CHALMERS UNIVERSITY OF TECHNOLOGY | UNIVERSITY OF GOTHENBURG

SOCIAL NETWORK ANALYSIS

Bachelor of Science Thesis in Software Engineering and Management

MALCOLM KENTE

Department of Computer Science and Engineering
UNIVERSITY OF GOTHENBURG
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2017

Social Network Analysis

MALCOLM KENTE

Supervisor: HÅKAN BURDEN
Examiner: BOBAN VESIN

University of Gothenburg
Chalmers University of Technology
Department of Computer Science and Engineering
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Department of Computer Science and Engineering
UNIVERSITY OF GOTHENBURG
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2017

Social Network Analysis

Malcolm Kente
University of Gothenburg
Software Engineering and Management
Gothenburg, Sweden
gusnyema@student.gu.se

Abstract

This thesis describes an exploratory case study that looks to understand user engagement and migration within online communities. Considering the structures of interaction provided within online communities on Facebook, the objective is to discover (1) the influence on the level of user activity within existing Social Network Sites (SNSs) communities, and (2) migration patterns of users across these communities. Implemented for this study is an Activity Analysis, which looks at the engagement level of users within those data sets. Additionally, user migration is explored through a Network Analysis so as to identify possible migration patterns that may exist between forums.

Keywords: *Social Network Sites; Data Mining; Social Media; User Engagement & Migration; Facebook Graph API.*

1 Introduction

The internet is an essential component of navigation in everyday life [17]. As a prominent social arena, the internet provides people the opportunity to meet and interact with each other. Social life on the internet is built on social components of interaction known as Social Media. Social media refers to the many, relatively inexpensive and widely accessible, electronic tools that enable people to publish and access information, collaborate on a common effort, or build relationships [24]. Tools such as chat forums, blogs, newsgroups, and social networks can all be considered as social media.

Social networks, also referred to as social network sites (SNS), are web-based services that allow an individual to (1) create a profile within a defined system, (2) establish connections with other users, and (3) observe their connections and those made by others within the system [3]. SNSs contain millions of registered users due to their attributed networking prowess. Huge amounts of data are constantly distributed within social media platforms among millions of users. Many of the users integrate these sites into their daily practices [3] because of the freedom to connect, express themselves, and share content. [39].

In today's society, one of the constantly growing SNSs is known as Facebook. The Facebook platform provides vast possibilities due to the abundance of data stored about individuals, groups, and products. Considered the heart of the social web [30], over 1 billion users worldwide are active daily on this platform. Social interactions such as status updates, chatting, posting of photos and/or videos, and playing games all contribute to the richness of data stored within Facebook. Facebook attributes interaction structures Information with tangible insight is amalgamated from this data by use of Facebook's Graph API [30].

This paper is an exploratory study that looks to understand user engagement and migration within online communities. Considering the structures of interaction provided within social media platforms, the objective is to discover (1) the influence on the level of user activity within existing SNS communities, and (2) migration patterns of users across these communities. The SNS analyzed for this study is Facebook, where data from the data sets Comments and Likes found in the online communities: Cabotagestudien (CS) [8], Här Stannar Sverige (HSS) [11], and Vi Som Vill Ha Våra Jobb Kvar (VSVHVJK) [14] are chosen for investigation. This study executes an Activity Analysis which looks at the engagement level of users within those data sets. Additionally, user migration is explored through a Network Analysis so as to identify possible migration patterns that may exist between forums. The monitoring of user engagement contributes in the attempt to map user movement patterns across SNS communities.

The results of this case study are aimed to contribute to the greater discussion of user engagement and migration within social media. For the chosen online communities and online communities across social media, the benefits and contributions of this study are presented in the Discussion section. For the software engineering society, contributions are meant towards data mining techniques and handling of mined data. Additionally, the providing of strategy towards mapping user migration. Thus, the contributions to the IT society include:

- Differentiating and grouping users based on their level of engagement within and across SNS communities.
- Presenting diverse appliances for mining data from SNSs.
- Visualizing basic patterns for user migration across SNS communities.

The structural set up of this thesis is described in the upcoming sections. Section 2 highlights relevant literature found in relation to this study. Section 3 is the method(s) of choice used with this study. The methodology introduces the research question and the case subject, as well as the cross referenced case subject. Section 4 is the collection of data where data collection tools are introduced and their uses described. The analysis methods are addressed in Section 5. Section 6 presents the results uncovered after data analysis. Discussions to the results are presented in Section 7 while the limitations and validity threats are pointed out in Section 8. Finally, the conclusions and suggestions of future work are presented in Section 9.

2. Related Work

The study of Cvijikj and Michahelles [5] looks at the effects of posts shared by a Page Administrator, within a Facebook brand page. Similar to this thesis, their study takes into account the post characteristics and and analyzes user interaction in terms of the number of likes and comments for each wall post. Cvijikj and Michahelles look at Facebook post types factors that can influence user interaction. This thesis considers the post types as cogent variables to aide in research of user activity within the chosen communities. Results from their study show that posting type and category to have significant effect over user interaction. Moreover, effects over comments and likes ratios prove larger when compared to the effects over interaction duration. These results are presented so as to identify the implications for social media marketing.

Gerolimos [15] conducts a study that examines and analyzes user's comments on the Facebook pages of 20 American academic libraries. From his research, Gerolimos observes a higher ratio of user activity where Likes stand for 82% of user participation over Comments. In relation to this thesis, the Likes and Comments of wall posts are explored for discovery of variation in user engagement. Understanding the difference within each of the structures can enlighten on the influence of social media on user activity.

Other studies that look at user interaction and migration include the Benevenuto et al [1] and Kumar, Zafarani, and Liu [21]. Similar to this thesis, both studies investigate the notion of user migration across online platforms with the consideration of user activity as key factor. Benevenuto et al investigates the interaction and navigation of users when they connect to SNSs, whereas, Kumar, Zafarani, and Liu conduct a study focused on user migration patterns within social media.

Based on detailed clickstream data, the analysis done by Benevenueto et al reveals evidence on the frequency of the connection between users and social networks. The types and series of activities that users are also discovered. Kumar, Zafarani, and Liu investigate the possibility and feasibility of user migration of user attention between seven social media sites. Their study shows that (1) the studying of user migration across social media is indeed feasible, (2) migration patterns can be identified, and (3) the possibility to act on migration patterns by monitoring high net worth users.

3. Methodology

This section describes the methodological approach chosen for this study. It includes a comprehensive description of research question, the applied method, collection of data, and data analysis.

3.1. Research Question

Social media's capacity for joining a diversity of people across the globe toward a shared goal is unprecedented. These engineered environments are renowned for cultivating forms of online sociality [38] and creating a cluster of online communities tied around similar interests. However, user engagement within social media remains highly interesting. Not enough attention is being paid to mechanisms by which social media govern their networks [37]. Interaction dynamics offered through social media platforms, like Facebook, need to be investigated in order to show how social media leverages user engagement within online communities.

This study attempts to identify user engagement and movement within the Facebook Social Network. By considering the site's structures of interaction, the objective is to discover (1) the influence on the level of user activity in Facebook online communities, and (2) possible migration patterns of users across these communities. Thus, the following research questions are to be answered:

RQ 1: *How do the structures of interaction within social media platforms influence the level of user activity in online communities?*

RQ 2: *How can user activity aid in mapping user migration across online communities?*

In order answer the research question, appropriate data collection methods are implemented. Data is mined from Facebook's Graph API [33] via appliances such as Graph API Explorer [10] , Facepager application [19], and Java-written methods using the RestFB client [34]. Additionally, the data is analyzed against similar groups in order to increase validity and reduce biased results [28]. The following chapter discusses the methods and tools chosen for data collection and analysis.

3.2. Research Method

This thesis is conducted as an exploratory research. This strategy is chosen due to its insightful advantages. Following the classification done by Robson [27], exploratory research is good for finding out what is happening, seeking new insights, and generating ideas/hypotheses for new research. Additionally, the approach is credited to embody inductive characteristics [29] that help to further understand the factors that promote the social impact of the data sets.

For this study, focus is placed on data found within the structures of interaction provided in social network platforms. Channels that allow user interaction within Facebook, such as Comments and Likes are defined as the structures of interaction. From these structures, data sets within the Facebook communities of CS [8], VSVHVJK [14], and HSS [11] are investigated in order to discover the levels of user engagement. Data sets are collections of data that are propitiated and structured by a number and types of attributes or variables [25]. A cross-comparison of all the communities is also implemented in order to discover prevalent users in these groups, their social functionality, and migration patterns.

3.3. Case Companies

Cabotagestudien (CS) is a research study that examines international haulage in Scandinavia. Its purpose is to produce facts on haulage carried within Scandinavia. Data for CS is collected through crowdsourcing [32, 31], where volunteers/users contribute vehicle position data to the CS mobile app and Facebook Page.

Vi Som Vill Ha Våra Jobb Kvar (VSVHVJK) [14] is an online community that deals with the haulage in Sweden. The community sheds light on the issues within the Swedish road freight transport and work to improve conditions for Swedish domestic drivers. VSVHVJK, which openly supports the work of CS, connects to users through a Facebook Page.

The Här Stannar Sverige (HSS) [11] movement has grown into a online community of users concerned with road safety for haulage within Sweden. HSS has established a Facebook Group where information and ideas are presented and discussed between different users.

The chosen online communities are considered to be part of a limited scope where the users and the content are guaranteed to focus on similar objectives. Each community deals with a similar concern which is haulage and maintain an active online population. Making the plausibility of discovering user activity within these communities high. Also taking into account the geographical focus being Sweden/Scandinavia, it is more likely to discover user movement across the communities.

These factors leverage the potential of receiving insightful data from all communities. This study will focus on the activity within the Facebook entities of CS, VSVHVJK, and HSS. The aim is to discover factors that may influence user engagement in online communities. Additionally, this study looks at user migration across Facebook communities of and attempts to discover possible migration patterns that may exist between online communities.

3.4. Planning for the Case Subject

The defining of this case study follows the guidelines presented in Runeson and Host [28], where the necessity of a plan for research is recognized and developed. Similar to the elements described by Robson [26], a simple scheme is composed for this study and looks to cover:

- Objective - What is to be achieved?
- Research Question - What is to be answered?
- Method(s) - What approach is to be used?

From the highlighted elements a protocol (figure 7, check appendix) is developed for assisting with data collection and analysis. Runeson et al [29] describe a case study protocol as a continuously changing document that stays updated because of changes made to it throughout the research period. Thus, this study's protocol was initially designed to cover components of research that would easily be updated without affecting the research process.

4. Data Collection

The importance of obtaining accurate and reliable information is recognized when collecting data for analysis. The papers of [22, 29] point out the need for the organizing of large raw data after initial collection. Structurally organizing data contributes to alleviating the refining and analysis of data. Runeson et al [29] state that organizing raw data helps the researcher assure the quality of the study by ensuring that data isn't overlooked through disorganization. Therefore, this study implements member checking [16] to help improve the accuracy and validity of results. For each data set retrieved, ideas or theories of significance are considered. The retrieved data is then examined and interpreted alongside those theories. This process can either validate or oppose the theories inferred, while also providing new data to research. Member checking allows for researchers to correct errors perceived as wrong interpretations, as well as, the opportunity to volunteer additional information [16].

The collected data for this thesis derives from the Facebook Graph API, and is recognized as vital towards the progress of this research. The Graph API provides data sets of metadata tags that are used as base reference to the activity and network analysis. Upcoming is a thorough description of the Graph API and data collection methods used for this study.

4.1. Data Collection

The Graph API

Facebook's Graph API is a low-level HTTP based API that is used as the main method for getting data in and out of Facebook's Social Graph [30]. The Social Graph depicts social interactions through the composition of nodes, edges and fields [33].

- Nodes represent objects such as people, photos, pages and comments,
- Edges are the connections between these objects
- Fields provide information about the objects.

Graph API requests are made in order to access within Facebook's Social Graph. These requests require the use of an access token in order obtain the information stored in Facebook's Social Graph. An access token is an opaque string that recognizes a user, app, or Page and is used to make Graph API calls [7]. Access tokens, also considered OAuth tokens [30], are provided in order for users to gain temporary, but secure access to Facebook APIs. Facebook provides the Graph API Explorer (Figure 1) as a tool to quickly generate tokens and navigate the Social Graph. Additionally, access tokens can be securely obtained through third party applications due to the OAuth 2.0 authentication and authorization permissions leveraged by Facebook [20, 30]. Finally, Facebook provides different access token types such as User, App, and Page Access Tokens to support different use cases [7].

