The Influence of Community Interactions on User Affinity in Social Networks: A Facebook Case Study

Authors:

M. Senaweera	14020752
M.M.N.D. Chamindi	14020084
S.A.D.Y.R. Dissanayake	14020191

Supervisor:
Dr. C.I. Keppitiyagama

Co-Supervisors:
Dr. M.I.E. Wickramasinghe
Dr. P.V.K.G. Gunawardana

This dissertation is submitted to the University of Colombo School of Computing in partial fulfillment of the requirements for the Degree of Bachelor of Science Honours in Information Systems.

University of Colombo School of Computing
35, Reid Avenue, Colombo 07,
Sri Lanka

June 2019

Declaration

I, M. Senaweera (14020752), hereby certify that this dissertation entitled The Influence of Community Interactions on User Affinity in Social Networks: A Facebook Case Study is entirely my own work and it has never been submitted nor is currently been submitted for any other degree.

Signature:
Date: Saturday 8th June, 2019

I, M.M.N.D. Chamindi (14020084), hereby certify that this dissertation entitled The Influence of Community Interactions on User Affinity in Social Networks: A Facebook Case Study is entirely my own work and it has never been submitted nor is currently been submitted for any other degree.

Signature:
Date: Saturday 8th June, 2019

I, S.A.D.Y.R. Dissanayake (14020191), hereby certify that this dissertation entitled The Influence of Community Interactions on User Affinity in Social Networks: A Facebook Case Study is entirely my own work and it has never been submitted nor is currently been submitted for any other degree.

Signature:
Date: Saturday 8th June, 2019

I, C.I. Keppitiyagama, certify that I supervised this dissertation entitled The Influence of Community Interactions on User Affinity in Social Networks: A Facebook Case Study conducted by M.Senaweera, M.M.N.D. Chamindi and S.A.D.Y.R. Dissanayake in partial fulfillment of the requirements for the degree of Bachelor of Science Honours in Information Systems.

Signature:
Date: Saturday 8th June, 2019

I, M.I.E. Wickramasinghe, certify that I supervised this dissertation entitled The Influence of Community Interactions on User Affinity in Social Networks: A Facebook Case Study conducted by M.Senaweera, M.M.N.D. Chamindi and S.A.D.Y.R. Dissanayake in partial fulfillment of the requirements for the degree of Bachelor of Science Honours in Information Systems.

Signature:
Date: Saturday 8th June, 2019

I, P.V.K.G. Gunawardana, certify that I supervised this dissertation entitled The Influence of Community Interactions on User Affinity in Social Networks: A Facebook Case Study conducted by M.Senaweera, M.M.N.D. Chamindi and S.A.D.Y.R. Dissanayake in partial fulfillment of the requirements for the degree of Bachelor of Science Honours in Information Systems.

Signature:
Date: Saturday 8th June, 2019

Acknowledgements

We would like to take this opportunity to express our gratitude to the people who have lent their support all along with this thesis and thus enabled me to concentrate on our work.

Foremost, we would like to express our sincere gratitude to our advisers Dr. C.I. Keppitiyagama, Dr. M.I.E. Wickramasinghe and Dr. P.V.K.G. Gunawardana for the continuous support of our final year research, for their patience, motivation, enthusiasm, and immense knowledge. Their guidance helped us in all the time of research and writing of this thesis.

Besides our advisers, we would like to thank Mr. Primal Wijesekera, Mr. Sameera Horawalavithana, Mr. Charith Elvitigala and Mr. Anupa Shyamlal for their insightful comments, hard questions, immense support and encouragement.

We thank our 4th year fellow mates, for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last year.

Last but not the least, our families and friends for continuous support and encouragement throughout the year.

Abstract

With the advent of social media, one of the biggest concerns have been its suspected impact on democracy by influencing users' opinions during elections. Impact of the social networks (and/or media) has been the center of that discussion due to well-known cases of using social networks to sway people's opinion in events crucial to the democracy such as the election. The accusations vary from the propagation of fake news to concerns about Facebook having unfettered power over its users' content.

As a step towards understanding the true nature of social media's influence, we started collecting data prior to Local Election in Sri Lanka, 2018 and continued the collection until the election was over; we collected the data through Facebook API and all of the data is completely anonymized. The dataset covers 44K users from Sri Lanka and their interactions with 44 Facebook groups. The dataset also includes all of the associated events related to each group. As a preliminary step, we have analyzed how the user affinity surrounding these groups have changed surrounding the period of the local election. Our analysis also gives a concrete and quantitative evidence of how users are communicating on Facebook and how active they are surrounding a sensitive event.

Further, there is a significant change of affinity of a set of individuals over the time corresponding to the external event. Additionally, we find that there are users migrating among the Facebook groups during the election period. Further, we show that when the migration network, viewed as a weighted network, displays features which are significantly different from a comparable random network. In particular, we show that user migrations within Sri Lankan political groups during the election period show a non-random behavior potentially motivated by the real world events.

We believe our analysis is the first of its kind in providing scientific analysis of the influence of social media in Sri Lanka.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Social networks, one of the most controversial medium for information dissemination and communication is believed to influence people's lifestyles and perceptions both directly and indirectly [1]. The users are exposed to a variety of content such as personal write-ups, images, videos, GIFs and many more shared within their own personal network of users. The continuous interaction with these content is believed to have influenced the perception of users even though there is few substantial evidence backing this belief [2]. There were many uproars around the world (Sri Lanka is a good example of the actions taken to ban several social networks and communication platforms due to racial riots [3]), on the mere belief that users' perceptions are changed with the content of online social networks.

It is important to identify and quantify the actual impact posed on the users by interacting with the content in social networks. As a fundamental step in reaching the aforementioned goal, we look into identifying the changes in user affinity of the users corresponding to content interaction within a social network. This initiation opens doors to a large diversity of research in the field of influence on users in social media. Furthermore, it enables to educate the society on the impact of interacting with the content of social networks towards them.

1.1 Background of the Problem

The recent advancements in technology brought about various social networks facilitating the society for communication [4], resource sharing, networking and more. Facebook, Twitter, Instagram, Pinterest are few examples of such social networks. Out of all socially available networking sites, Facebook is ranked as the most popular social networking site as of January 2018 with a user base of 2,167 million [5]. A user on Facebook is able to set up a new profile and add their personal information with required privacy settings. Thereafter, a bi-directional connection [6] can be established within two users called a friendship. A single user can have a variable number of friendships with the other users forming a network of friends around the user. A Facebook page allows entities, figures, businesses and organizations to be authentic and public whereas a group brings together individuals with common interests to share their opinions [7].

An individual user has the capability to share their opinions and ideas as a post on their personal profiles, pages or groups. The post owner and the other users get to react or comment to the content of the post with regard to the privacy settings of the user profiles or the settings of a group or page. A user can react to the post in five ways; like, love, haha, wow and sad. One user gets to react with only one reaction for a single post but can comment and reply to comments as many times as required. A post can be shared. Thereby a user can interact with the content in a social network through liking, commenting or sharing the content.

It is evident that Facebook has provided its community with a large array of features to express their ideas and opinions freely. Many entities and individuals in the society use this platform for communication, networking, marketing and campaigning [8]. However, with the power of features provided by Facebook, individuals or groups of individuals have found ways and means to negatively affect its community. The Presidential Election of the United States in 2016 is believed to have deeply influenced by the fake news propagation within Facebook and other social networking platforms [9, 10]. Another example is the Brexit election [11]. These topics came into attention within the society because people believed that the content on Facebook or other social networking sites did influence its users' affinity. Another example would be the Presidential Election 2015 of Sri Lanka. Many people believed that Facebook was able to sway the mindset of Sri Lankan users and entirely affect the results of the election. This election was called the 'first-ever cyber election of Sri Lanka' [12]. In all these events, there is a general belief that the results of the election were most likely due to the influence posed on the user affinity.

What is *affinity*? It is the tight bond a consumer holds towards a particular brand [13]. A brand represents a business, organization, product, service or even political parties. In the context of this research study, we refer to the affinity or the tight bond of the user of social networks holds towards political parties within the network. A user either has a high or low affinity towards a political party. There can be changes to their level of affinity due to internal or external factors. These affinity changes in political views of the users due to the interactions in social networks is one of the most talked and controversial problems speculated by society. There is no ground truth or evidence to prove this speculation. However, the society hangs tightly onto it. People believe social networks can be used to change the affinities of its users, especially in the political domain through Facebook or other social networking sites. Various election candidates of different countries carried out large-scale political campaigns on the mere belief that affinities can be varied through user interactions with the content [14]. Sri Lanka is capable of providing another example for this scenario, the ban on several social networking sites like Facebook and communication platforms implemented by the Sri Lankan Government for nearly three days due to racial riots in March 2018 [3]. This ban was imposed because the Government officials believed that the online user interaction could influence user affinity hence limited their interactions through the ban. There was no evidence that backed up the claim made by these officials. It is clear this issue has outgrown to lengths such that it has started to affect the democracy of countries as well. Yet, no concise explanation or justification provides substantial information proving that online user interactions do influence changes in the user affinity.

Facebook as a whole is a complex and sophisticated platform. It consists of personal profiles, groups and pages. We narrow it down to Facebook groups. It is notable that groups are places where people with common interests come together to share their viewpoints [7]. In a group, we can find three kinds of users according to their interactions or consumption of the content. There are active contributors, passive users and consumers of content [15]. These users consume a considerable amount of content on a daily basis. People believe that this exposure to data can change the affinity of a user. Thereby, it is important to identify the actual impact of the user interactions towards their own affinity.

The existing literature does not provide justice to the presented area of research. Many research conducted in the social networks field concentrate on areas such as fake news propagation [9, 10], influence maximization, content personalization [16], online user behavior [17], sentiment analysis [18] and more. Very few research is available where it explains the effect on a user affinity due to the impact of interacting with the online content. People need to know the real impact as we see many decisions and actions are being solely taken on the belief that online interactions change the user affinity. The decision-making process should be informative and use accurate facts and figures. We believe this research enables to create the pathway in recognizing the effect of user interactions.

A crucial measure for this study is a measure to quantify user affinity. Affinity is an intangible abstract quality, therefore, it cannot be directly measured from the network. Previous research [19] have considered secondary measurements obtained from a graph of the network such as the shortest path length of the users and the number of edge-disjoint paths between users as a measure of closeness. However, it is doubtful whether these measures are good enough to quantify affinity. To this end, we look at the migrations of users within the Facebook groups. The rationale here is that most user migrations are a result of a prior change in affinity. A migration happens when a user supporting one entity and interacting with the online content of the given entity's group starts interacting with an entity's content which is contrasting to the ideas and opinions of the initial entity.

It is interesting that approximately 82% of Sri Lankan internet users are using Facebook. It is the most used social media platform in Sri Lanka according to the Statcounter statistics [20]. We collected data on user interactions from 44 Facebook groups where the topic is dominantly politics with regard to the Local Government Election 2018 held on the 10th of February 2018. The dataset contains data from both the pre-election and post-election period. All the personally identifiable information is removed from the dataset. We study this data in three phases; micro-level analysis, macro-level analysis and user content coding. The micro-level analysis is where every single user is considered as the unit of study. Facebook groups are used as the unit of study in macro-level analysis. User content coding is conducted to evaluate the results. At each of the analysis stages, we look into changes of user affinity through user migrations via a model introduced in this research.

1.2 Problem Statement

The society looks at the social networks with the mere belief that it influences the users with its content and accomplish to change their user affinity regarding political matters. Past research has not addressed this particular area of study adequately. There exist many pieces of research which have look into influencing a person to interact with the content, personalizing content on social networks, fake news detection and many more aspects of social media [9, 10, 16–18]. However, the influence on the user's affinity after interacting with content is scarcely studied. It is observed that decisions and actions were made blindly based on the mere belief, in the absence of clear or incontrovertible evidence. We observe a clear gap of knowledge that needs to be analyzed and understood precisely. This study is of high significance as understanding the influence on the user affinity by interacting with the content alone provides ways and means to look at all the opportunities and threats that it poses on the society. The outcome of the research will assist in various decision-making processes of authorities and entities.

1.3 Research Questions

Our research is directed to analyze the impact on user affinity exerted by the interactions of the users within a community in social media. With this intention, we have planned to answer the question, **how does the community interactions on social media influence the user affinity?** . This finding will be vital as it will verify many ideologies held by the public.

We look into identifying the relationship between the community interactions and the user affinity.

In the process of filling the gap of knowledge and getting a sound understanding of the influence of community interactions on the user affinity, it is important to precisely understand about the user affinity. Identifying the affinity changes is one of the most critical tasks that is achieved during the study. This quantification assists to get a broader view on the impact of community interactions on the user affinity. The user affinity quantification also becomes useful since it allows to derive a precise depiction on its fluctuations with regards to interactions of a user with various content available in the communities of social networks. Therefore, the sub research question, **how to measure the user affinity changes in relation to community interactions on Facebook?** is answered through this research.

After the identification of a method to measure user affinity changes, it is possible to look at its evolution with time in the context of a given social phenomenon. We do not believe it is possible to generalize the findings for all the phenomena but able to provide evidence of what happens to the affinity with relation to the interaction of content on social networks. It is possible to generalize the method used in capturing the evolving user affinity. Thus the sub research question, **how does the online user affinity evolve corresponding to social phenomena depicted in Facebook in association to community interactions?**, is answered. Understanding the fluctuations of the user affinity with time helps to weigh the effect of community interactions and identify the ways it affects various individuals.

1.4 Goals and Objectives

1.4.1 Goal

Identify the influence of community interactions on user affinity in social networks

1.4.2 Objectives

- Analyze the evolving user affinity due to interactions in Facebook groups
- Identify a metric for user affinity in relation to the community interactions on Facebook
- Create a model to identify the changes in user affinity

1.5 Significance of the Research

During the racial riots that happened in 2018 [3] in Sri Lanka, the government restricted access to social media without substantial evidence under the assumption that social media fueled these riots. While we are not suggesting that such actions are ineffective, having insights on the impact of social interactions on user affinity will assist to make informed decisions as the macro-level can be understood with the information.

People tend to use social media as a platform for manipulative marketing where they try to influence people by boosted posts or else fake likes. Having insights into human behavior within the social media platform, we have a single source of truth to identify this kind of manipulative marketing or campaigning at early ages. This identification can be used as an indirect way of determining the authenticity of the information.

This research lays the foundation for various other research as we contribute knowledge on user affinity and community interactions which is scarce in the research community.

1.6 Research Approach

The research will follow a nonexperimental descriptive research design approach using quantitative methods which will bring out effective and reliable results for the study. We plan to follow nonexperimental results as the data collection will be carried out on Facebook from public groups and pages which no manipulation of any sort can be done [21, 22]. In this research we intend to study the relationship between the social interactions of online social network users and their user affinity, it is evident that the purpose of the research is descriptive [23–25]. However, the data that we collect is qualitative in nature, but we will use quantitative methods to analyze it using graph theory and statistical methods to reap the best results.

