comment (albeit with no content).

**Individual sustainability:** Since the calculation was done using the aggregated data of all six target groups, the combined sustainability emerged as a result—which had been the purpose of this work. However, even though the overall state turned out to be sustainable, it should not be considered individually true for the groups in question, since the lower level on one's part could have been nullified by a higher sustainability in another.

## 6.2 Future Work

Multiple studies on the sustainability of groups belonging to different sectors, paradigms, or programming languages can be performed, which will be able to provide insights as to the state of various subgroups. A possible future endeavor includes an endeavor to interview members of the target groups and obtain decisive data on a much smaller yet intricate scale as to how sustainable the members themselves see a group as. This will also be able to provide answers about how to welcome more to the community and retain those that have already found it a safe haven.

# Bibliography

- [1] D. R. Millen and J. F. Patterson, “Stimulating social engagement in a community network”, in *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, 2002, pp. 306–313.
- [2] G. von Krogh, S. Spaeth, and K. R. Lakhani, “Community, joining, and specialization in open source software innovation: A case study”, 7, vol. 32, 2003, pp. 1217–1241.
- [3] C. Lampe and E. Johnston, “Follow the (slash) dot: Effects of feedback on new members in an online community”, in *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, 2005, pp. 11–20.
- [4] M. Burke, C. Marlow, and T. Lento, “Feed me: Motivating newcomer contribution in social network sites”, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 945–954.
- [5] C. Bird, “Sociotechnical coordination and collaboration in open source software”, in *Software Maintenance (ICSM), 2011 27th IEEE International Conference on IEEE*, 2011, pp. 568–573.
- [6] H. LI, F. Cai, and Z. Liao, “Content-based filtering recommendation algorithm using hmm”, in *2012 Fourth International Conference on Computational and Information Sciences*, 2012, pp. 275–277. DOI: 10.1109/ICCIS.2012.112.
- [7] D. Tewksbury, “Online-offline knowledge sharing in the occupy movement: Howtooccupy.org and discursive communities of practice”, *American Communication Journal*, vol. 15, no. 1, pp. 11–23, 2013.
- [8] P. S. Yadav and P. Srivastava, “A statistical analysis of impact of social networking media on higher education”, in *2013 2nd International Conference on Information Management in the Knowledge Economy*, 2013, pp. 9–13.
- [9] H. K. Chan, E. Lacka, R. W. Y. Yee, and M. K. Lim, “A case study on mining social media data”, in *2014 IEEE International Conference on Industrial Engineering and Engineering Management*, 2014, pp. 593–596.
- [10] I. Steinmacher, A. P. C. Steinmacher, T. Conte, and M. A. Gerosa, “Preliminary empirical identification of barriers faced by newcomers to open source software projects”, in *Brazilian Symposium on Software Engineering (SBES 2014)*, 2014. DOI: 10.1109/SBES.2014.9.
- [11] S. Angelopoulos and Y. Merali, “Bridging the divide between virtual and embodied spaces: Exploring the effect of offline interactions on the sociability of participants of topic-specific online communities”, in *2015 48th Hawaii International Conference on System Sciences (HICSS)*, 2015, pp. 1994–2002.

- [12] L. Jin, R. D., and M.-C. Boudreau, “The nature of hybrid community: An exploratory study of open source software user groups”, *The Journal of View publication stats Community Informatics*, vol. 11, no. 1, 2015.
- [13] C. López, Y.-R. Lin, and R. Farzan, “What makes hyper-local online discussion forums sustainable?”, vol. 2015, Mar. 2015. DOI: 10.1109/HICSS.2015.292.
- [14] P. Mukherjee and B. J. Jansen, “Correlation of brand mentions in social media and web searching before and after real life events, Phase analysis of social media and search data for super bowl 2015 commercials”, in *2015 IEEE 15th International Conference on Data Mining Workshops*, 2015, pp. 21–26. DOI: 10.1109/ICDMW.2015.60.
- [15] D. Chaki, Z. Moinul, and A. Ali, “Understanding complex social network of government officials in decision making”, Sep. 2017. DOI: 10.2139/ssrn.3044296.
- [16] K. Cheevasuntorn, B. Watanapa, S. Funilkul, and W. Krathu, “Factors influencing teenagers behavior on posting and sharing messages via facebook”, in *2017 2nd International Conference on Information Technology (INCIT)*, 2017, pp. 1–4.
- [17] R. Farzan and C. Jonassaint, “Exploring dynamics of facebook health support groups: A leadership perspective”, in *Proceedings of the 50th Hawaii International Conference on System Sciences — 2017*, 2017.
- [18] S. Savaş and N. Topaloglu, “Crime intelligence from social media: A case study”, in *2017 IEEE 14th International Scientific Conference on Informatics*, 2017, pp. 313–317. DOI: 10.1109/INFORMATICS.2017.8327266.
- [19] N. Straton, R. Vatrapu, and R. R. Mukkamala, “Facebook and public health: A study to understand facebook post performance with organizations’ strategy”, Dec. 2017, pp. 3123–3132. DOI: 10.1109/BigData.2017.8258288.
- [20] D. Showkat, “Determining newcomers barrier in software development: An it industry based investigation”, in *CSCW ’18 Companion*, Jersey City, NJ, USA, 2018. DOI: 10.1145/3272973.3274046.
- [21] R. Farzan, Y.-R. Lin, and D. Lu, “What happens offline stays offline? examining sustainability of a hybrid social system”, in *Proceedings of the ACM conference on Hypertext and Social Media*, Halifax, Canada, 2019.
- [22] U. Tanuja, H. L. Gururaj, and V. Janhavi, “An exploratory analysis on data features and analysis techniques in social networks”, in *2019 11th International Conference on Communication Systems Networks (COMSNETS)*, 2019, pp. 535–537.
- [23] S. Balali, I. Steinmacher, U. Annamalai, A. Sarma, and M. A. Gerosa, “Newcomers barriers... is that all? an analysis of mentors and newcomers barriers in oss projects”, *Computer Supported Cooperative Work (CSCW) (2018)*, vol. 27, pp. 679–714, DOI: 10.1007/s10606-018-9310-8. [Online]. Available: <https://doi.org/10.1007/s10606-018-9310-8>.

# Appendix

## Hyperlinks to selected Facebook groups

1. NodeJS Bangladesh  
<https://www.facebook.com/groups/nodejsbd>
2. Bangladesh Java Developers  
<https://www.facebook.com/groups/300432120041516>
3. Python Bangladesh  
<https://www.facebook.com/groups/pythonbd>
4. Dart/Flutter Programmers in Bangladesh  
<https://www.facebook.com/groups/804731159871538>
5. Programming Bangladesh  
<https://www.facebook.com/groups/10ms.programming>
6. PHP Programmers BD  
<https://www.facebook.com/groups/pprogrammers.bd>