RestFB and Facepager

RestFB is an open source, Facebook Graph API and Old REST API client written in Java [6, 34]. Attributing simplicity and flexibility, it is a minimal method for fetching and publishing new items to Facebook [34].Developed by Till Keyling and Jakob Jünger, this tool fetches any accessible JSON data from social media platforms such as Facebook, Twitter, Google+, etc. All retrieved data is stored in a SQLite database that is connected to the application.

The basic layout and concepts of Facepager include [18] (Figure 2):

- *Nodes View* - The objects of data collection: Facebook feed, Twitter tweets, and any other object an API is able to return.
- *Data View* - Shows all data of the selected object or row in Nodes View.
- *Column Setup* - Controls which data is shown in Nodes View.
- *Query View* - Starting point for fetching data from an API.
- *Status View* - For logging and debugging purposes. The messages displayed in this region inform about the details of collection process or errors.

4.2. Using the Data Collection Appliances

The Role of the Graph API Explorer

In the beginning of the research process, the Graph API Explorer [10] is used to retrieve data from the CS Facebook page. A Page access token [7] is generated in order to gain

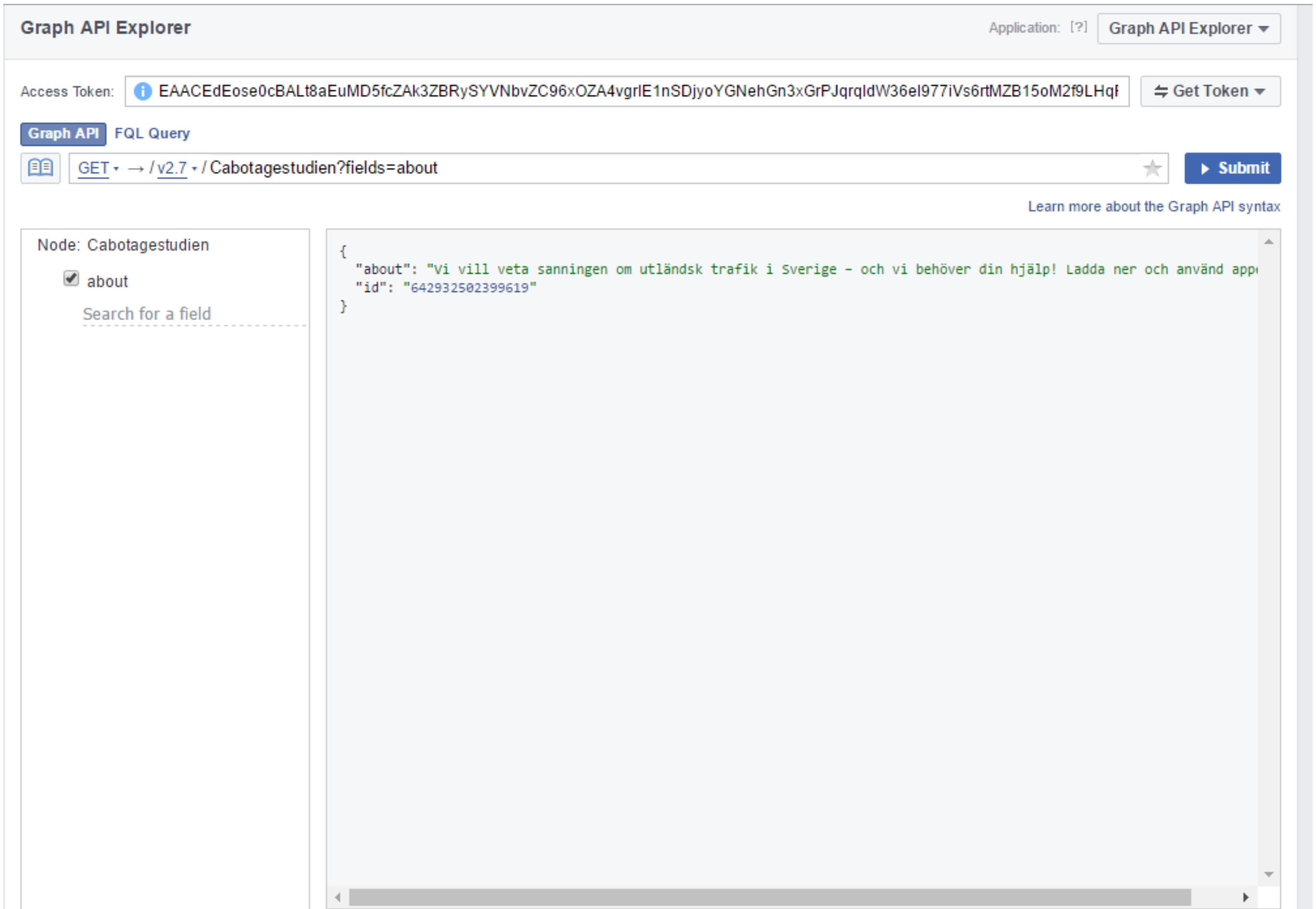

Figure 1. The Graph API Explorer

access to the Graph API. The *page-id* node is recognized as a representation of a Facebook Page with fields such as *name, id* [12].

From the *page-id* node, the metadata tag *page-id/feed*, also known as an edge, is the first recognized feasible data-set. This is considered a criterion that allows for research moving forward. It's feasibility is attributed the social connection it displays between the Facebook page and a Facebook page feed. The *page-id/feed* tag returns the posts (including status updates) and links published by a page, or by others on a page [13] (Figure 3).

However, the obtaining of data from the *page-id/feed* data-set on the Graph API Explorer is limited to 100 posts. Any attempt over the given number, an error message is displayed. Given the *page-id/feed* limitations, data collection from any of the online communities would require reevaluating so as to obtain the needed data-sets. As stated by Runeson et al[29], the antecedent data collection process for studies usually return without clear objective. A more eloquent selection strategy is to be implemented if feasible data is to be extracted.

The Role of the Facepager Application

After the initial trial with the Graph API Explorer, the Facepager application is used to first extract data from the CS Facebook page. The reasoning being the scope of which the limitations extend with the application. Unlike the Graph Explorer, Facepager can retrieve more than 100 posts in one query. The main work Facepager does, consists of three steps for every node (Figure 3): (1) It assembles URLs from node data and the query setup, (2) data is requested from the web by using these URLs, and (3) the downloaded data is sliced up and put into the database as new nodes – ready for export or as new starting points for fetching data [18].

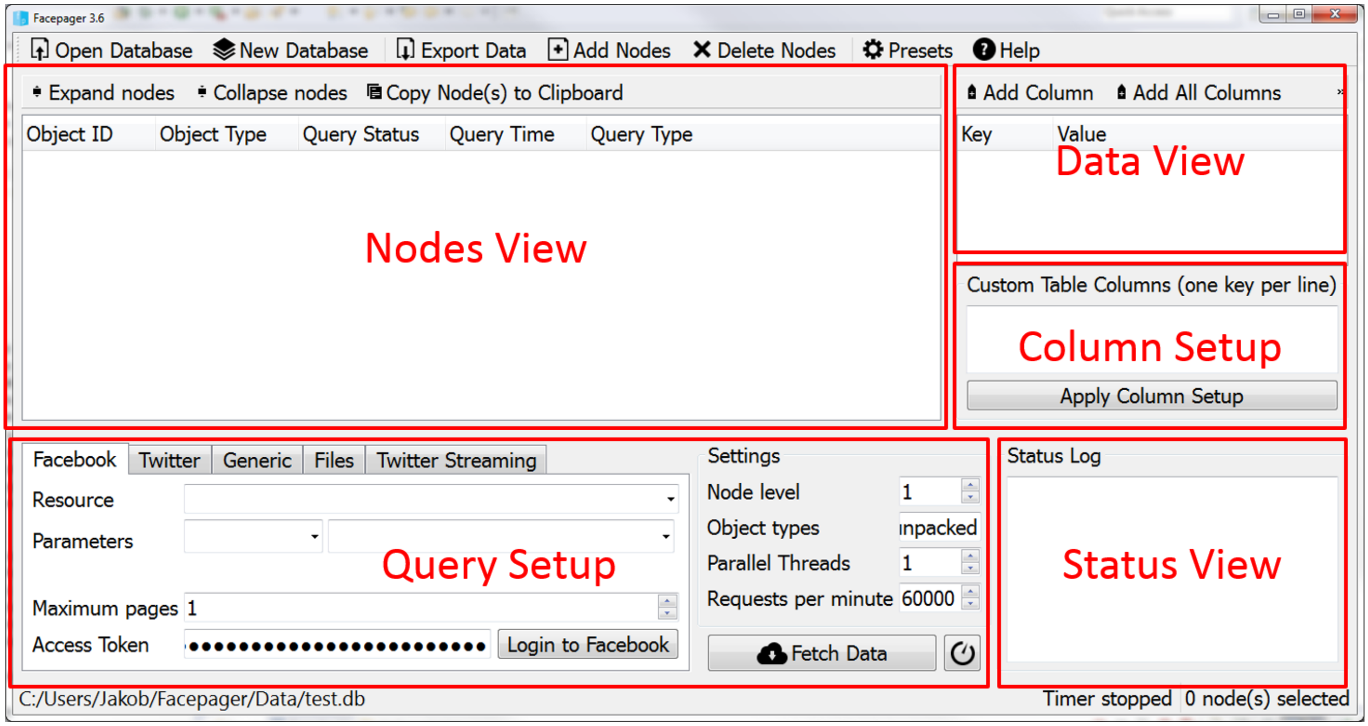

Figure 2. The Facepager Application, see [19]

Similar to Graph Explorer, the Facebook module available in Facepager is also connected to Facebook's Graph API. This means that through Facepager, data can be accessed in the same way as the Graph Explorer. All retrieved data is similarly presented in a JSON-format and can be simply monitored in the Data View section (refer to Figure 2) of the application. However, unlike the Graph Explorer, Facepager's connection to SQLite database makes the retrieving and saving of data much easier.

The Facepager process of retrieving *page-id/feed* from the Cabotagestudien Page, is described below [18]:

1. Clicking a New Database option on the toolbar to create a blank SQLIte database.
2. Clicking Add Nodes (Facebook ID's) on the toolbar to create a new node instance of the User/Page/Group being mined. Facebook ID's for nodes are usually last part of entity's url. In the case of this study 'Cabotagestudien' taken from 'https://www.facebook.com/Cabotagestudien/' is the Facebook ID for our page.
3. Selecting the Facebook tab in Query Setup view and adding the proper resource, parameters, and maximum pages for the type of data being mined. For this study, the resource selected for Cabotagestudien is ¡page¿/feed, with the parameters Object ID and the maximum of 30 pages.
4. Obtaining an access token with the Facepager requires one to login to Facebook, with the Login to Facebook feature provided on the interface.
5. Fetching data is then done by selecting the entered Node (Facebook ID) and clicking on Fetch Data.
6. Once the data is fetched, the parent nodes can be extended so as to view the child nodes by clicking Expand Nodes. The Data View tab shows the detailed JSON file retrieved.
7. Columns can then be changed to the users desired need for mining. For this research, columns for the created time and object type are selected for focus.