We will use the Local Government Election 2018 of Sri Lanka held on $10^{\text{th}}$of February 2018 as the case study to get data for analysis[26]. The data on the Facebook user communities and the interactions within the time period, 1^{st} of December 2017 to the 20^{th} of February 2018 will be collected. The data collection spans over few weeks prior to the election and a few weeks after the election is over with the intention of getting a complete dataset with pre-election and post-election activities.

The dataset will consist of community interactions within public groups and pages available on Facebook where the election was the main topic of the discussions. Primary data of the research was collected through Facebook API[27] extraction where it generates basic data relevant to community interactions such as user ID, post ID, comments, number of likes and number of shares of a particular post etc.

1.7 Limitations, Delimitations and Assumptions

1.7.1 Limitations

The dataset does not include all the Facebook user interactions that happened during the time period the data was collected. Some Facebook posts were deleted by users or page/group admins or reported and taken down by Facebook[28] which resulted in creating gaps in the dataset during the data collection. Presence of this deleted data might be able to provide more useful and better insights for the analysis. We consider this as a major limitation of our research.

It is certainly possible that the factors that drive the changes in user affinities were caused by user interactions that happened outside of the time period that the data was collected. Further, it is possible that some users do not tend to interact with Facebook groups even though they were influenced by the content. This research limits itself to analyze the online user interactions that happened within a selected number of Facebook groups from 1st of December 2017 to the 20th of February 2018.

1.7.2 Delimitations

We do not analyze the content of the Facebook posts or comments even though they were extracted from the selected groups/pages. There is related research that was carried out by means of content analysis[29], however, it is deemed rather difficult in this research due to the nature of the content in posts and comments. Majority of the post and comment content is in Sinhala or mixture of Sinhala and English.

Sinhala, being a morphologically rich language, makes it harder to apply the Natural Language Processing techniques with sentimental analysis that are available to content written in English. There is a lack of tools and techniques tailored to Sinhala content analysis in the research community. To this end, we chose not to include automated content analysis as another path on this research.

1.7.3 Assumptions

The time a particular interaction occurred is required for temporal analysis of user interactions. While it is possible to get the time a certain comment is made on a post, Facebook API does not provide timestamps for reactions. Due to this limitation in the Facebook API, we assume that the reaction timestamp is equivalent to the timestamp of the published post the reaction was made to.

1.8 Contributions

We fill an existing knowledge gap by contributing the following,

- To the best of our knowledge, we are the first to analyze how user affinity evolves surrounding a key event such as an election in Sri Lanka.
- To the best of our knowledge, we are the first to analyze user interactions among Sri Lankan Facebook user groups and pages surrounding political topics.

We have contributed to the research community with two publications as of the writing of this thesis.

- "A Weighted Network Analysis of User Migrations in a Social Network" Presented at 2018 International Conference on Advances in ICT for Emerging Regions (ICTer)
- "The Influence of Community Interactions on User Affinity in Social Networks: A Facebook Case Study" Presented at 2018 National Information Technology Conference (NITC)

Further, we deliver the following to the research community for future research in this area,

- An anonymized dataset of Sri Lankan Facebook public groups and pages during Local Government Election 2018
- Scripts to collect data from Facebook groups and pages
- An open-source toolset to analyze and visualize evolving user interactions in Facebook

Chapter 2

Background

2.1 Social Influence

Social influence as defined by Rashotte is "the change in an individual's thoughts, feelings, attitudes, or behaviors that result from interaction with another individual or a group" [30, p. 4426]. Contextualizing social influence is a tedious task since it is necessary to take into account that there are many sources and targets of influence and time [31]. It is important to understand that social influence is multi-directional and dynamic. In the book [32], Fogas and Williams have written, they have described influence to be rather indirect than direct since at times, individuals are not even aware of been influenced by someone or something.

It is observed that social influence plays a major role in social networks today. Many researchers have studied the social influence and incorporated its properties to derive various applications. One of the major applications is the influence maximization. In a network of a given number of nodes, influence maximization identifies the set of nodes that set the maximum influence within the network provided that both the Linear Threshold and Independent Cascade Models are satisfied by the nodes [33]. This concept has been widely used in the fine-tuning viral marketing [34, 35] and heuristic algorithms [36]. In the study conducted by Bond et al. the impact online social influence on offline behavior was investigated [37]. Many concept models have been produced with regard to the social influence and its impact on social networks online [38, 39] and offline [40].

We can see that there are various types of models and algorithms are present to identify the influential nodes. This study emphasizes understanding the factors that cause these nodes to be influential. It is important to look at the microscopic level of the network to identify the factors. Techniques such as actor-oriented models, statistical analysis such as cumulative distribution function and pattern recognition can be used to investigate these factors.

2.2 Social Affinity

Social affinity can be explained in various ways as discussed by Vela-McConnell in his book [41]. Affinity can be explained as either resemblance or fondness or imply kinship with individuals. A research was carried out by Panigrahy et al. where they introduced a method to quantify the social affinity between users in a social network which intermediary to both the shortest path and number of paths. In order to understand the mechanism of the viral marketing, affinity was used by Irribarren and Moro [42]. They explained that the affinity between the message content of the campaign and personal preferences drove the mechanism. However, in our research, we plan to measure the user affinity of a user with the political entity of their preference. It is not possible to use measurements such as shortest distance or number of paths to the entity in our dataset since the edges are multidimensional as it consists of many properties within itself. Thereby, we had to include edges and node properties together with temporal data and network communities in order to quantify the affinity of the users.

2.3 Social Network Analysis (SNA)

A social network can be online and digital networks or face-to-face conversations, political associations and connections within or outside nations, and economic transactions within businesses or organizations [43]. Social Network Analysis is the process of investigating social structures through the use of networks and graph theory according to Otte et al [44]. It is interdisciplinary as it brings together social theory and its applications, statistical and mathematical disciplines and computing methodology. The process of social network analysis is not merely looking at pictures and diagrams of networks to identify patterns but also including a network's structural properties and its implications. SNA uses relational data; it requires data of the relationships between the analysis units to proceed with the analysis. In the process of analyzing networks, as John explained it is important to consider the following pointers [45];

- The study requires relational or structural data
- The actors and their actions are interdependent
- Network models targeting individuals are seen as opportunities or constraints for an individual's action in the structure
- Network models conceptualize its structure using the relationships between its actors

SNA has been used in many areas of work to understand, describe and explain networks;

- Migrant behavior [46]
- Textual analysis applications [47]
- Internet applications [48, 49]
- Computer-supported collaborative learning [50]
- World political and economic system [51]
- Social Influence [52]
- Spread of Obesity [53, 54]
- Social Capital [55, 56]

Network analysis can be represented mainly using three forms [45];

- Graph Theory: Graph theory is used in this analysis [57].
- Sociometric: It is used in studying equivalence and block models. The actors are represented in a two-way matrix called a sociomatrix. Most of the software use this method in analyzing network data [58].
- Algebraic: This method is used in role and positional structure and relations within a network by using various algebraic techniques. It is useful in looking at multiple relations [59].

Graph theory has been widely used in studying network models. One of the appealing reason is the vocabulary that provides a precise explanation to social structural properties [45, 60]. It uses mathematical and statistical approaches so that an objective view of the network can be obtained. In addition, graph theory presents ways to prove and justify theorems. It simply provides an easy way to represent a network and quantify its properties without complications.

2.3.1 Networks and Basic Representations

In the literature, networks have been treated as a static structure where all the calculations were derived from artificial graph structures. The advent of the World Wide Web has resulted in rapid growth of social networks and the application of graphs to the real work scenarios [61]. It is important to take into consideration the following measures when experimenting with real work models at present:

- The inclusion of both the theoretical and empirical data to structure the graphs
- Non-bias data collection methods

In light of the current development of social networks, it is important to understand the evolving and dynamic nature of networks. In social networks, nodes and edges keep getting added or removed with time, thus it is mandatory to capture the temporal changes taking place in the network when studying it. Furthermore, the properties affecting the network structure should be thoroughly studied.

A network is a collection of vertices connected to each other, which is also called a graph in the mathematical literature [62]. The vertices (V) are called nodes and the connections within the nodes are called edges (E). Thereby, a graph G) can be represented as $G = (V, E)$.

Few of the metrics that are used in social network analysis can be listed as follows.

- *Degree* is the number of edges connected to a vertex.
- *Degree sequence* is the summary of the sum of edges per vertex in the graph.
- *Degree distribution* is the probability distribution of the degrees over the network.
- *Clustering coefficient* is a measure of the degree to which nodes in a graph tend to cluster together.
- *Network diameter* is the shortest distance between the two most distant nodes in the network.
- *Average path length* is the average number of steps along the shortest paths for all possible pairs of network nodes.
- *Modularity* is a measure of the strength in division of a network into modules.

2.3.2 Models for Complex Networks

Models are used to mimic the connections within a network which assisted to understand a network's structure. We discuss a few general models which are used widely.

Random Graphs

Erdos and Renyi posit that given a set of all possible graphs that can be created using n vertices and N edges; a random graph is an element selected randomly from the set where each of the graphs has an equal probability of being selected [63]. It is represented as G

= *(n,N)*. On the other hand, Gilbert points out a deducing a random graph by randomly choosing the edges with an independent probability (p) for a given set of vertices [64]. In Gilbert's model, a random graph is represented as $G = (n,p)$.

The concept of random graphs is widely used in various fields to study complex networks such as brain structures [65], wireless networks [66], virus propagation and epidemic disease propagation.

However, as large size random graphs consist of Poisson degree distribution, random graphs are a poor depiction of the real-world networks as these networks consist of highly skewed degree distribution [67, 62]. Due to this complication, quite a number of research has been conducted to develop models to forego the shortcoming of the degree distribution. One such research [68] is where a realistic model is developed to include the properties; logarithmic average distance, high clustering and power-law degree distribution, to construct the graph based on a constructive model representing the real-world scenario and bring out a simple definition for the model such that intuitions and proofs of its properties can be produced easily.

Newman et al. developed a model such that the real-world properties of complex networks which are highly skewed degree distributions and difference from Poisson degree are co-existing while it depicts the real-world networks. This is enabled when the degree distribution in the model is specified, which in turn provides a solvable for the other properties [69].

Small-World Model

In the article written by Watts and Strogatz [70], they explained that there are real-world instances (biological, technological and social networks) where the graph had high clustering as regular lattices and small characteristic path lengths as random graphs which depicted the 'small-world' phenomenon. If a large network has got a small diameter, the network is considered to be a small-world network model.

Ever since, small-world model has been used in many studies related to social networks such as enhancing the cache replacement mechanism of Freenet [71], Jovanović's successful attempt to model peer-to-peer network topologies using small-world and scale-free model [72] and many more [73, 74]

Scale-free networks

In large networks which keep growing it was observed that [75],

- The expansion occurs due to the addition of vertices and
- The new vertices get connected preferentially with vertices that are well connected

These two properties in networks were described to be scale-free networks which forego the properties of individual systems and many. Furthermore, this depicts that the scaling of scale-free networks follows a power-law distribution. Barabási [76] further studied on its properties and introduced a model that depicts the World Wide Web as a self-organizing scale-free in spite of its characteristics of being a random graph. It has also been used to study computer virus propagation [77], generate models to further understand the growth of a scale-free network itself [78, 79] and various other research.

2.4 Evolving Social Networks

It is important to understand different means that researchers have followed to understand evolving social networks with time. Dynamics of such evolution can be studies considering the network at Micro-level and Macro-level. Micro-level studies involve looking at the network with a high granularity. Often these studies look at how individual nodes behave with respect to time. In contrast, in Macro level, we study how larger components of the network like communities within the network evolves over time. However, there are occasions where both levels are considered in a single research. There studies which look at both the levels which are discussed in the mixed evolution methods

2.4.1 Macro-level Evolution Methods

Kumar et al. studied metadata with a timestamp of every event from two real-world networks of Flicker and Yahoo 360 exceeding 5 million and 10 million friendship links respectively [49]. They separated the users as *singleton*, ones who are not participating in the network, *giant component*, the active users in the network participating in the activities and are connected to one another, or through someone else, *isolated communities*, these communities are in the middle region, they do communicate but as small groups and do not engage in participating the larger network. These user components are studied using migration patterns, structural evolution, measurements such as average degree, degree distribution and time graph properties. It was found out that the isolated communities consist of star formation and slowly grow and merge with the giant component. The middle region also consists of the star formation. This star formation was explained using the invitations within the network. Furthermore, the average distance between users within the giant component reduces over time. The social networks grow over time in different stages. A simple model was developed based on biased preferential attachment to understand the component growth accurately. We can use this is to understand different communities within a real-world network, it provides us the understanding of where we can approximately predict the evolution of a network by looking at the initial structure of it.

Berger-Wolf and Jared introduce a framework to investigate the dynamic interactions of social networks over time [80]. They introduce the concept of Meta-Groups (MG). Their process involves two main steps to turn a dataset with temporal data to a set of partitioned groups.

- Given a set of partitions of a dataset, calculate the similarity of the partitions to one another. They have used Jaccard Coefficient to find the similarity of partitions. Meta-Group is a subset of partitions which has similarity measure greater than a certain threshold value
- An individual node in the network is a member of a meta group, if the number of partitions the node belong to, is greater than a certain threshold.

This framework makes it possible in finding the most persistent and the largest social structures in a network, as well as social structures that encompass a set of specified groups of individuals.

2.4.2 Micro-level Evolution Methods

A similar network partitioning technique to Kumar et al. [49] has been followed by Vicario et al. on studying the structural patterns of the Occupy movement in Facebook [81]. They studied the Facebook user activities of 179 US Facebook public pages during the Occupy movement, i.e. an international protest movement against social and economic inequality organized online at a city level. Similar to Kumar et al. [49], they separated the users into 2 categories, habitual and occasional depending on the number of likes a user has made. However, these two pieces of research differ in the granularity of how they look at the network.

According to Vicario et al. [81], a user is considered habitual if 95% of his/her likes are made to a single particular page. Rest of the users are categorized as occasional. Then the structures of the networks formed by these two categories of users were analyzed and compared with respect to their geographical patterns in the information diffusion process. They discover that there is a correlation between the geographical location of the activities and the distributions of likes, comments and posts, which implies that the activities of Occupy movement are geographically coordinated.

Leskovec et al. have carried out a detailed study on the network evolution taking into consideration four large online social networks namely Flickr, Delicious, Yahoo and

LinkedIn [82]. Full temporal information about individual node arrival and edge creation processes have been analyzed which collectively lead to macroscopic properties of the network. A methodology based on the maximum-likelihood principle was followed to investigate several network formation strategies in order to support their conclusion that edge locality plays a critical role in the evolution of networks. With the received conclusions via the investigations. They have developed a complete model of network evolution, which nodes arrive at a pre-specified rate and given chance to select their own lifetimes. It is figured out that the combination of the gap distribution and node lifetime leads to a power law out-degree distribution for all the four taken networks. Although there are several edge creation processes carried out by several researchers, they all lead to heavy-tailed degree distributions and hardly can apply to real-world. Therefore, this research has tried to capture the reality of networks in the best way. The microscopic node behavior is studied in this research. They have used a series of snapshots to consider patterns.