The Role of the RestFB Client

The RestFB API client package comes well equipped with a convenient library. The API client, which doesn't require additional libraries included, bolsters its ability to to have zero run-time dependencies [34]. RestFB accesses data from the Facebook Graph API using the format of the Java programming language. RestFB users are able to not only access and fetch data from the Graph API, but also log, search, publish, delete and operate FQL query's through

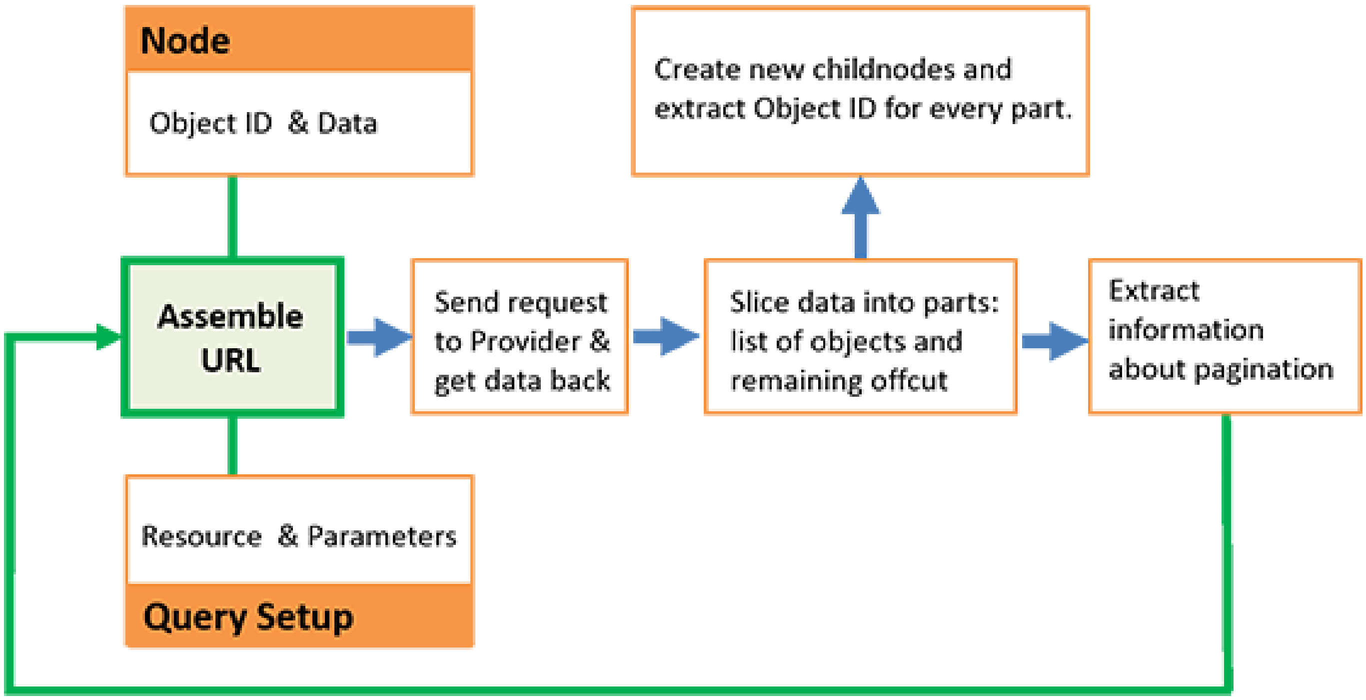

Figure 3. Facepager Process Model

java written code [35]. Finally, the parsing capabilities of JSON objects provided by RestFB from the messy files provided by the Graph API makes for good versatility for this client package.

For fetching purposes the RestFB client package provides the ***fetchObject*** method for retrieving a single object from Facebook. Similarly, ***fetchObjects*** method is provided for the fetching of multiple objects. For all API calls, RestFB needs to be informed of how to turn the JSON returned by Facebook into Java objects [35]

This study utilizes the ***fetchObjects*** method for accessing data from Facebook. The method collects JSON data from *page-id/feed* data-set, provided by the Graph API, which is parsed by RestFB's JSON package [36].

5. Data Analysis

Yin [40] states that the chain of evidence should allow a reader to deduce results and conclusions from the collected data. Thus, this study performs qualitative data analysis in order to acquire concise data from a clear chain of evidence. As suggested by Runeson and Höst [28], the analysis is performed in parallel with collection of data so as to always infer new insights. Although, unlike the suggestion of [28], the analysis for this study was conducted by a single researcher. Following a research protocol and plan, the importance of data validity is taken into account when analyzing the findings of this study.

5.1. Activity Analysis

Investigating the user engagement levels within the chosen online communities requires proper acumen into the dynamics of user activity. In order to properly understand how engaged users are within these communities, the *page-id/feed* data is to be examined and analyzed. For this study, the examining and analyzing of this data is referred to as the Activity Analysis.

The RestFB package [34] is used in providing a Java compatible Facebook client (figure 4) that is implemented into Java methods which fetch, read and parse the *page-id/feed* data. User activity within the online communities of CS, HSS, and VSVHVJK [8, 11, 14] is investigated. This is done in order to gain results from multiple data sources that will contribute to a more impartial conclusion [29].

Posts, Post Types, and Structures of Interaction

Posts: To get a fair insight into the user engagement, this study concentrates on content available in the feed within each forum. Feeds are posts (including status updates) and links that can be published by normal users, page, and/or group administrators [13]. For Facebook Pages and Groups, posts can be published either by a page/group administrator or normal users. Facebook categorizes posts into Post Types known as (1) Events, (2) Links, (3) Photo, (4) Status, and (5) Video [12]. The feed from each community contains various posts of different types thus making the Post Type a vital factor of consideration when conducting the investigation.

Post Types: A Post contains Comments, Likes, and Shares where user activity is present. Users are able to comment on the posted content in the Comments, show approval by clicking the 'thumbs up' button in Likes, and finally share the posted content in Share. The Comments field contains Reply and Likes. The Reply allows users to reply to the existing comment in the post, while the Likes allows users to show approval/appreciation by clicking "thumbs up". For this study, the Reply and Likes fields found within Comments are classified as Comments Reply and Comments Like.

Structures of Interaction: The analysis addresses user engagement levels by examining user activity within interaction structures of the chosen online communities. This study classifies the Comments and Likes of a post, as those structures that provide room for user interaction within a community. The Shares of a post, are considered to encourage interaction from outside the community, and can therefore skew the intended results. Thus, the Activity Analysis mainly focuses on Comments and Likes.

Classification of Users

Active Users: One aspect that contributes to the Activity Analysis is the classification of users according to the amount of activity demonstrated across the interaction structures. A data set of Active Users (*AUs*) is to be established for each community. For this study, *AUs* are defined as those users that are either active in one or more of the interaction structures.

The process of identifying Active Users considers attributes contained within Comments, Likes, Comment Replies, and Comment Likes. Research from this study reveals that each interaction structure contains multiple attributes that are part of their build-up. Considering each data set as a node [33], these attributes are the fields [33] within each node. Accessing these fields is achieved through clustering [2] of data in order to identify the relevant piece of data required for this analysis. The field *from*, which represents the profile of a user [9], is identified as the essential piece of data required for conducting the analysis. The *id* and *name* values are extracted and a profile for a user is stored so as to later be used in calculating the amount of user activity within the online communities.

Most Active Users: By establishing *AUs*, an investigation into the users most active within all four fields is considered. Users active within all four fields are placed into a Most Active Users (*MAUs*) data. As a criterion to establish the *MAUs*, users must observe activity in all of the chosen interaction structures.

The Activity Analysis is directed towards user engagement within online communities. Thus, results from this analysis look to answer the first research question:

RQ 1: *How do the structures of interaction within social media influence the level of user activity in online communities?*

5.2. Network Analysis

For user migration and the recognition of migration patterns across online communities a Network Analysis is carried out on the online communities of CS, HSS, and VSVHVJK [8, 11, 14]. This analysis uses the *AUs* data set to identify Common Users (*CUs*) across the online communities and map user movements between the communities. This study focus on site migration concept similar to the one mentioned in [21].

Common Users: The Common Users (*CUs*) data set is the collection of the recurring users found in each of the selected online communities. The existence of *CUs* within these data sets is measured through their user engagement in the structures for interaction within each community.

The Network Analysis seeks to provide answers for the research question:

RQ 2: *How can user activity aid in mapping user migration across online communities?*

6. Results

The results for this study are collected and stored in a cloud storage service as well as a SQLite database. For convenience during the analysis, retrieved data is stored in .csv files that are later grouped and analyzed using Microsoft Excel [23]. Initially, in order for get results from the analyses, the Object ID of *page-id/feed* is exported from Facepager. The RestFB Client is used to formulate Java code (Figure 4) that mines data from the data sets connected to *page-id/feed* tag from the Graph API.

```java
        // Restfb's Facebook client.
        FacebookClient fbClient = new DefaultFacebookClient(accesstoken, Version.VERSION_2_0);

        // Retrieves JSON list of all comments associated to the given
        // post.
        JsonObject fetchCommentsResults = fbClient.fetchObjects(Arrays.asList(feedID), JsonObject.class,
                Parameter.with("fields", "comments.limit(5000)")); // name,id

        // Retrieves all JSON Objects under the first JSON Object in the
        // JSON list.
        JsonObject fromUser = fetchCommentsResults.getJsonObject(feedID);

        /**
         * Checks if JSON Object 'comments' exists: If yes, commentary
         * data is extracted. If no, a sys.out function alerts of
         * non-existence.
         **/

        if (fromUser.has("comments")) {

            // Retrieves the JSON Object 'comments'.
            JsonObject fromUs = fromUser.getJsonObject("comments");
            System.out.println("From comments: \n" + fromUs + "\n");

            // Retrieves the JSON Array 'data'.
            JsonArray lang = (JsonArray) fromUs.get("data");

```

Figure 4. RestFB Facebook Client used to extract data from Comments of a Facebook post.

6.1. Activity Analysis Results

For each of the online communities, the feed extraction time-line extends from the date of the first published post to the most recent one. Thus, 614 posts are extracted from CS between 2013-2016, 38131 posts are extracted from HSS between 2014-2016, and 453 posts are extracted from VSVHVJK between 2012-2016.

Comments, Likes, and Post Types

From the extracted posts, seven post types are recognized. These post types are classified as: Photos, Status, Links, Video, Event, Music, and Note. Noted from the results, the Music and Note post types are recognized to not be defined in [9]. Table 1 displays the extracted posts according to the post types. The amount of extracted post for each post type is portrayed, alongside percentages that detail their usage. The total amount of extracted posts account for 100% of all data retrieved from the online communities.

After extracting all the posts and discovering the amount of posts types, the total count of Comments and Likes for posts within each community is examined. Also, for each post type, the shares count is compared alongside comments and like count. See Appendix for charts, figures 8-19 display the results.

Total Active Users

In order to further understand user engagement, the classification of users is applied where *TAUs* data set is established from the extracted data. The data for user activity within the interaction structures Comment, Like, Comment Reply, and Comment Like is examined. Table 2 shows the total number of users discovered within each structure. The amount of users for each interaction structure is portrayed, alongside their percentages. The total amount of active users account for 100% of all data retrieved from the online communities.

Unique Active Users

When combining the total amount of users from each of the structures, the repetitiveness of users is considered. Since user engagement isn't limited to a single particular field, the Total includes duplicates of users with multiple engagements across the different fields. In order to attain better validity to the Total, the duplicated values are removed and a refined set of Active Users (*AUs*) is established.

The refined set operates under the criterion of accepting a user (u) as active if they are present in at least one of the structures. Thus, Active Users (AUs) = $u \in$ Comments (C), or $u \in$ Likes (L), or $u \in$ Comments Reply (CR), or $u \in$ Comments Likes (CL). This refined data set contains **6535 users** from CS, **29630 users** from HSS, and **13334 users** from VSVHVJK.

Most Active Users

By establishing *AUs*, an investigation into the users most active within all four structures is considered. This is done so as to distinguish users that are consistently active to those

Post Type	CS	HSS	VSVHVJK
Photo	195 *(32%)*	12855 *(34%)*	75 *(17%)*
Status	305 *(50%)*	19768 *(52%)*	128 *(28%)*
Links	104 *(17%)*	4572 *(12%)*	233 *(51%)*
Video	7 *(1%)*	771 *(2%)*	16 *(4%)*
Event	3 *(0.5%)*	163 *(0.4%)*	1 *(0.2%)*
Music	0 *(0%)*	1 *(0.002%)*	0 *(0%)*
Note	0 *(0%)*	1 *(0.002%)*	0 *(0%)*
TOTAL	**614**	**38131**	**453**

Table 1. Extracted Posts

	Comment	Likes	Comment Reply	Comment Like	TOTAL
CS	1021 *(12%)*	6011 *(71%)*	90 *(1%)*	1310 *(16%)*	**8432**
HSS	9106 *(18%)*	27835 *(56%)*	1280 *(3%)*	11461 *(23%)*	**49682**
VSVHVJK	1500 *(10%)*	12394 *(78%)*	43 *(0.3%)*	1920 *(12%)*	**15857**

Table 2. Total Active Users Discovered

that aren't. The *AUs* data set is used as a base of reference for discovering those users found in repeatedly in the Comments, Likes, Comment Reply, Comment Likes structures. Thus, Most Active Users (*MAUs*) = $u \in$ Comments (C), and $u \in$ Likes (L), and $u \in$ Comments Reply (CR), and $u \in$ Comments Likes (CL) The *MAUs* data set then exhibits **55 users** from CS, **1122 users** from HSS, and **17 users** from VSVHVJK.

TAUs, AUs & MAUs

The Total Active Users within each structure of interaction represent the density of user engagement within the chosen online communities. For this reason it is vital to understand how Unique Active Users and Most Active Users scale in comparison to the Total Active Users. Table 3, displays the discovered amount of users for each of the data sets alongside the percentages of AU and MAU to the totality of TAU.

6.2. Network Analysis Results

This analysis aimed to recognize recurring users and attempt to visualize user movement across the similar entities. To accomplish this, the set criterion was to discover at least one networking connection between the three chosen communities.

Common Users

2036 users are established in the Common Users (CUs) data set from the Active Users (AUS) set that contained 49499 users. After identifying the common users within all three online communities, migration patterns through user activity are investigated. For the time period of this study, the results from the migration research are limited to only early migration of users between communities. Early migration for users considers the first time a user administers any type of activity in one of the online communities. The user's activity is then traced from that time stamp to the next online community where they administer their first action. Figure 5, displays the user's early migration between all three communities (VSVHVJK is represented as VSV in the figure).

7. Discussion

In the discussion, connection and comparison of the findings are made toward the research questions.

7.1. Activity Analysis

The activity analysis performed for the chosen Facebook communities looked at the structures of interaction within social networks. The analysis attempts to provide answers for **RQ 1:**

How do the structures of interaction within social media influence the level of user activity in online communities?

Post Types & User Activity

Looking at the post types within each community allows the freedom of understanding what users engage in. The post types show what each community bolsters as significant user interaction. As previously mentioned, Facebook officially categorizes posts into Post Types known as (1)

	CS	HSS	VSVHVJK
TAU	8432	49602	15857
AU	6535 *(77%)*	29630 *(60%)*	13334 *(84%)*
MAU	55 *(1%)*	1122 *(2.3%)*	17 *(0.1%)*

Table 3. TAUs, AUs, and MAUs within each community

```
CS -> HSS -> VSV has : 6 users
CS -> VSV -> HSS has : 24 users

HSS -> CS -> VSV has : 0 users
HSS -> VSV -> CS has : 3 users

VSV -> CS -> HSS has : 78 users
VSV -> HSS -> CS has : 0 users
```

Figure 5. User's Early Migration

Events, (2) Links, (3) Photo, (4) Status, and (5) Video [12]. Posts within the chosen communities are considered to be of the aforementioned five types.

The initial extraction from the communities revealed the existence of additional post types called Music and Note. The posts associated to these post types are rather few and based only in the HSS community, thus making their significance for this study minimal. However, the curiosity factor that follows this discovery implies that further investigation may lead to more information towards the classification of post types within Facebook. This type of further research can increase the scope of how post types can/should be classified when extracting data from social media.

After the initial revelation of the two extra post types, the other posts found are examined so as to discover which post types hold highest amount in respect to totality. Considering the highest ranked post types in these online communities reflects the posts that seem to be successful in getting users engaged. Thus, the higher the total amount of a post types extracted, the better likelihood that users are more engaged in the associated posts. This study discovers that of all posts found within the chosen online communities, Status, Photos, and Links post types have the highest amount in respect to totality. This means that the posts associated to the aforementioned post types are more likely to have more user engagement. Considering this, a look into the counts of user engagement within each of the post types is done.

Looking at the results for Active and Most Active Users we are able to see how engaged users get on posts within each community. Results shown in figures 13-18 (check appendix) represent the share, comments, and likes count for each post type. The total count found for each of the post type show how engaged users get depending on the type of post made. Users actively engage in commenting, liking, and even sharing posts with Status, Photos, and Links types. Considering the results, it's safe to assume that posts with these post types encourage users to engage more than posts with the post types Events, Video, Music, or Note. Additionally, considering the percentage of engagement for each post type shown in Table 1, it seems that posting items with types Photo, Status, and Links seem to influence more user engagement than any other types. Therefore, it seems that the post types Status, Photo, and Links can be considered structures that influence user activity in online communities.

Likes vs Comments

Considering the online communities anatomy as macro-structures of interaction, sub-structures of interaction within these communities are identified. This study looks at Comments, Likes, Comment Reply, and Comment Likes as structures of interaction that generate user engagement.

For study, discovered data shows that Likes exhibit higher user activity than Comments. Subsequently, the results reflect that Comment Likes also hold a higher level in user activity over Comment Replies. In his study, Gerolimos[15] argues the significance of Likes to be lesser than participation via Comments. This study, considers the weight in amount of user engagement within each structure.

It seems that Likes and Comments Likes influence users to engage more than Comments and Comment Replies in these online communities. The significance of Likes and Comment Likes is considered much higher than the other structures. Considering the retrieved data, it's apparent that users engage more with Likes rather than Comments. Thus, when considering the first research question, it would seem likely that users are influenced to interact and engage with the Likes structure than Comments.

7.2. Network Analysis

The network analysis performed for the chosen Facebook communities attempts to visualize user movement across the similar entities. The analysis attempts to provide answers for **RQ 2:**

How can user activity aid in mapping user migration across online communities?

Common Users & Early User Migration

Studies such as, Benevenuto et al [1] and Kumar, Zafarani, and Liu [21] all look into user migration across social media. Considering the previous work done in this area of research, this study looks at user movement from the perspective of user activity and attempts, similar to Kumar, Zafarani, and Liu [21], to recognize some pattern associated with migration.

In order to achieve the intended purpose, the mapping of user movement across communities is heavily dependent on the engagement of users via means of Comments, Likes, Comments Reply, and Comment Likes. The discovery of Unique Active Users (AUs) and Most Active Users (MAUs) gives way to recognizing Common Users (CUs) whose activity data is investigated.The investigation into Common Users *CUs* data set is done by associating [4] values within the established AUs data set in each online community. The *id* value is used as a common identifier between the three *AUs* data sets. By determining the presence of *id* frequent in all three lists, the user profile associated with the *id* is added to the Common Users.

Migration pattern

Utilizing the unique *id* and *name* values in *from*, a *CUs* data set between the 3 communities is established. By following the date of which a user is first active in each community and connecting the same user across other communities, a pattern of activity is discovered (figure 5). This pattern (1) proves the existence of user movement across communities (2) maps out how user migration can be traced through user activity.

For the length of the research to this study, a simple pattern of migration is proven to exist across user communities. Furthermore, this pattern is achieved through investigation of user activity data present for CUs in the structures Comments, Likes, Comment Reply, and Comment Like. Thus, with consideration of the achieved data, it would seem that user movement is indeed achievable through concentration on user activity. This entails that the chosen process as a viable answer to the second research question which looks at how user activity aides in mapping user migration.

7.3 Benefits of study

Given the results from the Activity Analysis, the moderators of within these communities gain a lot by taking into account that which gets users engaged. Considering the discovery of post types and their ranking, it would seem ideal for moderators engage their audience more with posts that are of types Status, Photos, and Links. As shown in figures 13-18 (appendix), there is a much higher count for Comments, Likes, and Shares for these post types in comparison to the others. Furthermore, across the 2 of the 3 chosen communities users are clearly shown to be more engaged in commenting and liking when the post is of type Photo. For CSS and VSVHVJK, moderators can take into consideration the usage of posts that contain more photos so as to get users even further engaged. Given the results form the Network Analysis, the chosen online communities gain a view of how users can move across communities. The feasibility of mapping user migration can help the communities understand (1) if there are users that openly move across communities, (2) how these users are moving, and (3) what their engagement is when moving across communities.

For other online communities, results from this study can serve as aide when considering user activity and the interaction structures influence engagement. Similarly to CS, HSS, VSVHVJK, other communities can benefit by assessing the post types and levels of user engagement when liking, commenting, and/or sharing. Following this type of process, it would seem that community moderators could, with certainty, discover the types of posts that (1) appeal to users and (2) engage users to comment, like, and/or share posts.

8. Limitations and Validity Threats

Recognizing the difference in dates of establishment of the online communities when extracting data is recognized as potential skewer of results. Although relatively close, each community is established at time periods with a year gap in between each other VSVHVJK (2012), CS (2013), and HSS (2014). Another threats to validity include the inadequate research into *Shares* which includes the attribute shared posts. Shared posts reveal the profiles of users that share *page-id/feed* items.

9. Conclusion and Future Work

This study was conducted in order to understand user engagement and migration within online communities. By looking Facebook communities Cabotagestudien, Vi Som Vill Ha Våra Jobb Kvar, and Har Stannar Sverige. This study executes an Activity Analysis which looks at the engagement level of users within those data sets. Additionally, user migration is explored through a Network Analysis so as to identify possible migration patterns that exist between forums.

Through the findings, this study answers the first research question that looks at the structures of interaction and their influence on user activity. Results show that structures of interaction that influence user activity include post types Status, Photo, and Links. Additionally, that other structures such as Likes and Comments Likes influence users to engage more than Comments and Comment Replies in these online communities. Finally, user movement is proved to exist between the online communities with a simple migration pattern.

For the second research question, this study looks at how user activity can be used to aid in user migration across communities. Results show that by considering Common Users (CUs) within each community, it then possible to examine their user activity in structures like Comments and Likes. Within these structures, the time stamps for earliest activity are then collectively assessed in order to create a pattern of migration from one community to the next. Thus, discovering (1) if there are users that openly move across communities, (2) how these users are moving, and (3) what their engagement is when moving across communities.

Future work for this study is to further continue the investigation into shared posts field within *Share* data set, as well as, answering for further user migration. Proper migration research may reveal a better understanding of Common User operations within these online communities. Also, a look into larger communities may reveal a different outcome in the migration of users.

10. Acknowledgments

The author would like to thank my supervisor Håkan Burden for his assistance, patience, and invaluable feedback during the course of this research.

References

[1] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizing user navigation and interactions in online social networks. *Information Sciences*, 195:1–24, 2012.

[2] P. Berkhin. A survey of clustering data mining techniques. In *Grouping multidimensional data*, pages 25–71. Springer, 2006.

[3] D. Boyd and N. Ellison. Social network sites: definition, history, and scholarship. *IEEE Engineering Management Review*, 3(38):16–31, 2010.

[4] M. Brown. Data mining techniques, 2016.

[5] I. P. Cvijikj and F. Michahelles. *A case study of the effects of moderator posts within a facebook brand page*. Springer, 2011.

[6] N. Dey and S. Bhattacharya. Performance analysis of cloud databases handling social networking data. In *Cloud Computing in Emerging Markets (CCEM), 2013 IEEE International Conference on*, pages 1–4. IEEE, 2013.

[7] Facebook. Access tokens. `https://developers.facebook.com/docs/facebook-login/access-tokens/`.

[8] Facebook. Cabotagestudien. `https://www.facebook.com/Cabotagestudien/`.

[9] Facebook. Facebook posts. `https://developers.facebook.com/docs/graph-api/reference/v2.6/post`.

[10] Facebook. Graph api explorer. `https://developers.facebook.com/tools/explorer/145634995501895`.

[11] Facebook. Här stannar sverige. `https://www.facebook.com/groups/1597037703849966/`.

[12] Facebook. Page. `https://developers.facebook.com/docs/graph-api/reference/page`.

[13] Facebook. Pagefeed. `https://developers.facebook.com/docs/graph-api/reference/v2.6/page/feed/`.

[14] Facebook. Vi som vill ha våra jobb kvar. `https://www.facebook.com/visomvillhavarajobbkvar/`.

[15] M. Gerolimos. Academic libraries on facebook: An analysis of users' comments. *D-Lib Magazine*, 17(11):4, 2011.

[16] E. G. Guba. Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Technology Research and Development*, 29(2):75–91, 1981.

[17] D. J. Hughes, M. Rowe, M. Batey, and A. Lee. A tale of two sites: Twitter vs. facebook and the personality predictors of social media usage. *Computers in Human Behavior*, 28(2):561–569, 2012.

[18] T. Keyling and J. Jünger. Facepager. `https://github.com/strohne/Facepager/blob/master/src/help/help.html\_Toc405676326`.

[19] T. Keyling and J. Jünger. Facepager application. `https://github.com/strohne/Facepager`.

[20] M. N. Ko, G. P. Cheek, M. Shehab, and R. Sandhu. Social-networks connect services. *Computer*, (8):37–43, 2010.

[21] S. Kumar, R. Zafarani, and H. Liu. Understanding user migration patterns in social media. In *AAAI*, 2011.

[22] T. C. Lethbridge, S. E. Sim, and J. Singer. Studying software engineers: Data collection techniques for software field studies. *Empirical software engineering*, 10(3):311–341, 2005.