In this research, the edge arrival sequence that is directly responsible for global network patterns. Twenty-five models were selected based upon the three core processes namely node arrival process, edge initiation process and edge destination selection process. With the dataset, they have shown that inherently non-local nature of preferential attachment is fundamentally unable to capture important characteristics in those networks. The selected network was evolved edge by edge and for every edge that arrives into the network, likelihood that particular edge endpoints were chosen under some model. The product of likelihoods over all edges gives the likelihood of the model. A higher likelihood was considered as a better model in the sense that it offers a more likely explanation of the observed data. Studying edge initiation is started by observing how long a node remains active in the social network and then studying the specific times at which node initiates new edges during the active life. The research concludes by proving that most new edges span very short distances, typically closing triangles. Node arrivals being more network-specific, the edge initiation process can be captured by exponential node lifetimes and a "gap" model based on a power law with exponential cutoff.

The model which has been developed produce true evolution of network properties such as clustering coefficient and diameter giving rise to accurate global properties. The model can be used in future to generate arbitrary-sized synthetic networks which closely describes the macroscopic characteristics of real social networks. It can be understood that user interaction generation in a social network can be identified by the lifetime of users in a specific network and macroscopic characteristics of these users in the social networks can be described with the use of the model that developed by the researchers.

Ahn et al. took an attempt at comparing the structures of the three online social networking sites; Cyworld, MySpace and Orkut each containing more than 10 million users [83]. Researchers try to show that they deviate from close-knit online social networks which show a similar degree correlation pattern to real-life social networks. The researchers try to answer the questions “What are the main characteristics of online social networks?”, “How representative is a sample network?” and “How does a social network evolve?” through putting interest into huge and those of magnitude have not yet been analyzed.

Four sets of online social network topology data from the above mentioned three popular SNS has been collected and the analysis was started by looking at the metrics; degree distributions and clustering coefficient. From a snowball sample network, researchers have measured the number of nodes at each round of breadth-first search. The average path length between nodes is estimated from sample networks. To confirm that the distribution of average path length eventually converges to complete network, researchers have looked at the incremental change that they can see when the number of sample networks increased.

Sample ration is considered as 0.33 percentage which resulted in 40,000 nodes in each sample network. It is proven that sample networks have varying smaller degree exponents than the complete network, but the multi-region scaling behavior is same. This has brought the knowledge on network evolution with the careful investigation on social network measurements like average degree distribution, clustering coefficient. In fact, there is a gap exist between this knowledge and the reasoning for these node properties from the perspective of user behavior in online social networks.

2.4.3 Mixed evolution methods

In a longitudinal study conducted for over 32 years to identify the spread of obesity over social networks, it was found out that obesity was more of a public health problem [53]. Christakis and Fowler used the data of a densely interconnected social network of 12,067 people collected at 3-year intervals from 1971 to 2003 which included three different generations. They identified *ego*, people whose behavior was analyzed and *alter*, individuals who influenced an ego’s behavior within the network. Data such as weight gain, gender influence, friends, family, environment changes and geographic distance between the domiciles of persons in the social network were collected which is available from the Framing-ham Heart Study. Since the network was interconnected by friendships, three types were identified; *ego-perceived friendship* where ego identifies and alters as a friend, *alter-perceived friendship* where alter identifies ego as a friend and *mutual friendship* where the friendship is identified by both the ego and the alter. Ego-perceived friendship was considered the most socially influencing in the network. The network data was structured using the Kamada-Kawai algorithm in Pajek software and videos of the network were generated using Social Network Image Animator (SoNIA) at time intervals. The derived models were then compared with theoretical models such as small-world, scale-free, and hierarchical types. The clustering of the network was explained using homophily, confounding properties and induction in the environment. The longitudinal logistic-regression model was used to statistically analyze the data of the network. With this quantitative analysis of the social network, the study suggests that obesity spreads with the network with the nature of the social ties. This finding can be applied to online networks as well. The proposed categorization of social ties would be helpful in studying the influence users can pose within a network, thus used in areas such as influence maximization or identification of influential nodes.

There are various models available explaining the influence and structural changes within social networks separately. However, actor-oriented model [84] looks at the longevity of the network and model it using both the co-evolution of social networks and individual behaviors. Burk et al. use Simulation Investigation for Empirical Network Analyses (SIENA) software to implement the model and test it with the data from a longitudinal sample of Swedish adolescents whose co-evolution of friendship networks and delinquent behaviors was collected. Selection of ties and influence process within the network was also assessed simultaneously. Process of modeling, the analysts used continuous-time Markov chains to ensure that changes of ties at a given instance occur due to the current network configuration and has nothing to do with previous configurations. The actor-oriented model looks at the network and individual behavior evolution separately in different transitions, but however models all possible variants that can occur simultaneously in a probable space. Three assumptions guide this model; continuous-time Markov process is used to depict changes, actors may change only one network tie or one level of behavior at any moment in time and actors do not react based on the prior agreement but on the behavior of each other. Thereby the model looks at every property and frequency of the network and behavioral micro steps of the individuals to obtain the model. Irrespective of the complexity of the resulting model, it assists in understanding both individual's decision regarding changes in social ties and the behavior of networks since these two processes relate with each other. Furthermore, internal changes within a network can also be simulated using the derived model from SIENA. This analysis provides a path for future researchers to evaluate the influence posed by individuals in the network in relation to their decisions on social ties. It would be important in identifying manipulative influencing within a network.

Users interact with the content in online social networks in various ways(wall posting, messaging, applications, photo uploads, and chat), thus an activity is performed in the network. Vishwanath et al. [85] studied the evolution of the activity network at both the

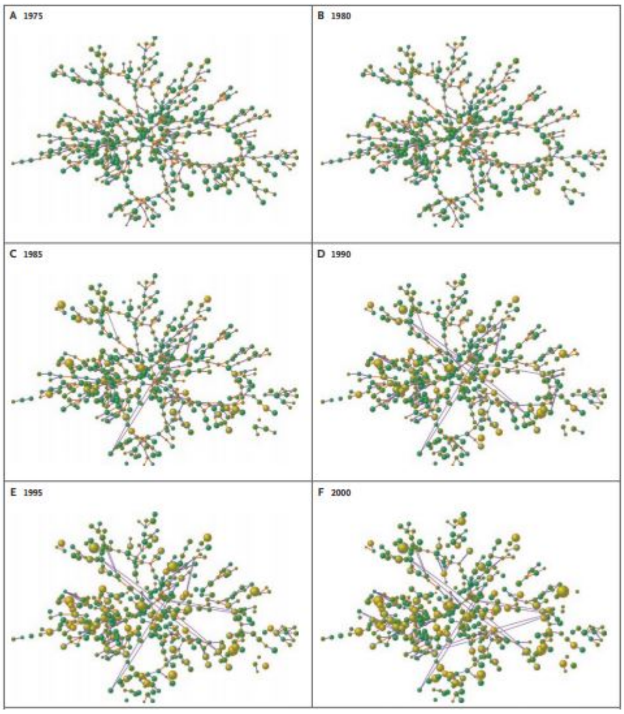


Fig. 2.1 Social network evolution of individuals from Framingham Heart Study with Information about Body-Mass Index According to Year as illustrated in [53]

micro and macro-levels. They used data from a subset of 60,000 Facebook users in New Orleans over 2 years. There were 800,000 wall postings which are considered as interactions in the study. Pairwise user interaction was analyzed at micro-level where the distribution of wall posting per social ties was examined and active pairs were identified using cumulative distribution function and pattern recognition. It shows that only a few user pairs interact highly while a majority of users interact rarely for occasions such as birthdays which accounts for a skewed distribution. At macro-level, the evolution of the whole network was observed by taking snapshots at 9 intervals during the period of 2 years and the quantitative overlap of consecutive snapshots was analyzed. Furthermore, metrics such as average node degree, clustering coefficient and average path length of the snapshots were calculated. This part of the research explains that irrespective of high activity levels between user pairs, the properties of the whole network remains stable over time.

2.5 Layout algorithms

Several force-directed layout algorithms for network visualization in popular use include Fruchterman-Reingold, ForceAtlas2, LGL and SFDP. In their paper, Jacomy et al. [86]

present the functioning and settings of ForceAtlas2, continuous graph layout algorithm which is designed within Gephi, social network analysis tool for the precise understanding of its' behavior to the users.

Most visualization techniques used with Online Social Networks employ a variant of force-directed graph layout according to Linton [87]. The positions of nodes in the drawing are determined by applying attracting forces between edges and repelling forces between nodes. Social networks graphs are commonly drawn using a class of algorithms called force-directed layout (also known as spring embedders) according to Linton.

The Fruchterman-Reingold Algorithm [88] is one of the best known force-based algorithms to represent complex networks. In this algorithm, they have provided a modification to the spring-embedder model of Eades for drawing undirected graphs with straight edges. The force exerted on a node is the vector sum of repulsion forces due to all other nodes and the sum of attractive due to the vertices connected to the vertex itself.

Though there are different layout algorithms, most of these are unable to use in large networks as it requires large memory. As a result, there is a need for powerful layout algorithms that can be used in any size network. The pursuit of less cluttered and more revealing visualizations has prompted further research on measures of graph layout quality according to Dunne et al. [89] There attempt was to create awareness about the effective graph drawing strategies. They have defined a new node and edge readability metrics to provide more localized identification.

2.6 Social Network Analysis Tools

There have been different discussions and viewpoints regarding the social network analysis tools that have been used to analyze the social networks answering different kinds of research questions. Martinez-Lopez in his research paper on Social Network Analysis tries to review the concepts and theoretical aspects related to the social network analysis and graph theory. The paper [90] discusses about open source SNA tools and SNA tools that can be used for academic purposes like Agna, Bianche, Cytoscape, FATCAT, Igraph, Iknow, KliqFinder, JUNG, Multinet, NetDraw, NEGOPY, Netvis, ORA, Pajek, PermNet, PGRAPH, Network Insight, StoCNET, STRUCTURE, VISIONE.

International Network for Social Network Analysis [91], the professional association for SNA researchers provides a platform to get access to different kinds of SNA tools and methods that are used in data collection. Also, it acts as a portal which provides publicly available datasets to the researchers interested in Social Network Analysis. Different research

Software	Main Functionality
Agna	Platform-independent application designed for SNA, sociometry and sequential analysis
Cytoscape	Network data integration, analysis and visualization
igraph	Network analysis package with emphasis on efficiency and portability
NetDraw	A free windows program for visualizing social network data
Negopy	Finds cliques, liaisons and isolates in networks
Netviz	A tool for visualizing real-time network or graph data
Ora	Dynamic network analysis tool which supports visualization
Pajek	Large network analysis tool for Windows platform
pgraph	A Python GUI for display of network topologies

Table 2.1 List of SNA tools

papers on SNA in several contexts have been added to the portal in order to support the researchers.

The Netvizz application [92] was initially developed by the author in 2009 as a practical attempt to study Facebook's API as a new media object in its own right and to gauge the potential of using natively digital methods to study Social Network Sites. The Netvizz application works as a tool that generates dumps of data from social media and these data works as an input to the social network visualization.

Graph visualization tools play a vital role in analyzing social networks. However, there are key issues that need to be considered when selecting visualization tools; the size of the network and usability [93]. When considering Gephi as a visualization tool for social network analysis, it has become a very time consuming and costly tool as the large networks take a lot of memory and space to visualize themselves via the tool.

Table 2.3 presents a list of network visualization tools in the arena of social network analysis.

2.7 Behavioral Models

Wellman [94] emphasizes the paradigm that "behavior is interpreted in terms of structural constraints on activity rather than in terms of inner forces within actors". In the book, "Six Degrees: The Science of The Connected Age" [95], Watts emphasizes that networks are the signature of social identity and the pattern of relations between individuals reflects the preferences and characteristics of individuals.

A mechanism is a theory or an explanation and what it explains is how an event causes another according to Kosowski. Strong tie hypothesis implies that one's close friends tend to move in the same circles that she/he does, while Weak tie hypothesis argues that weak ties are responsible for the majority of the embeddedness and structure of social networks in society as well as the transmission of information through these networks [96].

Tan et al [97] study how users' behaviors (actions) in a social network are influenced by various factors such as personal interests, social influence, and global trends. Here, they have proposed a Noise-Tolerant Time-varying Factor Graph Model (NTT-FGM) for modeling and predicting social actions. With the use of this model, researchers have tried to formalize the social action tracking problem exist in social networks.

Kossinets et al. [98] have proposed a temporal notion of distance in social graphs, by quantifying how long it takes for information to propagate along a given edge. They have taken an email communication system in a university as the case study and have observed the temporal dynamics of the communication. Their findings reveal that temporal measure provides structural insights that are not apparent from analyses of the pure network.

Cosley et al. [99] subsequently proposed a model of how influence propagates through a social network. In their research, they have considered two definitions for the influence based on a small set of "snapshot" observations of a social network and detailed temporal analysis. How influence spread is described using the above two methods. In our research, we followed the two definitions mentioned above but then tried to fit the results to an existing behavioral model.

Freeman [100], in his research paper, defines a person having high betweenness centrality is considered to be able to "influence the group by withholding or distorting information in transmission" because he or she is located as a passage point between different sections of a network. These users work as hubs of the network and further research should be carried out to quantify the actual influence on the other nodes from these hub nodes.

There has been a long-standing interest in identifying central individuals in a social network according to Bavelas [101]who make an influence on the others in a social network. Watts et al. say that it is important to underscore that highly central individuals are not by definition influential There are still questions left unanswered for users' behavior in social networks; whether they are influenced or not and the influential factors that affect them if there exists an influence for a user [95].

Chapter 3

Methodology and Design

3.1 Introduction

This chapter elaborates on the nonexperimental descriptive research design approach using quantitative methods employed in understanding the impact of social network interactions on user affinity relating to key events such as an election which happens outside the social media platforms. The study uses the Sri Lankan Local Government Election 2018 [26] as a case study to draw conclusions and answer the research questions driving the entire research. This approach will help to understand and describe the changes in user affinity relating to the interactions and gain a deeper understanding to the underlying mechanics to this phenomenon. This chapter breaks down the applicability of nonexperimental research design, data collection and analysis methods, evaluation of the results and the validity and reliability of the research in a comprehensive manner.

3.2 Research Questions

1. How does the community interactions on social media influence the user affinity?
2. How to measure the user affinity changes with relation to community interactions on Facebook?
3. How does the online user affinity evolve corresponding to social phenomena depicted in Facebook in association to community interactions?

3.3 Research Design

This research examines the interactions of users in social networks and the user affinity as the main variables. We adopted a nonexperimental research utilizing quantitative approach with an explanatory design.