[23] D. M. Levine, M. L. Berenson, D. Stephan, et al. *Statistics for managers using Microsoft Excel*, volume 660. Prentice Hall Upper Saddle River, NJ, 1999.
[24] D. Murthy. *Twitter: Social communication in the Twitter age*. John Wiley & Sons, 2013.
[25] E. S. PKDD., J. M. Zytkow, and M. Quafafou. *Principles of data mining and knowledge discovery*. Springer, 1998.
[26] C. Robson. Real world research. 2nd. *Edition. Blackwell Publishing. Malden*, 2002.
[27] C. Robson and K. McCartan. *Real world research*. Wiley, 2016.
[28] P. Runeson and M. Host. Guidelines for conducting and reporting case study research in software engineering. *Empirical software engineering*, 14(2):131–164, 2009.
[29] P. Runeson, M. Host, A. Rainer, and B. Regnell. *Case study research in software engineering: Guidelines and examples*. John Wiley & Sons, 2012.
[30] M. A. Russell. *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. O'Reilly Media, Inc., 2013.
[31] H. Sternberg, M. Filipiak, E. Hofmann, and D. Hellström. A study on trucking deregulation and cabotage in scandinavia and beyond. Technical report, Lund University, 2015.
[32] H. Sternberg, A. Holmberg, G. Lindqvist, and G. Prockl. A study on the movements of international vehicles in denmark. Technical report, Lund University, 2014.
[33] F. Team. Graph api overview. `https://developers.facebook.com/docs/graph-api/overview`.
[34] R. Team. Restfb. http://restfb.com/.
[35] R. Team. Restfb documentation. http://restfb.com/documentation/.
[36] R. Team. Restfb java documentation. http://restfb.com/javadoc/.
[37] N. Tempini. *GOVERNING SOCIAL MEDIA*. PhD thesis, London School of Economics and Political Science, 2014.
[38] J. Van Dijck. *The culture of connectivity: A critical history of social media*. Oxford University Press, 2013.
[39] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 37–42. ACM, 2009.
[40] R. K. Yin. *Case study research: Design and methods*. Sage publications, 2013.

THESIS PROTOCOL

Objective

What is to be achieved?

Project Objective/Summary: A social network analysis that explores the levels of user activity and user migration within three online communities. Thus, this investigation focuses and classifies research into the two aspects, User Engagement & User Migration.

USER ENGAGEMENT

Factors to consider include:

- Types of interaction structures provided by social network.
- Amount of user activity within each structure.
  - Total amount of active users.
  - The most active users.

USER MIGRATION

Factors to consider include:

- Common users across the online communities.
- Engagement of these users within each structure of interaction.

Research Question

What is to be answered?

Research Question 1: How do the structures of interaction within social media platforms influence the level of user activity in online communities?

Research Question 2: How can user activity aid in mapping user migration across online communities?

Methodology

What approach is to be used?

Exploratory research that explores user engagement within social networks. Data is to be extracted with chosen extraction tools.

Data Extraction Tools:

- Facebook Graph API – for retrieving user activity data from the structures of interaction found within Facebook.
- Facepager Application – retrieving and saving of user activity data to a MYSQL database.
- RestFB – Facebook API for data retrieval.

Figure 6. Thesis Protocol

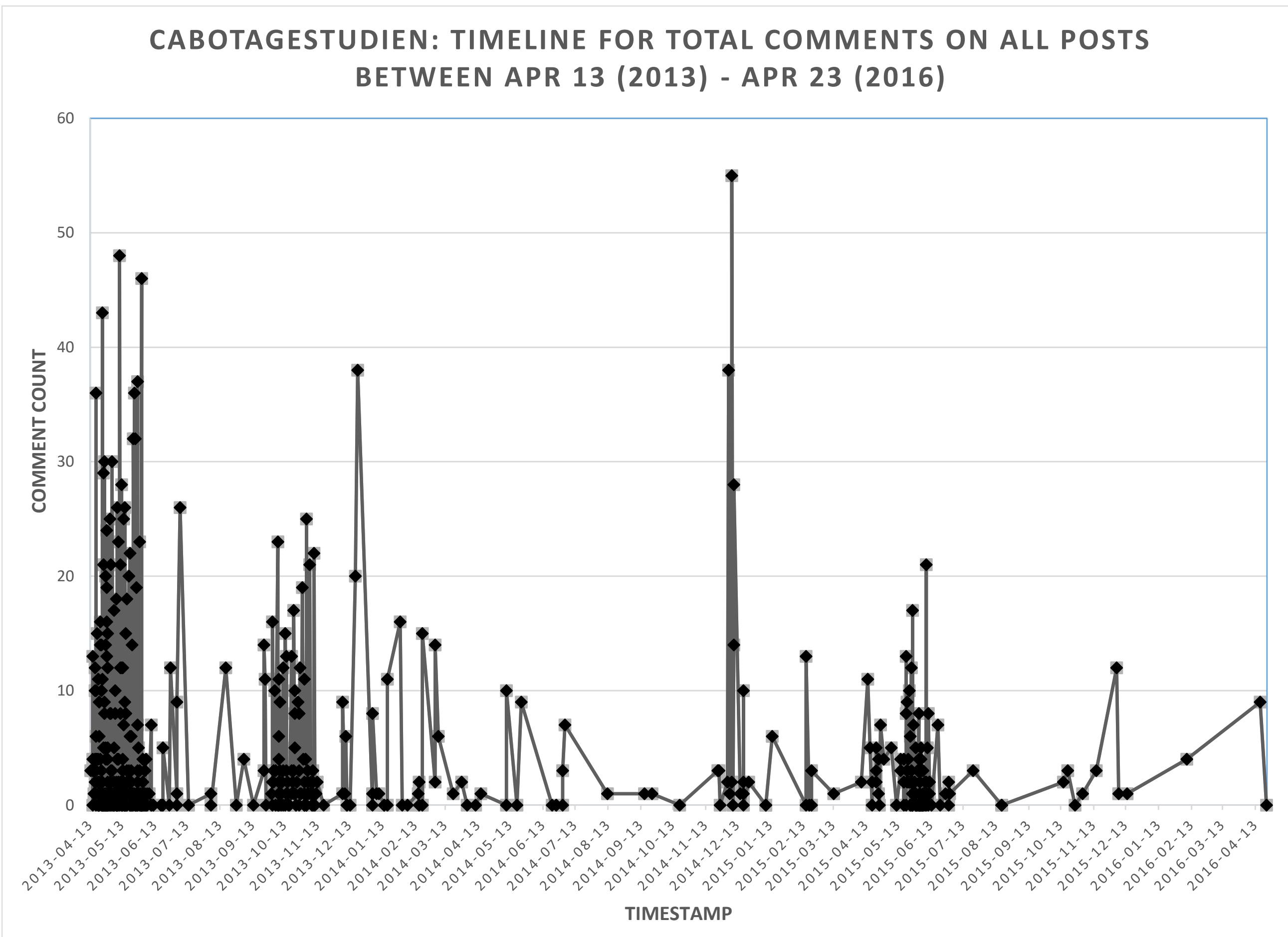

Figure 7. Total Number of Comments for Posts in CS

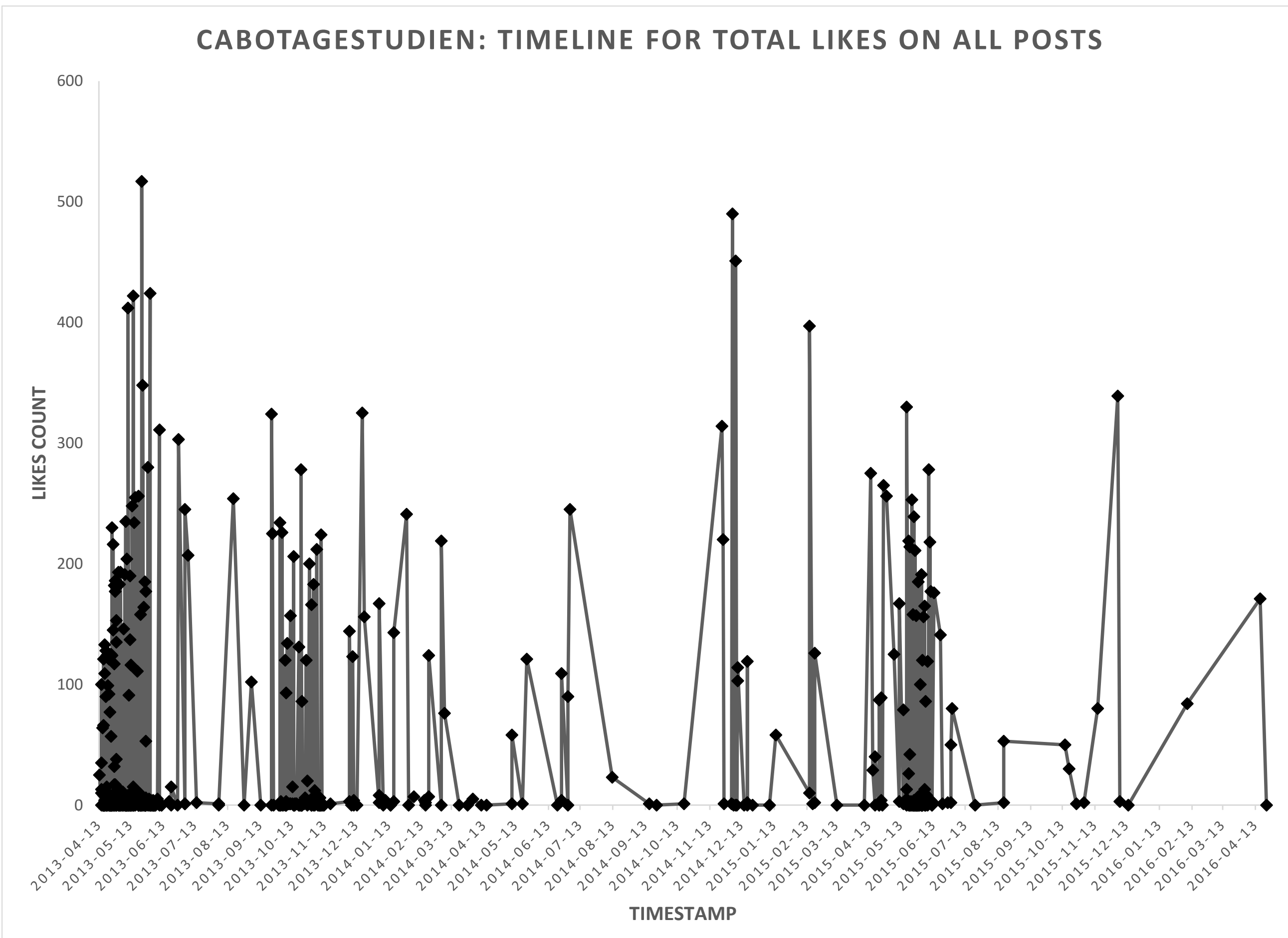

Figure 8. Total Number of Likes for Posts in CS

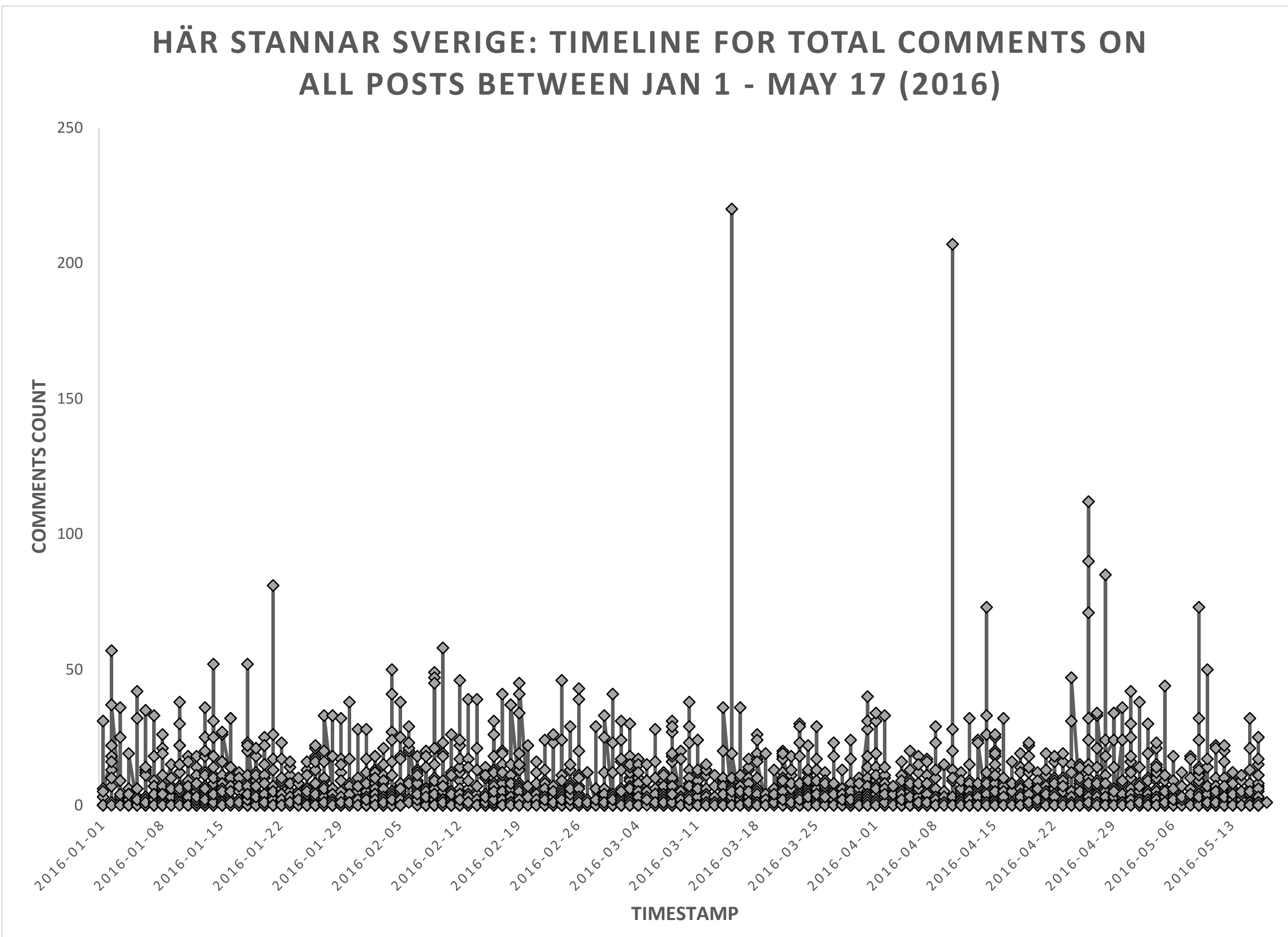

Figure 9. Total Number of Comments for Posts in HSS

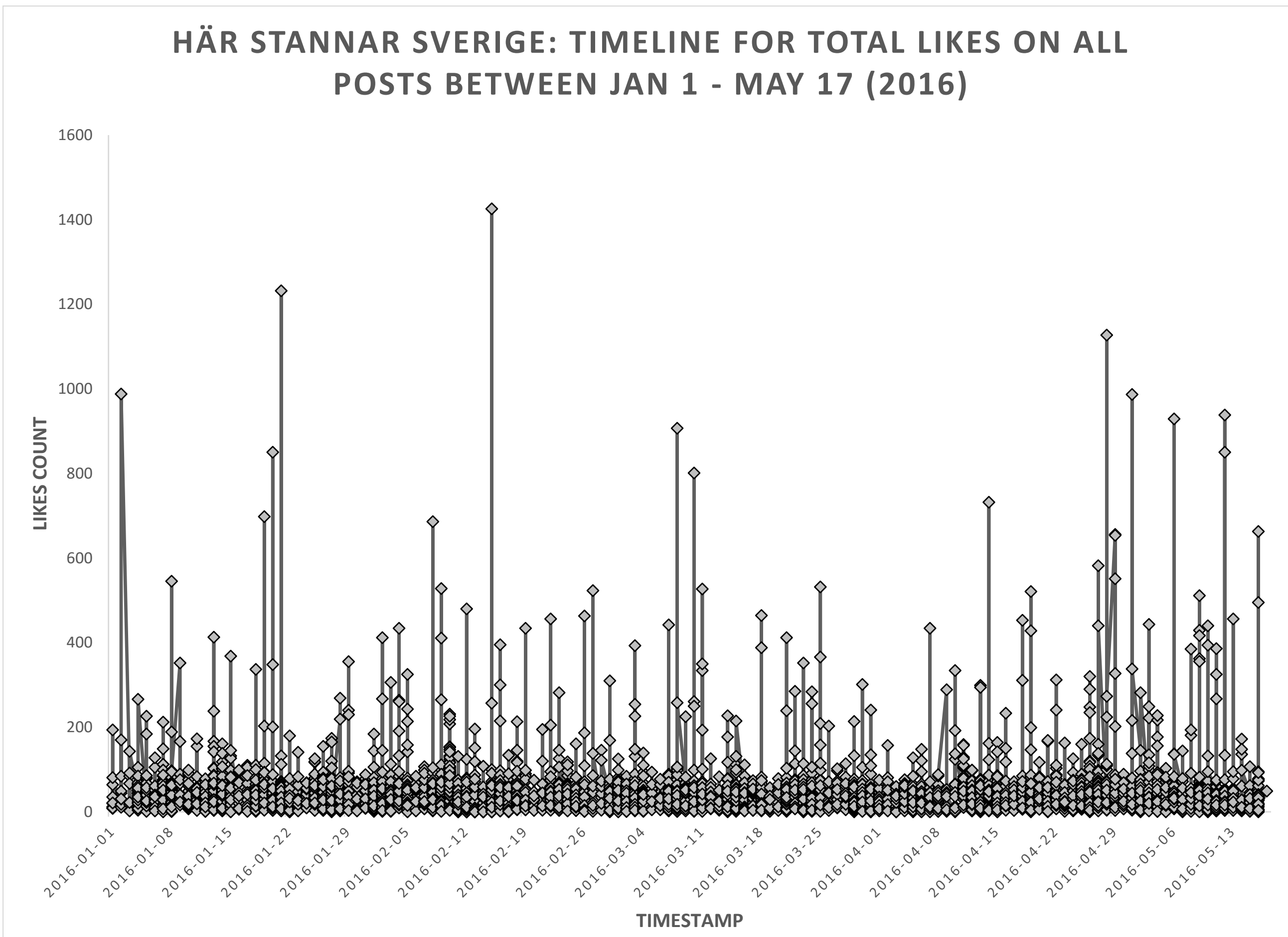

Figure 10. Total Number of Likes for Posts in HSS