The data must be able to reflect the natural and random behavior of Facebook users. The research questions look into analyzing the causal relationship between user affinity and the social interactions of the users. We are not able to manipulate the variables for the study, thus a nonexperimental research approach is followed. In nonexperimental research, the variables cannot be manipulated but studied as it stands [21]. The main considerations in applying a nonexperimental approach are the importance to capture the natural and random behavior of the users with external interference. Adding to it, it is unethical to manipulate the interactions in public discussions in a social network for the purpose of the as it will cascade incorrect information to the society [21, 22].

3.3.1 Research Purpose

The purpose of the research is an important dimension that must be considered in the research design. Fundamentally there are three types of purposes of carrying out a research; exploratory, descriptive and explanatory [23, 24]. In this research, we examine the relationship between social interactions and user affinity. The entire processes embody with describing and understanding the properties within the network and the relationships with various structural properties. It will, in turn, assist to understand the relation between the social network interactions and user affinity. Furthermore, we look into understanding the changes in user affinity with time. Thereby, it is evident that the purpose of the research is descriptive.

3.3.2 Research Approach

'Community interactions in social networks influence the user affinity' has become one of the growing ideologies in the society. In the process of understanding the above phenomena, it is important to review the literature to identify the existing knowledge in the domain. It was clear that the literature was not abundantly available for this domain of knowledge. A clear knowledge gap was identified. Therefore, we planned to analyze comprehensively collected large amount of data and get a generalized point of view of the phenomena.

David Thomas explained in his study the motivations for an inductive approach [102];

- to get a summarized format of the available data

- to get an unambiguous relationship and transparency between the findings and objectives of the research
- to model or theorize the processes or experiences that is evident in the collected data

In our research, as we look into a large collection of data and attempt in finding it relationships between its properties and produce a generalized view of the data through analysis. We adapted the inductive approach in this study. Furthermore, the data is analyzed using social network analysis using the graph theory representations [45]. The graph theory uses various mathematical concepts in the process of analysis. The study follows a quantitative analysis path with both the use of graph theory and statistical analysis techniques.

3.4 Data Collection

In this study, we look in the social interactions and the user affinity. The social interactions refer to the reactions, comments, and shares of posts made on Facebook groups or pages. We considered Facebook groups and pages which are public to Internet users. As we use, the Local Government Election 2018 of Sri Lanka to answer the research questions, initially we had to find Facebook groups and pages which were used as a platform to communicate about the aforementioned election. Snowball sampling was used to identify groups and pages where political discussions took place. Snowball sampling is also referred to as chain referral sampling where a sample is created using a series of referrals using entities or individuals who know one another [103]. An initial set of Facebook groups and pages relevant to the political discussions regarding the Local Government Election 2018 of Sri Lanka were identified using keywords. The dataset was expanded using snowball sampling to further identify other Facebook groups and pages. We excluded the Facebook groups and pages where there was no activity in the past 4 weeks. Despite this sampling method being non-probability and creating a biased sample [103], we followed this method to reach groups and pages that are actually in touch with the users and most likely to hold political discussions. Furthermore, it assisted to identify sample units which would remain unidentifiable if not for snowballing sampling. In order to reduce the bias, we included Facebook groups and pages which are politically independent and politically dependent in the initial sample.

We extracted the user interactions from the identified sample of Facebook groups and pages using Netvizz. Netvizz [104], a Facebook application which is built on top of Facebook graph API was used to extract anonymized author ids, post ids, comment ids, the content of posts and comments, timestamps, reaction counts, comment counts, group/ page id, number of shares of a particular post data in Facebook groups. Netvizz provides data as a set of tab-delimited and comma-separated files.

The final dataset contains pre-election and post-election data of user interactions from 46 different Facebook user groups and 76 public pages of Facebook collected during the period from 1st of December 2017 to 30th February 2018. We started the data collection few weeks prior to the election and continued the collection few weeks after the election was over with the intention of getting a complete dataset with pre-election and post-election activities. It was important to collect the data continuously during the period as certain users/admins tend to delete the content periodically or once the election results are announced. In the data extracted from groups alone, there are 44130 unique users with 192,958 interactions, out of which 4297 have interacted with more than one group.

3.5 Data Analysis

Analysis refers to identifying patterns and the reason for its existence within the data and quantitative analysis refers to use statistical methods to identify patterns in data [105]. The dataset collected is qualitative. We conducted quantitative analysis on this dataset using graph theory and statistical analysis techniques.

Graph Theory Techniques

The collected data is preprocessed to obtain a dataset without duplication, inconsistencies and missing data. Then dataset is fed into the Gephi [106], a network visualization and exploration tool which enables grasp unseen patterns and properties of a network.

We applied graph techniques considering two levels of granularity;

- Nodes of the graphs represent individual users and the edges represent the interactions between users which sum up the micro-level analysis. We modeled the graph using the force-directed layout in Gephi. This analysis looks into the temporal changes of the graph structure.

- Nodes are Facebook groups and the edges are aggregated number of migrated of users between two groups, which is the macro-level analysis. We use SNAP (Stanford Network Analysis Platform) to create the structure of the graph. SNAP is a general purpose network analysis and graph mining library used to analyze a large dataset [107]. Thereafter, the graph structure is visualized using Gephi. Furthermore, we compare the graphs of the user migrations with random graphs of similar properties to identify if the dataset depicts either random or non-random properties.

Statistical Analysis

We analyzed the reactions made by the Facebook users on the posts on the groups. It is a simple calculation of the aggregation of the types of reactions on the posts. Here the reaction types referred to are Haha, Wow, Angry, Like, Sad and Love. This aggregation is made for 02 user categories; users who interact in only one group, users who interact with multiple groups.

We used correlation analysis techniques to identify the relationship between the actual event which is Sri Lanka Local Government Election 2018 and analysis results derived from the graph techniques on the dataset collected from Facebook. However, correlation analysis doesn't determine the cause and effect relationships with the two entities.

3.6 Evaluation

It is important to ensure the correctness of the results by evaluating it. A comprehensive evaluation process needs to be defined in the early stages with the aim of exhaustively testing the results. We adapted manual content-coding for evaluation. We introduced a model to identify the changes in user affinity through the use of migrating users within Facebook groups. The evaluation looked into the effectiveness of this model by coding the comments of a sample of migrating users who have interacted with the Facebook groups within the timespan of the dataset. This sample of users is identified through the model. Manual content analysis, as opposed to automated text analysis was used due to the nature of the content. Post and comment texts largely consist of content written in Sinhala and sometimes as a mixture of Sinhala and English. Sinhala itself being a morphologically rich language, Sinhala and English together complicate the automated process of content coding. This makes it a difficult task to use Natural Language Programming techniques for content coding.

Content coding was conducted in order to show that,

- Users we considered as migrated users from one Facebook group to another Facebook group show actual changes in their political affinity
- Users we considered as non-migrated users who were supporting the same Facebook group throughout the data collection period does not show changes in affinity

Manual content coding was done with the help of 45 participants. We tried our level best to involve the participants from various backgrounds so that we can minimize the biases of opinions. 36 out of 45 content coders were male. They were tasked with labeling the interactions of the identified migrated users using a predefined criteria.

A random sample of migrating users and non-migrating users as predicted from our model was selected. This sample consisted of 21 migrating users and 24 non-migrating users. Each content coder was assigned with 3 users. And each Facebook user in the dataset was assigned to 3 different content coders. Content coders were given an anonymized dataset of interactions of the users assigned to him/her (See Appendix B).

We asked each participant to evaluate each interaction with two specific criteria as follows,

- If the interaction of the user is supportive to the general idea of the post, the participant has to put 1 and if not 0.
- If the comment the user for a particular post contains hate speech, the participant has to put 1 and if not 0.

Since 3 different content coders were involved to code the content of a single user, it was possible to choose the label the most content coders agreed upon. Indirectly, this helped to reduce the individual bias that could have been introduced by a single content coder. Table 3.1 contains a sample of aggregated labels of one such user.

With the results from content coding, it was possible to see if there is an affinity migration reflected in the interactions of each user. Then we matched the affinity migration status, as well as the migration directions of the users with the predictions from our model to evaluate the performance.

				Content coder 1		Content coder 2		Content coder 3		Final	
userid	commentText	postText	link	Suppor -tive?	hate Speech?	Suppor -tive?	hate Speech?	Suppor -tive?	hate Speech?	Suppor -tive?	hate Speech?
123	Athakota hamuda mulasthanya vikunuwea pot city ya liyala dunnea kauda bag	vikalpa netha pohottuwa pamanai	REDACTED	0	0	0	0	1	0	0	0

Table 3.1 Content coded interaction of a single user

3.7 Summary

The table 3.2 summarizes the entire research design used in our study.

Research Design Aspect	**Type**
Type	Nonexperimental
Purpose	Descriptive
Approach	Inductive
Data Collection tools	Netvizz, Python scripts
Data Analysis	Social network analysis using graph theory and statistical techniques, OpenSNA
Data Visualization Software	Gephi, NetworkX, SNAP,Matplotlib
Evaluation Technique	Manual Content Coding

Table 3.2 Summarized Research Design

Chapter 4

Results

We aimed at identifying the impact on user affinity through online user interactions. In order to discover the actual impact, we studied the changes in user affinity around a social event where we believe most of the social network users are active. We collected data from Facebook groups around the Local Government Election 2018 held in Sri Lanka.

We broke down the analysis into micro-analysis and macro-analysis where both the methods show that there are changes in user affinity with time through user migrations. The analysis methods consist of various graph techniques and statistics. We present a novel approach to identify user migrations within Facebook groups by bringing together social network analysis and sociology which is a significant milestone in this study. The model was evaluated using manual content coding which resulted in an accuracy of 82%.

This chapter unravels the results of the entire work process of the study. It provides a very detailed description of various analysis methods and results that are both reasonable and unreasonable for this study. We order the content as micro-analysis, macro-analysis, reaction analysis and user content coding.

4.1 Micro-level Analysis

This is the initial level of analysis where we looked into the lowest level of details in the dataset. We collected all the data using Netvizz and use graph modeling techniques in this stage and got the graph model using Gephi software. The nodes of the graph are the users and the edges between the nodes are the user interactions within themselves. An edge between two users is made when one user comments or reacts to a post of another user. Even though it is one of the simplest representations of the dataset, we got to study the most intricate details of the structure of the entire network we consider for the study. The addition of the temporal component for the graph modeling makes the micro-level analysis much more interesting and informative. It showcases the growth of a user together with the whole network around them throughout the entire time period. Furthermore, we drilled down into the dataset to understand the behavior of each user separately.

4.1.1 Data Preprocessing and Labelling

We combined the collected comment and reaction data of each group. Then the user hashes (Netvizz uses user hashes to represent a unique user) are replaced with incremental user ids for the ease of use. Multiple files are mapped with each other to create a dataset of dynamic user interactions in the following format: `[user1, user2, timestamp]`. The `timestamp` represents the time an interaction between the `user1` and `user2` takes place. Interactions could be either comments or reactions of user 1 to a post of user 2 or the other way around. However, due to the limitations of Facebook's graph API which does not provide the time a reaction was made, we chose to use the time the post was created for reactions. Duplicated records are identified and filtered out. The few incomplete records are removed.

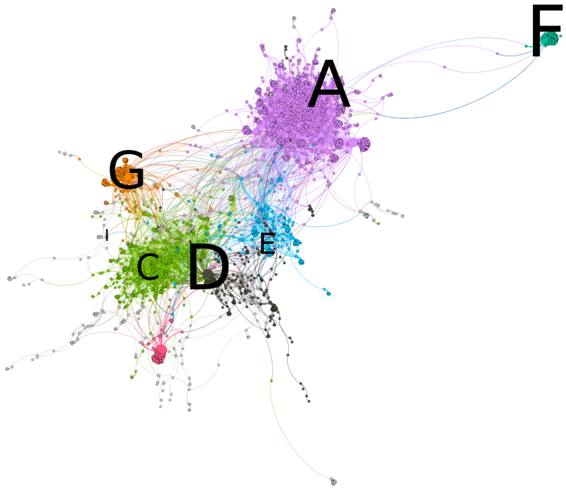
In addition to the above-mentioned network, an aggregated inter-group interaction graph is used to represent the user movements within groups.

4.1.2 Graph of Individual User Interactions

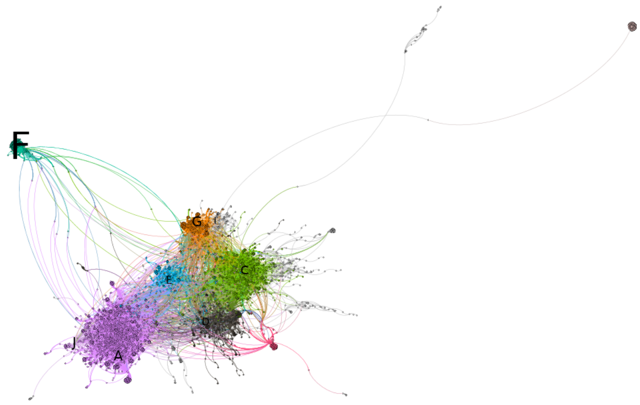
The first representation can be found in Fig. 4.1. It is possible to study the temporal evolution of interactions. We observed that the clusters of nodes are consistent with real-world Facebook groups. With time, new clusters enter into the network. Existing clusters grow in size. Furthermore, distances between certain clusters either reduce or increase. This is a visible indication of changing modularities of the network. Since modularity of the clusters is a representation of affinity, this is a visual indication that affinities do change.

Interestingly, during the course of the election period, we observed that the structure of the graph is vastly changing. Facebook communities with opposing political views sometimes came closer in the graph. Certain self-identified independent groups coalesced with political groups. While this could be an indication of a dramatic shift of the user's political opinions, it is not possible to establish it through this analysis alone.

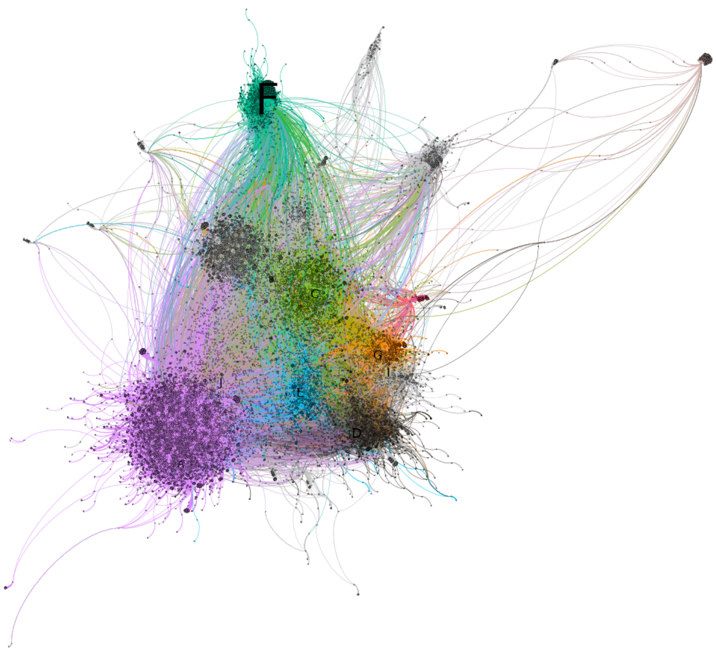
Fig. 4.2 shows the frequency of interactions of a single user across the groups he has interacted. The graph summarizes a user's activities in all the groups the user has been interacting during the time period. Each colored line represents a unique group. We saw that there are spikes of interactions in groups the user has not previously interacted substantially. This kind of behavior was observed among multiple users. These spikes in activities could have been a result of a real-world event. This is another interesting observation given that this



(a) 10-12-2017



(b) 21-12-2017



(c) 20-02-2018

Fig. 4.1 Micro-level interactions of users during the pre-election and post-election period

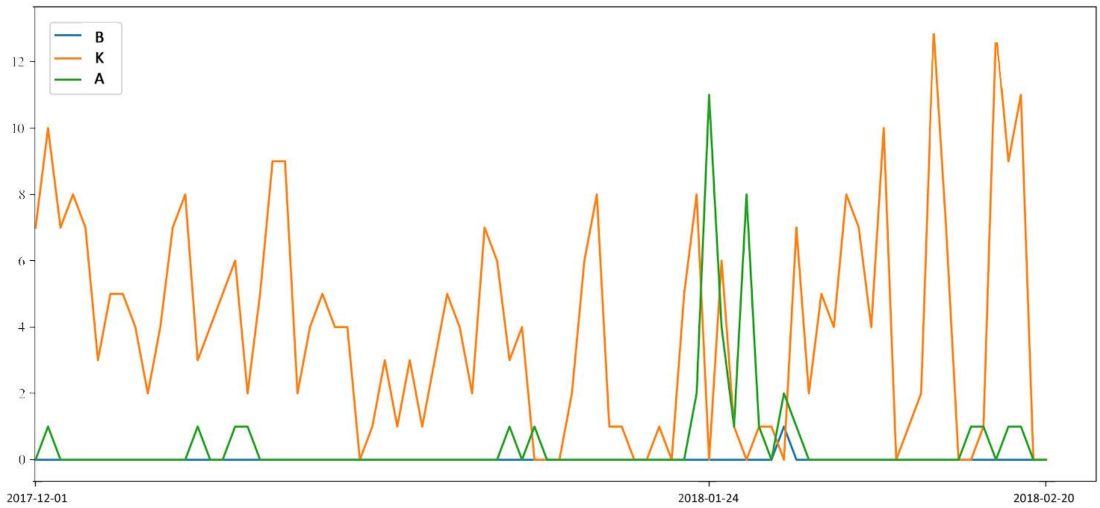


Fig. 4.2 User activity frequency of a single user across groups

re-confirms that the connection between the real-world event and social media interactions are a two-way street meaning that both have significant impacts towards each other. And the main spike is around the actual election day which is not surprising.

4.1.3 Results of Individual User Interactions

The 44 collected Facebook groups contain 44130 unique users who have interacted with them within the collection time period. Different groups have a varying degree of interactions with users. The most interactive group accounts for 37% of interactions and least interactive group accounts for 0.01% interactions. The likely scenario is that the majority of users are concentrated on very highly popular groups rather than equally spreading their attention across many groups. Thus, we envision further research is required to understand why a certain set of groups are popular while the rest getting substantially less attention. We found that 4297 users have interacted with more than one group. This accounts for 9.7% of users in the entire dataset. This re-affirms our previous observation that a significant majority of users interact with only a chosen single group.

Analyzing the 1st representation(Fig. 4.1), we report that the average degree of a node is ten. And the network diameter, the longest shortest distance between any two nodes of the graph is thirteen. We further report that the degree distribution follows a power law (Fig. 4.3), implying that the network is scale-free. This means the number of nodes with very low interactions is higher than the number of nodes with very high interactions. This observation could be due to the fact that most discussions are driven by the handful number of highly

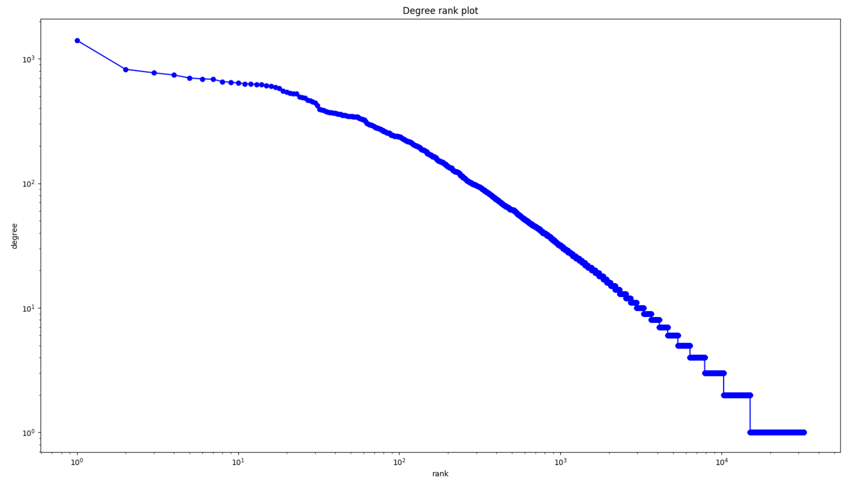


Fig. 4.3 Degree-Rank in log-log scale

motivated and/or outspoken individuals. In the context of the big picture of analyzing the impact of the social media on events like elections, this is a significant observation. If those highly outspoken individuals are malicious then few people can negatively influence the opinion of the many connections to relevant groups. However, if they have good intention, this might well be a positive push towards being transparent and getting people more informed. However, as it was with the previous observation, further research is required to verify our hypotheses.

The visualization in Fig 4.1 represents one of the most significant observations of our analysis. These snapshots show that the user interaction gets intense and the number of people communicating with each other significantly increased towards the election day and remain like that after the election. This observation does not explain the increased interactions, it confirms the long-standing belief that people are using social media activity during events like an election. Given the observed groups are politically motivated groups, the only plausible explanation for the increased interaction is the election.

4.2 Macro-level Analysis

This analysis acts as the next step in understanding the influence of social media on the users. The goal here is to identify any anomalies in the migration patterns which infers that several users or groups of users intentionally or naturally migrate within groups which hamper the expected use of Facebook. We analyzed the user migrations patterns within Facebook groups by representing the dataset using directed weighted(DW) graphs. The weighted edges (migration of a user) between the nodes (Facebook groups) of the graph is useful to understand the intensity and the strength of the migration. Furthermore, it assists in understanding the topological architecture of the network. We compared the derived DW graph with random graphs generated using weighted Erdos-Renyi graph model [108] with scale-free behavior and power-law distribution.

4.2.1 Data Preprocessing and Labelling

The public Facebook groups consisting of user interactions in our dataset can be classified as belonging to either a political party (promoting its propaganda) or as an independent community. The 11 Facebook groups with the highest user interactions within the period of data collection are anonymized with labels. Among those groups, the groups B, D, H, I, J can be identified as belonging to various political parties while groups A, C, E, F, G can be identified as independent communities.

Initially, we combined the comment and reaction data of each selected Facebook group obtained via Netvizz. Next, the user hash (Netvizz uses user hashes to represent a unique Facebook user) of every user is replaced with incremental user ids for the ease of use.

We removed duplicate records as well as few incomplete records in the collected dataset. We mapped multiple files with each other to create a data set of dynamic user interactions across multiple groups.

Then we partitioned the Facebook users into two groups.

- Active users - Users who have interacted in more than one group
- Non-active users - Users who have interacted only in one group

4.2.2 Inter-Group Interaction Model Generation

In order to study how Facebook user interactions between multiple groups evolve over time, we represented the inter-group interactions in a directed graph. Such a graph representation permits one to observe and measure user migrations at a macro-level between different Facebook groups.

We used Gephi as a visualization tool to represent the interaction. A directed edge between two nodes implies a migration of a user from the source group to the target group. The force-directed layout is used to represent these interactions in Gephi so that nodes with many interactions are much closer than the others.

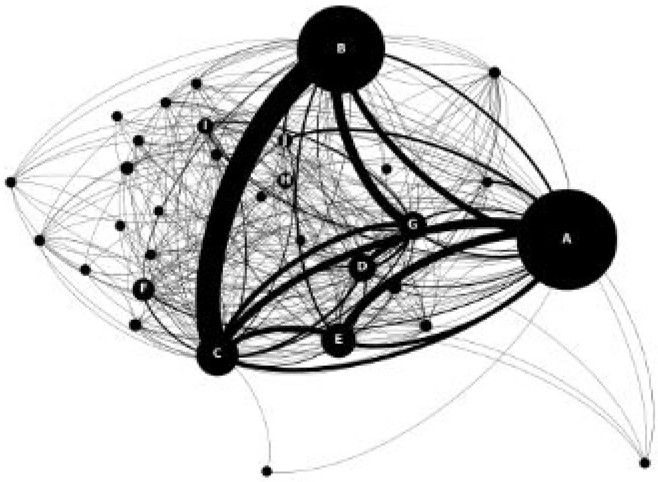


Fig. 4.4 User migrations among 44 Facebook groups

Size of the node is proportional to the number of interactions happened in the group. Self-edges are not represented on the graph, however, they still affect the size of the node. To understand how we represent a migration in the graph, consider the following inter-group interactions of a single user. A as the source group and B as the target group which the user migrated at the end. After looking at the interactions, if the following conditions are satisfied, then we identify that a user has migrated.

- Interactions of the user in group B has a later timestamp than that of group A.
- At least 20 percent of the user's interactions are in group B.
- Users total number of interactions in group A and B are greater than mean interactions of active users.

With the above-outlined representation, edge weight represents the number of migrated users from one group to another. Fig 4.4 depicts the above-explained graph with anonymized edge labels. Edge direction in the figure should be read in the clockwise direction. It can be observed from the graph that the edge weights are non-uniform.

4.2.3 Results of Inter-Group Interaction Model

We identify that there are 1051 users who satisfy the previously stated migration criteria out of 44098 total users. Fig 4.4 is a visualization of the migrations of 1051 users across the 44 groups. The thickness of an edge corresponds to the number of users migrated while the size of the node represents the number of interactions within the group. From the visualization

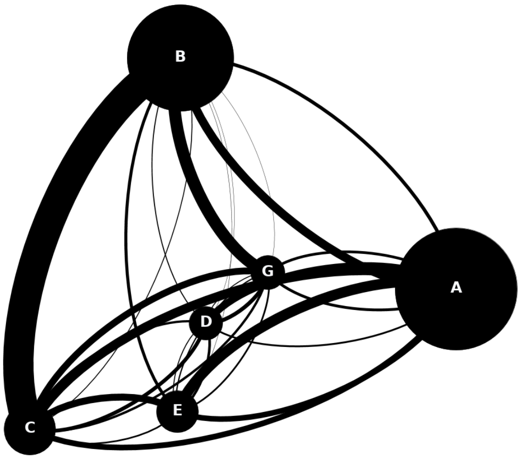


Fig. 4.5 User migrations among the top six most active groups

alone it is apparent that migrations are prominent within a few groups than the rest. Fig 4.5 shows the migrations among the top six groups where most interactions have happened.

4.2.4 OpenSNA

Due to the lack of existing tools to analyze social network data in the way we have in the above steps, we developed a toolkit, OpenSNA which is capable of doing higher-order operations on social media datasets. It is built upon networkX, a low-level library for graphs. OpenSNA supports various higher order operations like link prediction, network diameter calculations, modularity calculations and functionality can be extended via extensions. It is written in Python and available as an open-source software at *https://github.com/scorelab/OpenSNA*

4.2.5 Random Graph Modelling

We modeled a Random Graph keeping the number of edges and nodes constant, in order to compare the randomness of the interactions in the network that we were analyzing. We initiated the model with an Erdos-Renyi graph, which is a widely used model for generating random graphs. Then the edge weights were updated drawing out random samples from a power-law distribution where the exponent is 2.16. This exponent corresponds to the Pareto principle as derived from Staša Milojević [109] from Newmann's analysis of power laws [110]. Random scale-free networks generated in this manner have been previously used by Riccaboni and Schiavo in their studies. They show that this way of modeling successfully captures the properties of real-world networks [111].

This method has since been used in simulating the weighted scale-free networks to study evolution and properties of world trade webs [112] [113].

To study the interactions with a higher granularity, we partitioned the data set into ten-day intervals and investigated their properties separately.

We partition the network to ten-day partitions. For each partition of the network, we generate a random graph model. Properties of these two graphs are then compared. Fig. 4.6 shows edge weight distribution comparisons taken at different time intervals. Even though the edge weight distributions of the real dataset and modeled graph follow the same general power-law curve, it is clear that there is a significant variation among them. In all of the partitions of the real dataset and their modeled graph, it is observed that modeled random weights are skewed towards smaller numbers compared to the random model. Interestingly, this means that there is an abnormally high number of migrations in a few groups. While it is not clear why these migrations happened, they show a deviation from random behavior.

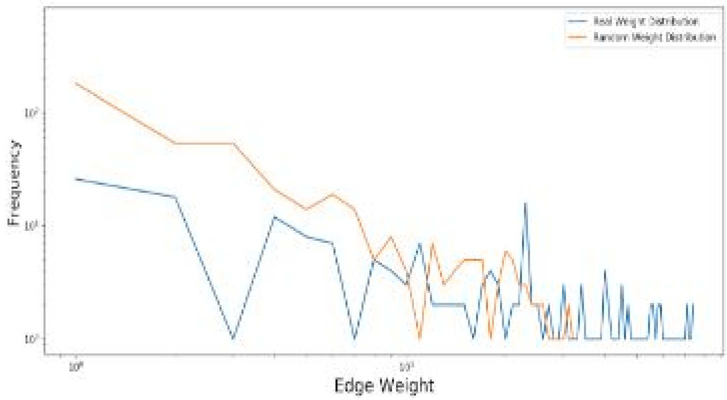
Fig. 4.7 shows a comparison of diameter, average path length and clustering coefficient of the real graph and the random graph. The diameter and average path length of the real graph are consistently higher across all partitions. This implies that the real graph is more spread out than the random graph model. However, the clustering coefficient is consistently higher than that of the random model. This is due to the fact that the real groups have more closed loops of edges between nodes than that of random. This is again evidence that the real network has behaved significantly different from random behavior in the period the dataset was collected.