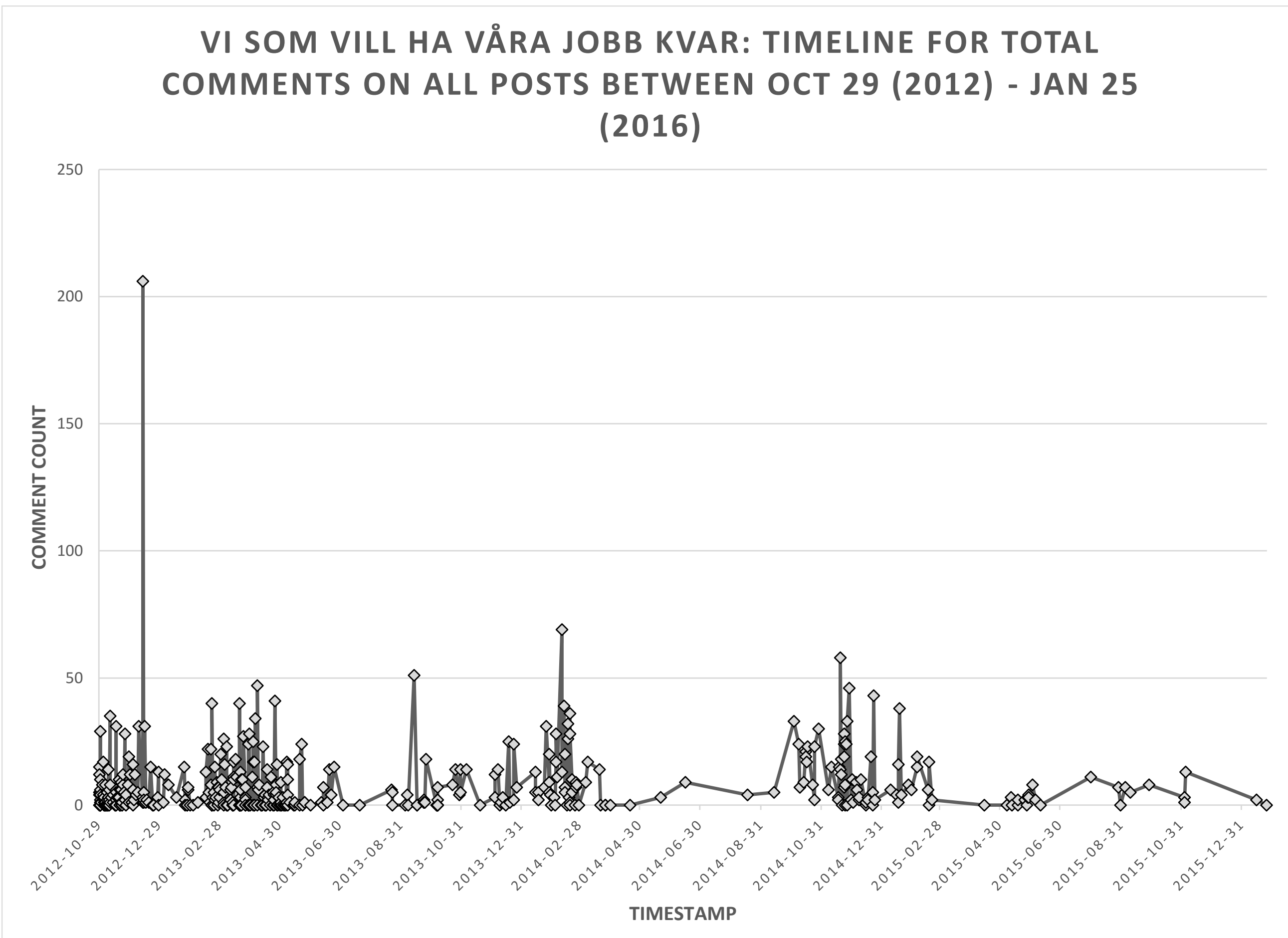

Figure 11. Total Number of Comments for Posts in VSVHVJK

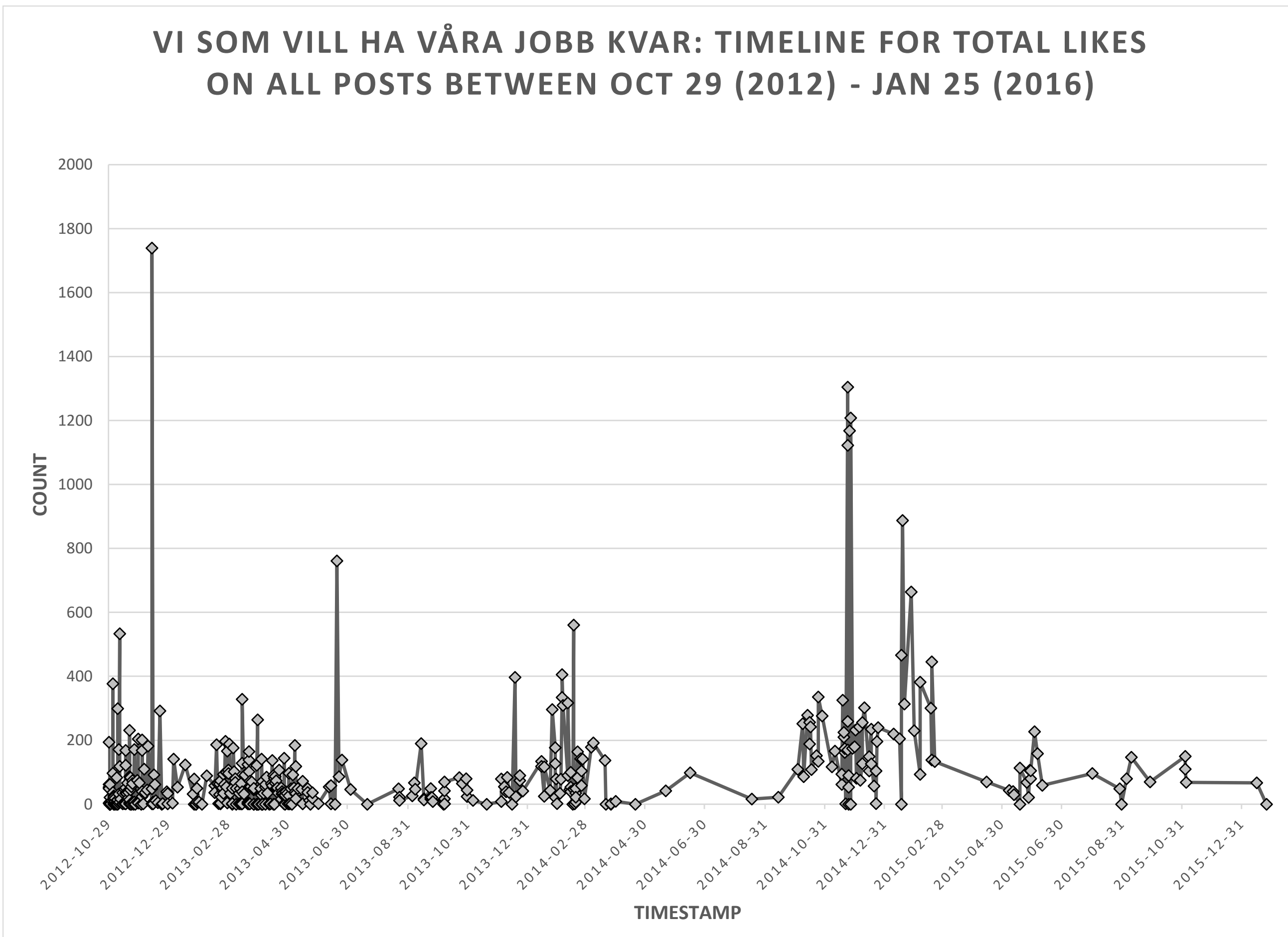

Figure 12. Total Number of Likes for Posts in VSVHVJK

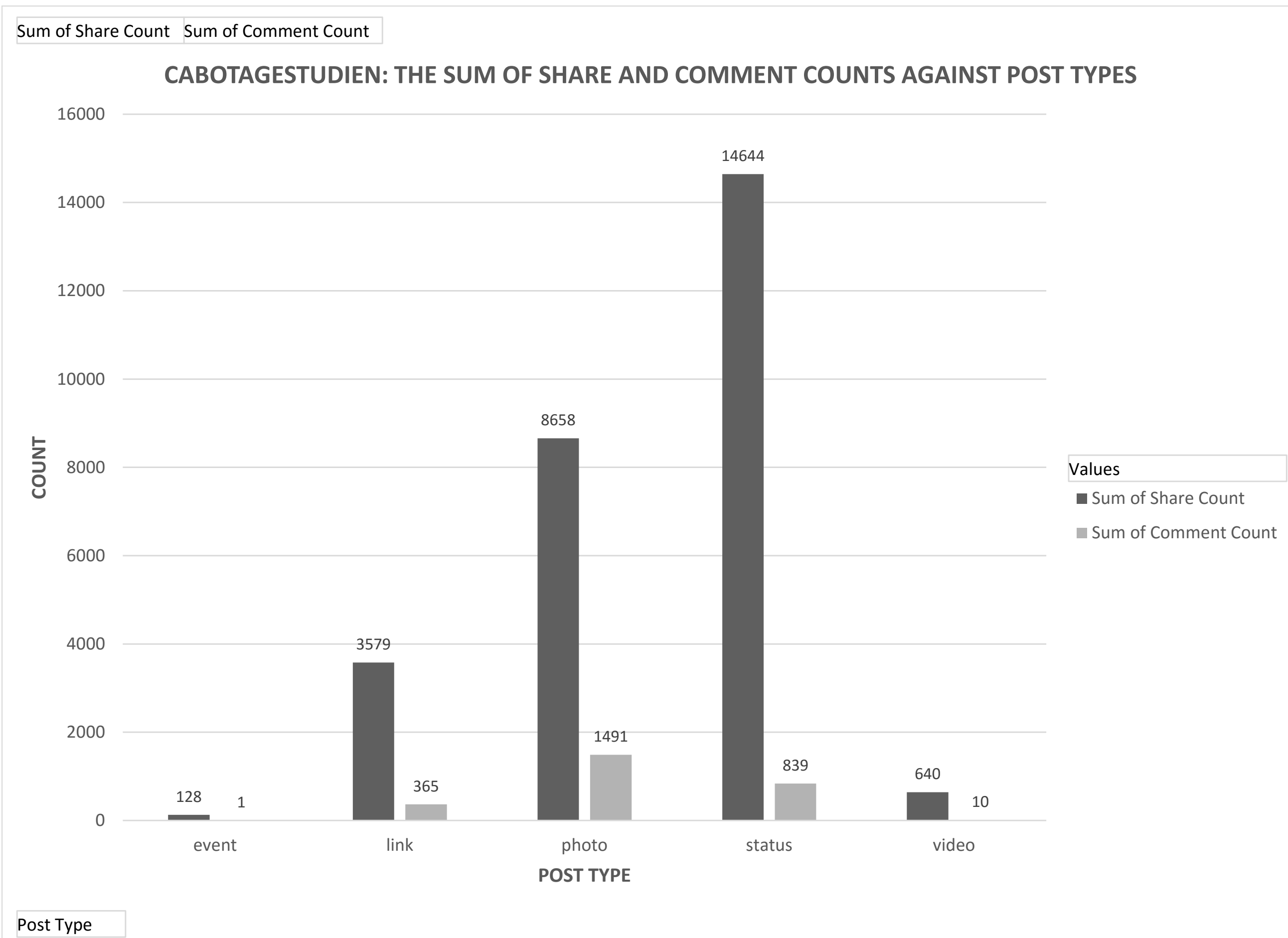

Figure 13. Post Type: Share Count x Comment Count CS

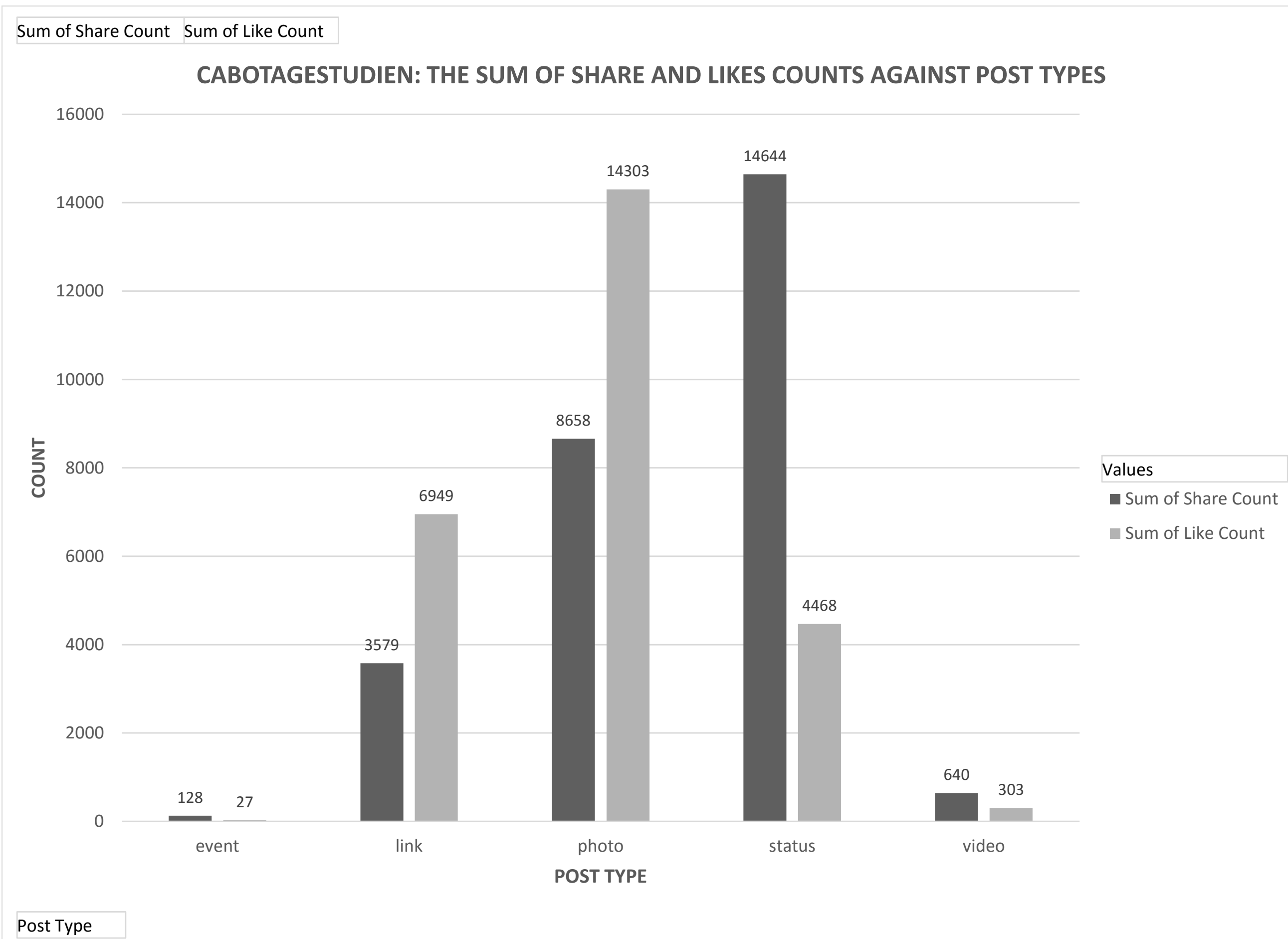

Figure 14. Post Type: Share Count x Likes Count CS