4.2.6 Analysis of User Activity Intensity

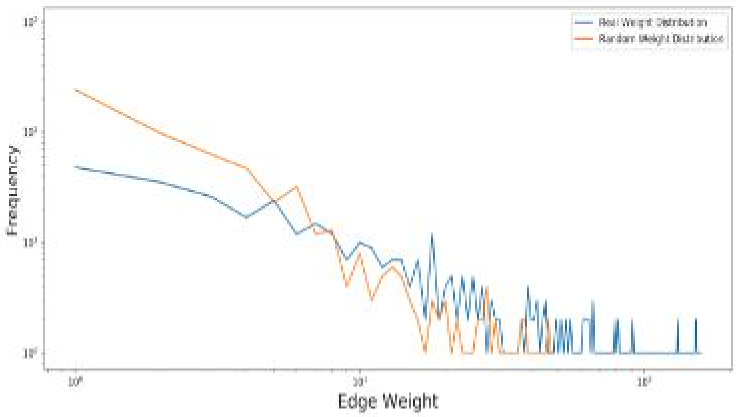
In order to see whether the non-randomness of user behavior is visible in another perspective, we looked at the cumulative user frequency graph and cumulative comment frequency graph of the dataset. Separate Python scripts were written to extract and transform the data into suitable formats. Matplotlib, a popular Python library was used for the visualizations.

Cumulative user frequency plot (Fig 4.8) represents the number of new users that have started interacting with for each day in the dataset. The frequency of new users interacting in the observed groups tends to grow linearly with time, however, a slight increment in the rate of increase is observed just before the election, which is expected.

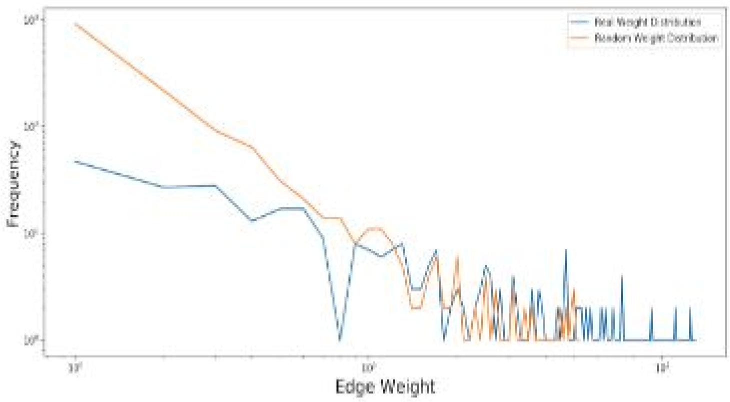
Cumulative comment frequency (Fig. 4.9) represents the growth of comments across time. Growth appears to be similar to what is seen in Fig. 4.8. Since not much new



(a) 09/01/2018



(b) 08/02/2018



(c) 28/02/2018

Fig. 4.6 Edge weight distribution comparison in log-log scale

A COMPARISON OF REAL AND RANDOM GRAPHS

	Diameter		Average Path Length		Clustering Coefficient	
	Real	Random	Real	Random	Real	Random
20/12/2017	3	2	1.98	0.98	0.13	0.06
30/12/2017	3	2	1.89	0.90	0.10	0.05
09/01/2018	3	1	1.89	0.90	0.09	0.05
19/01/2018	3	1	2.23	0.89	0.08	0.04
29/01/2018	4	1	1.96	0.89	0.10	0.04
08/02/2018	3	1	1.99	0.89	0.11	0.04
18/02/2018	4	1	1.90	0.89	0.14	0.04
28/02/2018	3	1	1.86	0.89	0.13	0.04

Fig. 4.7 User migrations among the top six most active groups

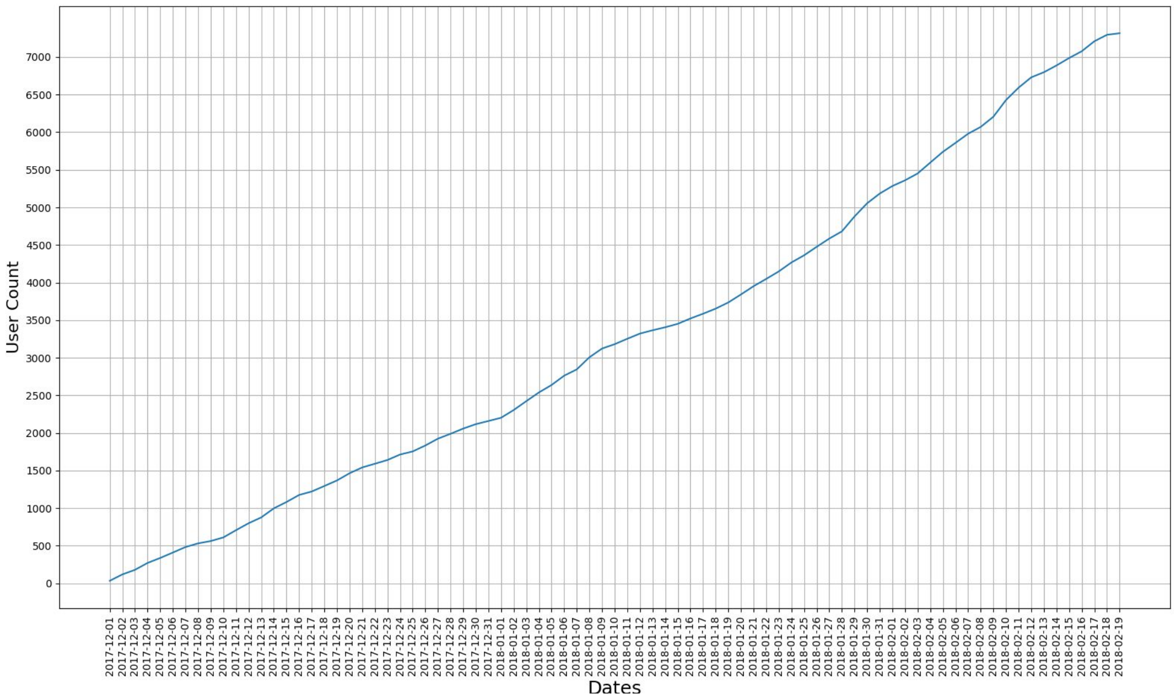


Fig. 4.8 Cumulative User Frequency Plot

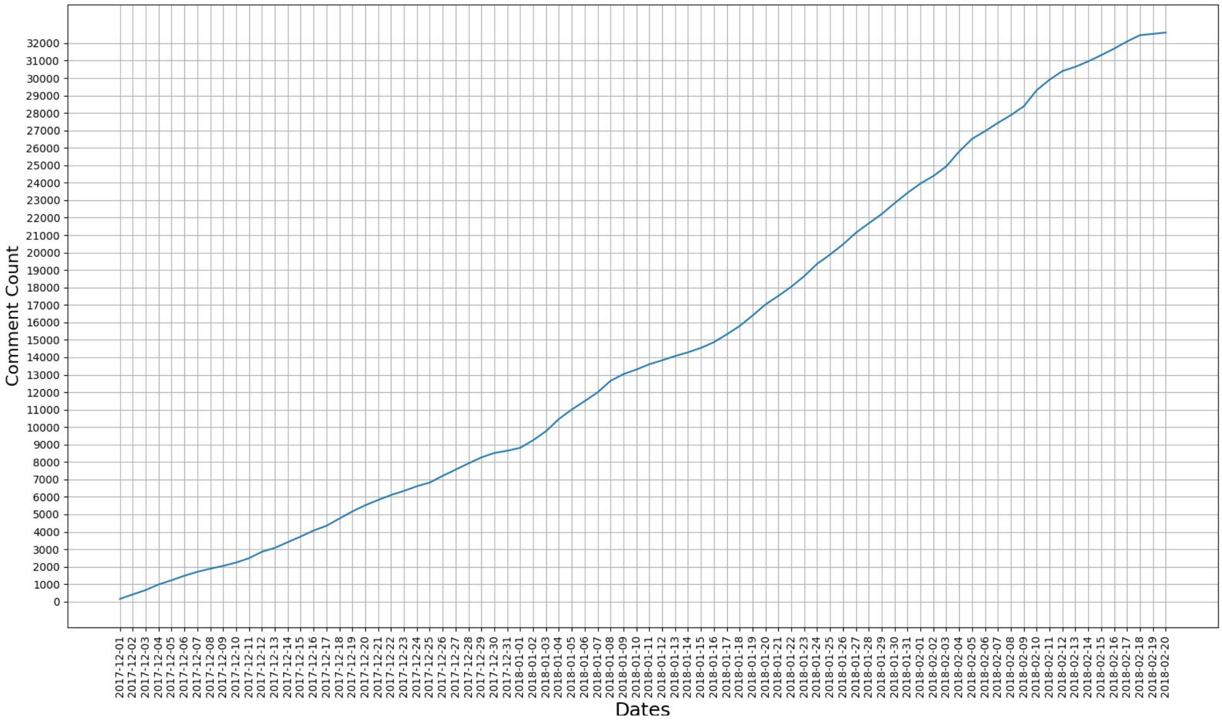


Fig. 4.9 Cumulative Comment Frequency Plot

information could be obtained from these graphs, we looked at the autocorrelation of user activity intensities.

Autocorrelation analysis can be used to find the non-randomness in time series data. Autocorrelation is a measure of correlation between the two values of the same variable at the times x_i and x_{i+k}. This can be calculated by the equation

$$R(x) = \frac{E[(X_t - \mu)(X_{t+r}) - \mu]}{\sigma^2} \tag{4.1}$$

The intensity of user activities per day can be calculated from the dataset. This information is then used to obtain the autocorrelation plot of user activities given in Fig. 4.10. The maximum lag of the plot is 81, this due to the fact that the dataset contains data for 81 days. When lag=1. the correlation value is 0.75, that means next day activities is strongly correlated with previous day activities when lag increases correlation decreases which is expected. However when lag >= 50 correlation starts increasing, and remain positive after lag>=70.

This positive increase in correlation maps with the day of the election. Thus, the change in correlation could be due to an external event, in this case, the election. This reconfirms our

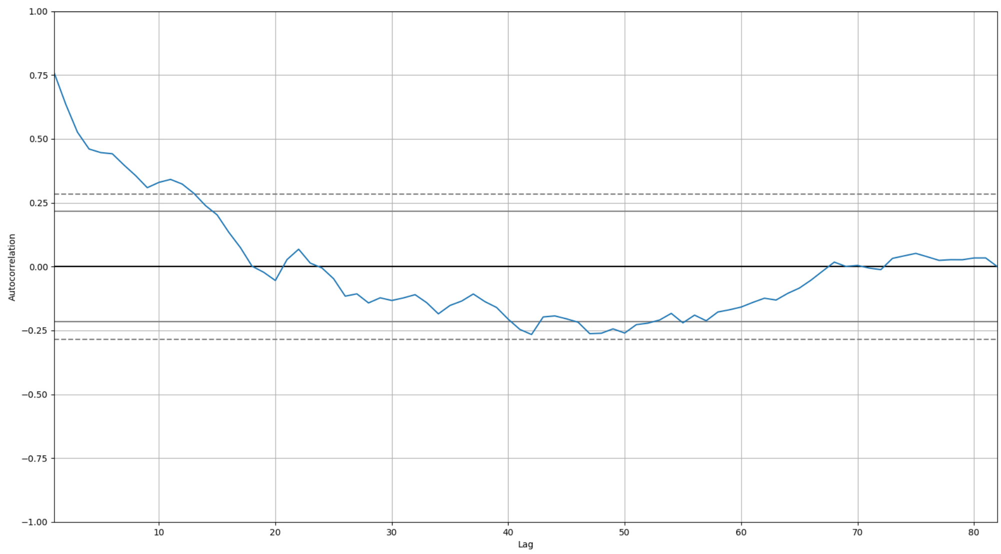


Fig. 4.10 Autocorrelation plot of daily user interactions

previous analysis that the user activities of Facebook during the Local Government Election 2018 shows a non-random process.

4.2.7 Correlation of migrations with election results

With the migration data from the microlevel analysis, we recorded the number of migrations inward and outward of each Facebook group in the dataset. We then aggregated the results according to the preferred political party of each group to get a summary of migrations of each political party as shown in Table 4.1.

With the above results, it is possible to find the net inward migrations of each party. We then ranked each political party with the based on the net inward migrations (Table 4.2). It can be observed that this rank correlates perfectly with the election results. We do not claim that the data of a single election is sufficient to justify the social media migrations depict the real-world political opinion migrations. However, existence of this correlation, supports in favor of the accuracy of our migration model.

Source Party	**Target Party**	**Count**	**Total Count**
UNP	SLPP	244	2345
	SLFP	135	
	Independent	1960	
	JVP	6	
JVP	UNP	398	762
	Independent	351	
	SLPP	13	
SLPP	UNP	183	1610
	Independent	1373	
	JVP	3	
	SLFP	51	
Independent	UNP	4074	9326
	SLPP	4375	
	SLFP	777	
	JVP	98	
	Social Democratic	2	
SLFP	UNP	17	62
	Independent	22	
	SLPP	23	

Table 4.1 Migration Counts of Political Parties

	UNP	SLPP	SLFP	JVP
Inward	4672	4655	1072	107
Outward	2345	1610	919	762
Net Inward	2327	3045	153	-655
Total Interactions in Groups of Each Party	19523	21375	1531	670
Normalized Net Inward	0.1191927	0.14245614	0.0999954	-0.97761194
Migration Rankings	2	1	3	4
Election Rankings	2	1	3	4

Table 4.2 Normalized Net Inward Migrations of Political Parties

Reaction Type	Reaction percentage	
	Interacted in a single group	Interacted in more than one group
LOVE	2.08%	2.64%
LIKE	89.08%	87.68%
WOW	0.39%	0.37%
HAHA	5.87%	6.64%
SAD	1.06%	1.11%
ANGRY	1.47%	1.55%

Table 4.3 Aggregated reaction frequencies

4.3 Reaction Analysis

Tian et al. [114] have shown that Facebook reactions are a projection of the user's sentiment. By studying temporal variation of sentiment, we can observe how users affinities change over time. Since Facebook API does not provide timestamps of likes and reactions, proceeding along this path is not possible. However, we can still study the aggregated reaction frequencies.

To this end, we compared the reactions of 39810 users who interact with a single group (Group A) and those 4288 users who interact in multiple groups (Group B). We found out that there is no substantial difference between the frequencies of reactions of these two types of users (Table 4.3). Any indication of a significant difference between the 2 groups would have denoted that user affinity changes or migrations which were previously observed could have been due to certain users intentionally trying to negatively influence the opinion of others. It is because of these reasons that it is important, we do a further thorough analysis to understand the motives behind migrations.

4.4 User Content Coding

To verify the migration model, we used manual content analysis. We extracted the comments and reactions of 21 migrating users and 24 non-migrating users which were identified using the introduce migration model. The content interactions (comments/likes) of these users were content coded to identify changes in affinity.

45 participants were involved in the task of content coding where each participant content coded the interactions of three users previously identified Facebook users. Each Facebook user's content was coded by three content coders. We tried to minimize the systematic bias that could be introduced due to content coder's background by taking into account both

male and female participants. Content coding participant pool largely consisted of university students.

After coding the interactions, we analyzed the results from coding to identify migrations. If migrations are identified, the direction of the migration was recorded. Then we compared the directions from coding and directions from the migration model to evaluate the predictions from the novel migration model. Table 4.4 shows a sample comparison that was made during this process.

Table 4.5 gives the confusion matrix of the results. It is important to note that the accuracies would be higher if we only consider whether the user is migrating or not (without considering the direction of migration).

Table 4.6 shows some of the metrics which shows the performance of the model.

Total accuracy of classifying migrating and non-migrating users is 82%. F-Score which is the harmonic mean of recall and precision gives a value of 78%. These performance measures validates the fact that migration model we developed is a good measure to identify user migrations in Facebook groups. Since there exist no other models for predicting online user migration behavior in the context of social media during elections, to best of our knowledge this is the best performing model to predict user migrations in social networks during the Local Government Election Sri Lanka 2018.

User	Group 01(From)		Group 02(To)		Heuristics		Match?
	Group Name	**Party**	**Group Name**	**Party**	**From**	**To**	
565323235656565	Group A	Party T	Group B	Independent	Group B	Group A	0
895302302556485	Group C	Independent	Group D	Party U	Group C	Group D	1
11209656201100	Group E	Party U	Group A	Party T	Group E	Group A	1

Table 4.4 Comparison of migration model and user content coding results

	Actual: Migrating	**Actual: Non-migrating**	
Predicted: Migrating	14	7	21
Predicted: Non-migrating	1	23	24
	15	30	45

Table 4.5 Confusion matrix

Performance Metric	**Value**
Accuracy	0.82
Recall (TPR)	0.93
Precision	0.67
FPR	0.23
Specificity(TNR)	0.77
F-Score	0.78

Table 4.6 Performance Metrics of Content Coding for Migrated Users

Chapter 5

Discussion

5.1 Introduction

The current growth in social networks has affected the society both in a positive and negative manner. People have started to speculate the influence of community interactions on social network or social media to influence the user affinity in areas such as politics. We thrived to understand this phenomenon through the analysis of a key social event on Facebook. The research was able to corroborate with the idea that user affinity towards Facebook groups does evolve over time. Furthermore, we introduced a novel model to identify user migrations within Facebook groups which assisted in understanding the changing or evolving user affinity. The study showed that the behavior of some users was non-random behavior within the time period of analysis.

5.2 Discussion

In the study of the understanding impact on user affinity by the community interactions, there were several major contributions. The main contribution of this research is understanding the inner dynamics of Facebook groups with political affiliations in order to understand the impact of the social network towards a key event such as an election. t assisted to understand the relationship between the community interactions and user affinity. To this end, we have conducted a micro-level analysis and a macro-level analysis. The results were evaluated with a manual content coding process. Our results show the following observations.

- Sri Lankan Facebook users are significantly more active on Facebook political groups during the election period

- An only handful number of Facebook users lead the interactions among all of the studied users
- There are users migrating among the 44 Facebook groups in the dataset that we collected
- These migrations show a significant variation from the migrations that could randomly occur in a social network
- Facebook comments and reaction behavior of the migrated users show changes in their political affinity

We represent the gathered dataset in two graph structures to carry out a micro-level and a macro-level analysis. We find that the graph structure in the first representation evolves with the time. In addition to the natural growth of clusters over time due to the increasing number of user involvement in groups as the election gets closer (if we normalize for the natural growth, the observed growth stands out significant and the election could explain the sudden growth), there are significant changes in overall positions of the clusters relative to other clusters. Two clusters get closer together if the forces between them increase. This happens when,

1. Existing users in one cluster start to interact with the users on the other cluster more often than before. These user interactions could be mostly based on the major event i.e., the election. We need more work to understand whether such interactions are to re-affirm their shared political beliefs or to counter argue with another. Eventually, it is imperative to understand how such interaction affects each other's own beliefs and subsequent interactions with the social network. Such an analysis conclusively says the nature of the impact of the social networks towards crucial events such as an election.
2. A completely new set of users who have not previously interacted with either of the groups start interacting with both. This is an interesting observation and we can hypothesize two scenarios; a) users are less likely to engage in political conversations are now drawn into the groups due to the intense political climate surrounding the election and, b) the other more concerning reason could be, these are bots (or trolls, based on fake accounts) trying to disseminate fake news and negatively influence the opinion of the general public. Hence it is important we do a further thorough analysis to understand the motives behind such new users.

A sudden spike of activities was observed on many users in the same day, one of which is shown in Fig. 4.2 Even though the reason for these spikes is not clear at this point, some external event could have triggered these activities. It is worth investigating the factors which have caused this scenario in follow up research. Especially, to see how people are reacting; we can only see the increased participation but we need to understand their motives as well.

Our micro-level analysis results show two more significant observations;

1. Sri Lankan Facebook users are significantly more active on political groups during elections

2. Only handful number of users lead the interactions among all of the users

Through the macro-level analysis, we observe that groups which correspond to clusters D and B (Fig 4.5) which self-identify themselves as having 2 different political opinions are much closer in the final graph, taken after the election. Among many hypotheses to explain this observation few plausible ones are, a) users of the group who actually won the election are now mingling the other group to intimidate, and/or b) users are reconciling their beliefs and now coming into terms.

Cluster E which is an independent community appears to be influenced by cluster D which is a group belonging to a certain political party. This observation could be more concerning since a politically motivated group is influencing a seemingly an independent group which is more likely to have a negative impact on the big picture of the impact of the social networks for an election.

It is observed in Fig. 4.5 that edge weight from D to B is higher than that of B to D. This implies that comparatively higher number of users have interacted in group B after interacting with group D. This is consistent with the observations in Fig. 4.1. Similarly, edge weight from G to D is again consistent with the Fig. 4.1. In addition, it can be noted that there is a considerable disparity in the edge weights from G to B and C to B than that of B to G and B to C. This is an indication that a lot of users who have previously interacted in groups C and G have later interacted with the group B. Therefore it can be suspected that the self-identifying independent groups C and G have had some sort of an influence from group B which is a group representing a major political party.

In trying to understand the rationale behind their migration, we found that the political affiliations of these groups and the respective user migrations align well with the outcome of the election that we have observed. The parties that did well in the elections saw their affiliated groups getting a lot of users migrated into compared to the other groups with different affiliations. This affirms that Facebook can be predictable of the outcome of the election. However, the key question is to understand what made these users migrate to a different group. Is it something happened in the real world or some other influence they had on the social network itself?

Further, after evaluating the migrations via content coding we observe that a considerable proportion of the migrated users show a change in affinity towards political parties. This confirms that observing the migration patterns is a viable and effective way to measure changes in user affinity.

Content coding that we did for the migrated users shows an accuracy of 82 percent and non-migrated users show an accuracy of 95.83 percent. So accuracy of the human-coded content model is high which means that the results we obtained via graphical methods are very much true.

Firstly we looked at the data set with a metric perspective via graph visualizations which can be subjective. During the content coding, we looked at the data set with a human perspective allowing human subjects to do the content-coding. Metric perspective showed that there is a migration within the Facebook groups and 82 percent accuracy in the human perspective confirms that metric perspective and human perspective tallies in this case allowing a true performance of the content coding model that we used.

We have shown that Facebook is a good representative measurement of key real-world events such as an election. However, the key question is whether Facebook is reactive to real-world events or proactive in swaying users opinions and in turn affecting the election. Evidence supporting the latter of the above is visible in the content coded user interactions. User migrations between opposing political parties were observed prior to the election and they imply changes in affinity. Therefore, it is evident that Facebook user behavior is a good measure to predict the outcome of an election. This claim needs to be validated further by observing the social network user dynamics of future elections. However, to understand whether there was an active manipulation of user opinion, content should be analyzed with much scrutiny in the future.

We also want to mention that during the course of the data collection, we only used publicly available APIs provided by Facebook to only access publicly available posts and other Facebook interactions. All of the collected data was properly anonymized to prevent identifying individuals behind each interaction.

5.3 Conclusion

We conducted this research with the primary purpose of understanding and describing the impact of user interactions in social networks and its influence on the user affinity. We attempted to understand the relationship between these two properties. Furthermore, we examined how the user affinity evolves over time around a key social event.

In process of tackling the research question, 'How does the community interactions on social media influence the user affinity?' we were able to get a sound understanding of the community interactions and the user affinity in the network. We used data from Facebook groups during the pre-election and post-election period of the Local Government Election 2018 in Sri Lanka. We identified ways to measure or capture user affinity using the community interactions. However, we were not able to answer the above-mentioned question to its entirety but we were able to lay the foundation and preliminary analysis to answer this question. It will be very beneficial in identifying the causal relationship between the community interactions as it will open doors to a huge arena of studies in the field of social networks. It will help the social network participants to carry out the daily activities in a more precautionary manner as well.

However, we were able to understand the evolving nature of user affinity. As means to achieving this goal, we first identified a way to quantify the user affinity. One of the most significant contributions of our study, the migration model was a novel approach in identifying the changes in user affinity. This model was evaluated with the use of content coding by 45 different users. During the study, we understood that user affinity can be measured using the community interactions itself which establishes a more solid understanding of the relationship between the community interactions and the user affinity. Furthermore, we identified the non-random nature of these migrations. We believe this finding will motivate the research community to identify the actual factors for changes in user affinity. It will create a deeper understanding on the actual nature of how user affinity is swayed within a key social event. It will assist to identify the true nature of social networks such as Facebook as to being either reactive or proactive to social events.

5.4 Future Work

As future work, a natural extension of this work entails the investigating the possibility of predicting user migrations with the use of data collected before the election. Graph analysis techniques like link prediction already exist which may assist in this task. Further, it is possible to uncover the network contagions that respond to external events in a similar manner.

The micro-level results pointed out that there are groups where interactions are very high while there are groups which get substantially very less attention from the users. It is an important observation which requires further research to identifying the underlying reasons for it.

With the basis we have established with the findings of this research, it is possible to conduct predictive and prescriptive analysis in future. Understanding the factors determining the fluctuations in user affinity aids in predicting the future state of a network of users given the initial network structure. The ability to predict such evolution of a community makes it possible to predict the general future opinion of the community with a certain degree of accuracy.

The migration model which is a core part of the research, has only been evaluated with the Local Government Election of Sri Lanka 2018. It is rather infeasible to verify the model against previous elections because social media content posted during the past elections may have been altered or deleted by now. However, it is possible to apply the model to upcoming elections and observe its performance.

References

[1] J. Chukwuere and P. Chukwuere, “The impact of social media on social lifestyle: A case study of university female students,” *Gender and Behaviour*, vol. 15, pp. 9966 – 9981, 12 2017.

[2] L. Margalit, “The psychology behind social media interactions,” 2014. [Online]. Available: https://www.psychologytoday.com/us/blog/behind-online-behavior/201408/the-psychology-behind-social-media-interactions

[3] A. J. News, “Did sri lanka’s facebook ban help quell anti-muslim violence? | news,” https://www.aljazeera.com/news/2018/03/sri-lanka-facebook-ban-quell-anti-muslim-violence-180314010521978.html, (Accessed on 05/20/2018).

[4] T. Baruah, “Effectiveness of social media as a tool of communication and its potential for te chnology enabled connections: A micro - level study,” *International Journal of Scientific and Research Publications*, vol. 2, no. 5, 2012. [Online]. Available: http://www.ijsrp.org/research_paper_may2012/ijsrp-may-2012-24.pdf

[5] 2018. [Online]. Available: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

[6] C. W. et al., “User interactions in social networks and their implications,” *Proceedings of the fourth ACM European conference on Computer systems*, pp. 205–218, 2009.

[7] M. Hicks, “Facebook tips: What’s the difference between a facebook page and group?” Feb 2010. [Online]. Available: https://www.facebook.com/notes/facebook/facebook-tips-whats-the-difference-between-a-facebook-page-and-group/324706977130/

[8] J. Raudeliuniene, V. Davidaviciene, M. Tvaronaviciene, and L. Jonuška, *Evaluation of Advertising Campaigns on Social Media Networks*, 2018. [Online]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwj_hP_Nrs7fAhULTo8KHV-zCPkQFjAAegQIChAC&url=http%3A%2F%2Fwww.mdpi.com%2F2071-1050%2F10%2F4%2F973%2Fpdf&usg=AOvVaw0ftozdi5XZo-xgKVnNapZ3

[9] P. N. Howard, B. Kollanyi, S. Bradshaw, and L.-M. Neudert, “Social media, news and political information during the us election: Was polarizing content concentrated in swing states?” *arXiv preprint arXiv:1802.03573*, 2018.

[10] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild *et al.*, "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[11] M. Bossetta, A. D. Segesten, and H.-J. Trenz, "Political participation on facebook during brexit," 2017.

[12] 2018. [Online]. Available: http://www.dailymirror.lk/85811/social-media-and-general-elecations-2015

[13] N. Merchant, "Customer affinity: How to get it, what to do with it." [Online]. Available: https://www.pragmaticmarketing.com/resources/articles/customer-affinity-how-to-get-it-what-to-do-with-it

[14] M. Renfree, "Campaign communications: A look at obama's social media success," 2016. [Online]. Available: https://www.prnewsonline.com/obama-social-media

[15] S. L. C. Wang and W. Fang, "Extending the technology acceptance model to mobile telecommunication innovation: The existence of network externalities," *Journal of Consumer Behaviour*, vol. 7, pp. 101–110, 2008.

[16] G. Kazai, I. Yusof, and D. Clarke, "Personalised news and blog recommendations based on user location, facebook and twitter user profiling," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 2016, pp. 1129–1132.

[17] M. A. Moreno, R. Kota, S. Schoohs, and J. M. Whitehill, "The facebook influence model: A concept mapping approach," *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 7, pp. 504–511, 2013.

[18] H. Patil and M. Atique, *Sentiment Analysis for Social Media: A Survey*. IEEE, 2015. [Online]. Available: https://ieeexplore.ieee.org/document/7371033

[19] R. Panigrahy, M. Najork, and Y. Xie, "How user behavior is related to social affinity," in *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*. ACM Press, 2012. [Online]. Available: https://doi.org/10.1145/2124295.2124379

[20] 2018. [Online]. Available: http://gs.statcounter.com/social-media-stats/all/sri-lanka

[21] *Research Essentials: An Introduction to Designs and Practices*. Jossey-Bass, 2009. [Online]. Available: https://www.amazon.com/Research-Essentials-Introduction-Designs-Practices/dp/0470181095?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0470181095

[22] P. Price, R. Jhangiani, and I. Chiang, *Research Methods in Psychology*. BCcampus, BC Open Textbook Project, 2014. [Online]. Available: https://books.google.lk/books?id=FEcDtQEACAAJ

[23] L. M. E, Ed., *The SAGE Encyclopedia of Qualitative Research Methods*. SAGE Publications, Inc, 2008. [Online]. Available: https://www.amazon.com/SAGE-Encyclopedia-Qualitative-Research-Methods/dp/1412941636?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=1412941636

[24] B. L. Berg, *Qualitative Research Methods for the Social Sciences (4th Edition)*. Allyn & Bacon, 2000. [Online]. Available: https://www.amazon.com/Qualitative-Research-Methods-Social-Sciences/dp/0205318479?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0205318479

[25] A. Bhattacherjee, *Social Science Research: Principles, Methods, and Practices*, 2012.