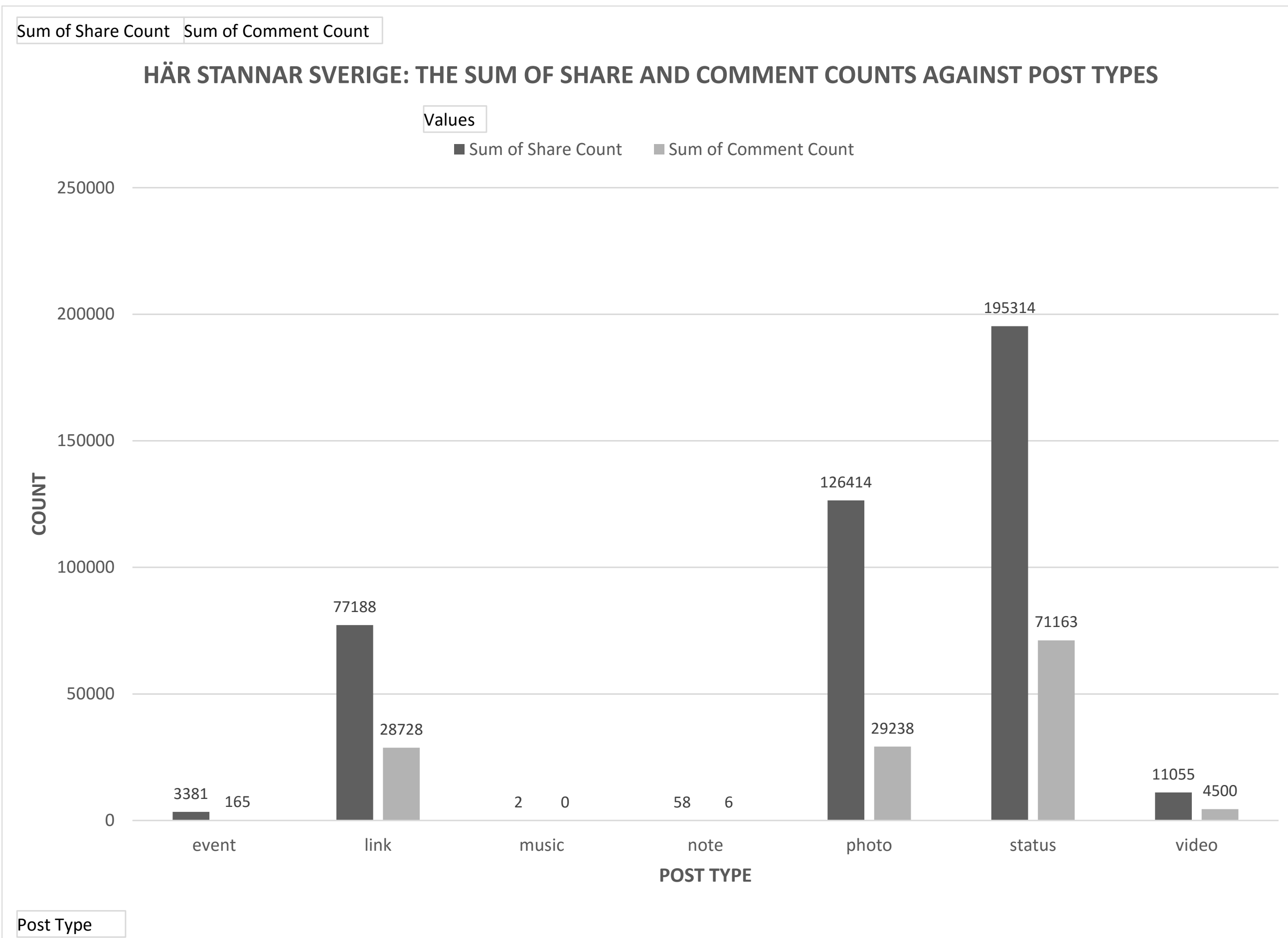

Figure 15. Post Type: Share Count x Comment Count HSS

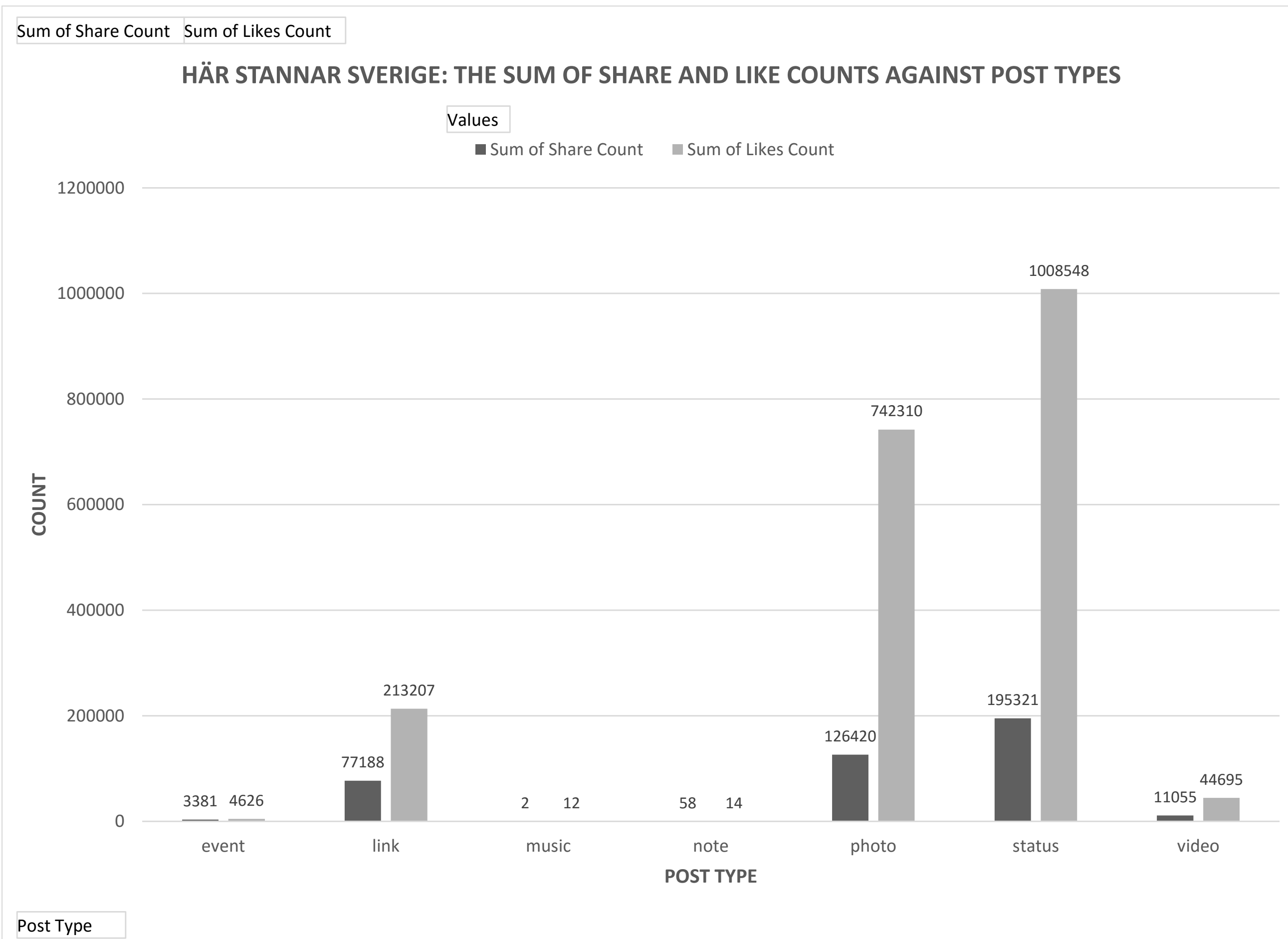

Figure 16. Post Type: Share Count x Likes Count HSS

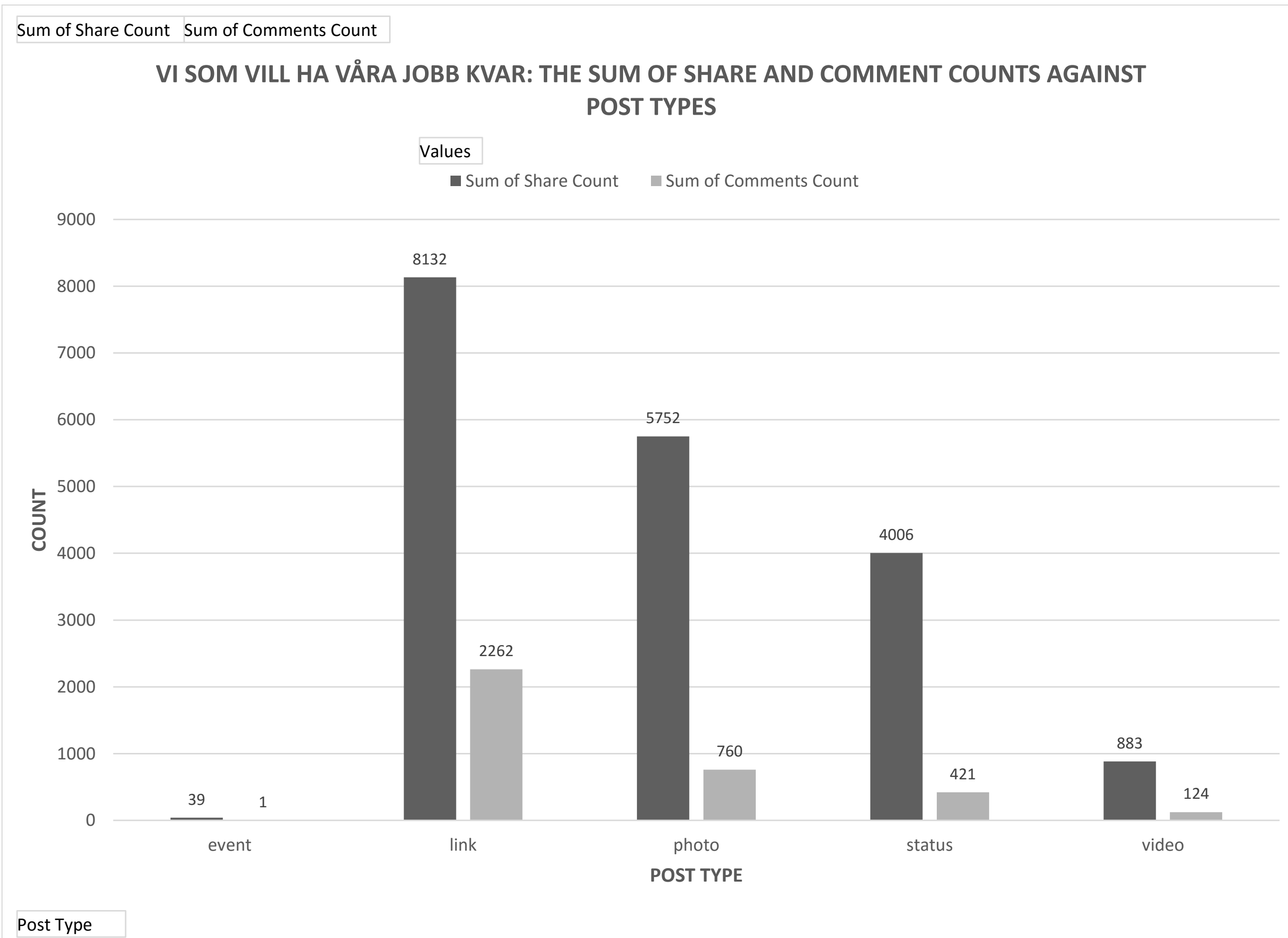

Figure 17. Post Type: Share Count x Comment Count VSVHVJK

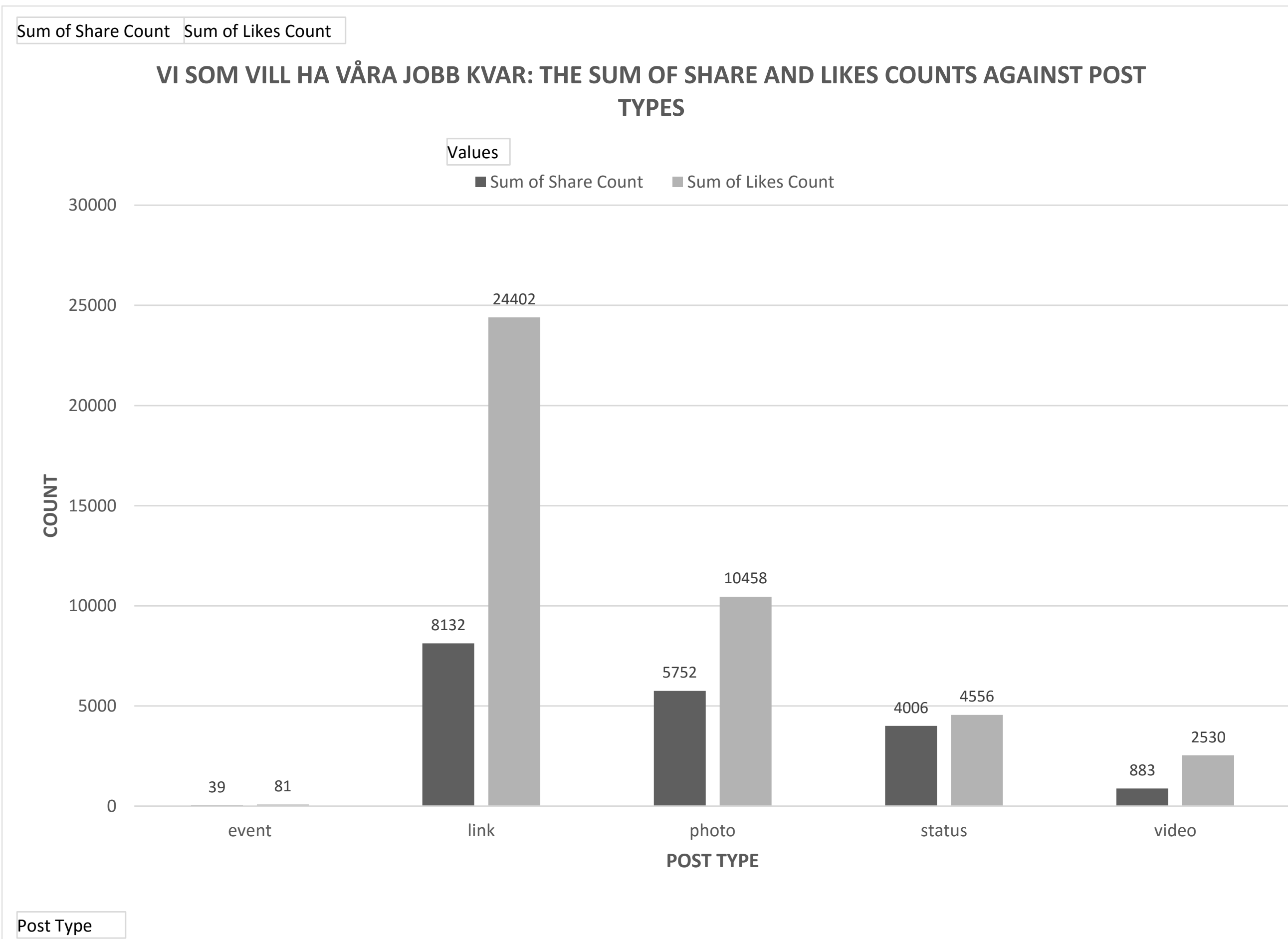

Figure 18. Post Type: Share Count x Likes Count VSVHVJK