[26] "Election losses test sri lanka's leader, and the country's direction - the new york times," https://www.nytimes.com/2018/02/11/world/asia/sri-lanka-elections.html, (Accessed on 05/20/2018).

[27] "Facebook graph api," 2018. [Online]. Available: https://developers.facebook.com/docs/graph-api

[28] A. J. News, "Did sri lanka's facebook ban help quell anti-muslim violence?" https://www.aljazeera.com/news/2018/03/sri-lanka-facebook-ban-quell-anti-muslim-violence-180314010521978.html, (Accessed on 05/20/2018).

[29] D. Terrana, A. Augello, and G. Pilato, "Analysis of Facebook Users' Relationships Through Sentiment Classification: A Case Study of Italian Politicians," *International Journal of Semantic Computing*, vol. 08, no. 03, pp. 301–317, 2014. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S1793351X14400108

[30] R. Lisa, *Social Influence*. American Cancer Society, 2007, pp. 4426–4429. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405165518.wbeoss154

[31] W. A. Mason, F. R. Conrey, and E. R. Smith, "Situating social influence processes: Dynamic, multidirectional flows of influence within social networks," *Personality and Social Psychology Review*, vol. 11, no. 3, pp. 279–300, aug 2007. [Online]. Available: https://doi.org/10.1177/1088868307301032

[32] J. P. Forgas and K. D. Williams, *Social influence: direct and indirect processes*. Psychology Press, 2001.

[33] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03*. ACM Press, 2003. [Online]. Available: https://doi.org/10.1145/956750.956769

[34] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan, "Influence maximization in social networks when negative opinions may emerge and propagate," in *Proceedings of the 2011 SIAM International*

Conference on Data Mining. Society for Industrial and Applied Mathematics, apr 2011, pp. 379–390. [Online]. Available: https://doi.org/10.1137/1.9781611972818.33

[35] E. Mossel and S. Roch, "On the submodularity of influence in social networks," in *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing*, ser. STOC '07. New York, NY, USA: ACM, 2007, pp. 128–134. [Online]. Available: http://doi.acm.org/10.1145/1250790.1250811

[36] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '09. New York, NY, USA: ACM, 2009, pp. 199–208. [Online]. Available: http://doi.acm.org/10.1145/1557019.1557047

[37] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler, "A 61-million-person experiment in social influence and political mobilization," *Nature*, vol. 489, no. 7415, pp. 295–298, sep 2012. [Online]. Available: https://doi.org/10.1038/nature11421

[38] M. A. Moreno, R. Kota, S. Schoohs, and J. M. Whitehill, "The facebook influence model: A concept mapping approach," *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 7, pp. 504–511, jul 2013. [Online]. Available: https://doi.org/10.1089/cyber.2013.0025

[39] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '09. New York, NY, USA: ACM, 2009, pp. 807–816. [Online]. Available: http://doi.acm.org/10.1145/1557019.1557108

[40] N. S. Contractor and L. A. DeChurch, "Integrating social networks and human social motives to achieve social influence at scale," *Proceedings of the National Academy of Sciences*, vol. 111, no. Supplement_4, pp. 13 650–13 657, sep 2014. [Online]. Available: https://doi.org/10.1073/pnas.1401211111

[41] J. A. Vela-McConnell, *Who Is My Neighbor?: Social Affinity in a Modern World*. State University of New York Press, 1999.

[42] J. L. Iribarren and E. Moro, "Affinity paths and information diffusion in social networks," *Social Networks*, vol. 33, no. 2, pp. 134 – 142, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378873310000596

[43] J. Scott, *Social Network Analysis*. SAGE Publications Ltd, mar 2017. [Online]. Available: https://www.xarg.org/ref/a/1473952115/

[44] E. Otte and R. Rousseau, "Social network analysis: a powerful strategy, also for the information sciences," 2002. [Online]. Available: http://journals.sagepub.com/doi/abs/10.1177/016555150202800601

[45] S. Wasserman, *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge University Press, nov 1994. [Online]. Available: https://www.xarg.org/ref/a/0521387078/

[46] I. Gečienė, "Social network analysis in migration studies," *Sociology. Thought and Action*, vol. 25, pp. 130–143, Dec. 2009. [Online]. Available: http://www.journals.vu.lt/sociologija-mintis-ir-veiksmas/article/view/6091

[47] *Applications of Social Media Text Analysis*. [Online]. Available: https://www.cs.cmu.edu/~ark/EMNLP-2015/proceedings/EMNLP-Tutorials/pdf/EMNLP-Tutorials06.pdf

[48] R. Guimera, S. Mossa, A. Turtschi, and L. A. N. Amaral, "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," *Proceedings of the National Academy of Sciences*, vol. 102, no. 22, pp. 7794–7799, may 2005. [Online]. Available: https://doi.org/10.1073/pnas.0407994102

[49] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '06. New York, NY, USA: ACM, 2006, pp. 611–617. [Online]. Available: http://doi.acm.org/10.1145/1150402.1150476

[50] M. Dado and D. Bodemer, "A review of methodological applications of social network analysis in computer-supported collaborative learning," *Educational Research Review*, vol. 22, pp. 159 – 180, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1747938X17300325

[51] D. Snyder and E. Kick, "Structural position in the world system and economic growth, 1955-1970: A multiple-network analysis of transnational interactions," *American Journal of Sociology - AMER J SOCIOL*, vol. 84, 03 1979.

[52] N. E. Friedkin and E. C. Johnsen, "Social positions in influence networks," *Social Networks*, vol. 19, no. 3, pp. 209 – 222, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378873396002985

[53] N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *New England Journal of Medicine*, vol. 357, no. 4, pp. 370–379, 2007, pMID: 17652652. [Online]. Available: https://doi.org/10.1056/NEJMsa066082

[54] D. B. Bahr, R. C. Browning, H. R. Wyatt, and J. O. Hill, "Exploiting social networks to mitigate the obesity epidemic," *Obesity*, vol. 17, no. 4, pp. 723–728. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1038/oby.2008.615

[55] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of facebook â€œfriends:â€ social capital and college studentsâ€™ use of online social network sites," *Journal of Computer-Mediated Communication*, vol. 12, no. 4, pp. 1143–1168. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1083-6101.2007.00367.x

[56] W. Tsai and S. Ghoshal, "Social capital and value creation: The role of intrafirm networks," *Academy of Management Journal*, vol. 41, no. 4, pp. 464–476, 1998. [Online]. Available: https://doi.org/10.5465/257085

[57] J. Barnes and F. Harary, "Graph theory in network analysis," *Social Networks*, vol. 5, no. 2, pp. 235 – 244, 1983. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0378873383900266

[58] S. Gozzo and V. Tomaselli, “Network analysis methods to measure sociometric status,” *Mathematics and Statistics*, vol. 4, pp. 81–94, 08 2016.

[59] V. Batagelj and S. Praprotnik, “An algebraic approach to temporal network analysis based on temporal quantities,” *CoRR*, vol. abs/1505.01569, 2015. [Online]. Available: http://arxiv.org/abs/1505.01569

[60] J. Barnes and F. Harary, “Graph theory in network analysis,” *Social Networks*, vol. 5, no. 2, pp. 235–244, jun 1983. [Online]. Available: https://doi.org/10.1016/0378-8733(83)90026-6

[61] M. Newman, A.-L. Barabasi, and D. J. Watts, *The Structure and Dynamics of Networks: (Princeton Studies in Complexity)*. Princeton, NJ, USA: Princeton University Press, 2006.

[62] M. E. J. Newman, *Networks*. Oxford University Press Inc., 2010.

[63] P. Erdős and A. Rényi, “On the evolution of random graphs,” in *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*, 1960, pp. 17–61.

[64] E. N. Gilbert, “"random graphs",” *"Annals of Mathematica Statistics"*, 1959.

[65] E. Bullmore and O. Sporns, “Complex brain networks: graph theoretical analysis of structural and functional systems,” *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, feb 2009. [Online]. Available: https://doi.org/10.1038/nrn2575

[66] W. Chih-Chien, L. Shao-Kang, and F. Wenchang, “Extending the technology acceptance model to mobile telecommunication innovation: The existence of network externalities,” *Journal of Consumer Behaviour*, vol. 7, no. 2, pp. 101–110. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cb.240

[67] S. Aouay, S. Jamoussi, F. Gargouri, and A. Abraham, “Modeling dynamics of social networks: A survey,” in *2014 6th International Conference on Computational Aspects of Social Networks*. IEEE, jul 2014. [Online]. Available: https://doi.org/10.1109/cason.2014.6920421

[68] J.-L. Guillaume and M. Latapy, “A realistic model for complex networks,” 2003.

[69] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, “Random graph models of social networks,” *Proceedings of the National Academy of Sciences*, vol. 99, no. Supplement 1, pp. 2566–2572, feb 2002. [Online]. Available: https://doi.org/10.1073/pnas.012582999

[70] D. Watts and S. Strogatz, “Collective dynamics of ‘small-world’ networks,” *nature*, vol. 393, pp. 440–442, 1998. [Online]. Available: https://www.nature.com/articles/30918

[71] H. Zhang, A. Goel, and R. Govindan, “Using the small-world model to improve freenet performance,” *Computer Networks*, vol. 46, no. 4, pp. 555 – 574, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128604001550

[72] M. Jovanovi, "Modeling peer-to-peer network topologies through " small-world " models and power laws," 2001.

[73] M. E. J. Newman and D. J. Watts, "Scaling and percolation in the small-world network model," *Phys. Rev. E*, vol. 60, pp. 7332–7342, Dec 1999. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.60.7332

[74] M. E. J. Newman, C. Moore, and D. J. Watts, "Mean-field solution of the small-world network model," *Phys. Rev. Lett.*, vol. 84, pp. 3201–3204, Apr 2000. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.84.3201

[75] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999. [Online]. Available: http://science.sciencemag.org/content/286/5439/509

[76] A.-L. Barabási, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: the topology of the world-wide web," *Physica A: Statistical Mechanics and its Applications*, vol. 281, no. 1, pp. 69 – 77, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437100000182

[77] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Phys. Rev. Lett.*, vol. 86, pp. 3200–3203, Apr 2001. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.86.3200

[78] S. Mossa, M. Barthélémy, H. Eugene Stanley, and L. A. Nunes Amaral, "Truncation of power law behavior in "scale-free" network models due to information filtering," *Phys. Rev. Lett.*, vol. 88, p. 138701, Mar 2002. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.88.138701

[79] S. S. Manna and P. Sen, "Modulated scale-free network in euclidean space," *Phys. Rev. E*, vol. 66, p. 066114, Dec 2002. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.66.066114

[80] T. Berger-Wolf and J. Saia, "A framework for analysis of dynamic social networks," *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 523–528, 2006. [Online]. Available: http://dl.acm.org/citation.cfm?id=1150462

[81] M. Del Vicario, Q. Zhang, A. Bessi, G. Caldarelli, and F. Zollo, "Structural patterns of the occupy movement on facebook," *Studies in Computational Intelligence*, vol. 693, pp. 595–606, 2017.

[82] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, "Microscopic evolution of social networks," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '08. New York, NY, USA: ACM, 2008, pp. 462–470. [Online]. Available: http://doi.acm.org/10.1145/1401890.1401948

[83] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," in *Proceedings of the 16th International Conference on World Wide Web*, ser. WWW

’07. New York, NY, USA: ACM, 2007, pp. 835–844. [Online]. Available: http://doi.acm.org/10.1145/1242572.1242685

[84] W. J. Burk, C. E. Steglich, and T. A. Snijders, “Beyond dyadic interdependence: Actor-oriented models for co-evolving social networks and individual behaviors,” *International Journal of Behavioral Development*, vol. 31, no. 4, pp. 397–404, 2007. [Online]. Available: https://doi.org/10.1177/0165025407077762

[85] 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*, ser. WOSN ’09. New York, NY, USA: ACM, 2009, pp. 37–42. [Online]. Available: http://doi.acm.org/10.1145/1592665.1592675

[86] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, “Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software,” *PLOS ONE*, vol. 9, no. 6, pp. 1–12, 06 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0098679

[87] L. Freeman, “Visualizing social networks,” *Journal of Social Structure*. [Online]. Available: http://social.cs.uiuc.edu/class/papers/freeman.pdf

[88] T. FRUCHTERMAN and E. REINGOLD, *Graph Drawing by Force-directed Placement*, 1991. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.13.8444&rep=rep1&type=pdf

[89] C. Dunne and B. Shneiderman, “Improving graph drawing readability by incorporating readability metrics: A software tool for network analysts,” *University of Maryland, HCIL Tech Report HCIL-2009-13*, 2009/// 2009.

[90] B. Martínez-López, A. M. Perez, and J. M. Sánchez-Vizcaíno, “Social network analysis. review of general concepts and use in preventive veterinary medicine,” *Transboundary and Emerging Diseases*, vol. 56, no. 4, pp. 109–120, 2009. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/19341388

[91] 2015. [Online]. Available: http://www.insna.org/

[92] R. Rogers, *Digital methods*. Mit Press, 2015.

[93] I. Herman, G. Melancon, and M. Marshall, “Graph visualization and navigation in information visualization: A survey,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 6, no. 1, pp. 24–43, 2000. [Online]. Available: https://doi.org/10.1109/2945.841119

[94] B. Wellman, *Structural analysis: from method and metaphor to theory and substance*. [Online]. Available: http://homepage.ntu.edu.tw/~khsu/network/reading/wellman2.pdf

[95] D. J. Watts, *Six degrees*. Vintage Books, 2004.

[96] M. Granovetter, “The strength of weak ties: A network theory revisited,” *Sociological Theory*, vol. 1, no. 78, p. 201, 1983. [Online]. Available: https://www.journals.uchicago.edu/doi/10.1086/225469

[97] C. Tan, J. Tang, J. Sun, Q. Lin, and F. Wang, *Social action tracking via noise tolerant time-varying factor graphs*. ACM, 2010. [Online]. Available: https://dl.acm.org/citation.cfm?id=1835936

[98] G. Kossinets, J. Kleinberg, and D. Watts, *The Structure of Inf ormation P athwa ys in a Social Comm unication Netw ork*. Cornell University, 2018. [Online]. Available: https://www.cs.cornell.edu/home/kleinber/kdd08-bb.pdf

[99] D. Cosley, D. Huttenlocher, J. Kleinberg, X. Lan, and S. Suri, *Sequential Influence Models in Social Networks*. [Online]. Available: https://www.cs.cornell.edu/home/kleinber/icwsm10-seq.pdf

[100] L. Freeman, *Centrality in Social Networks Conceptual Clarification*, 1st ed. Elsevier, 1979. [Online]. Available: http://leonidzhukov.net/hse/2014/socialnetworks/papers/freeman79-centrality.pdf

[101] A. Bavelas, “A mathematical model for group structures,” *Human Organization*, vol. 7, no. 3, pp. 16–30, 1948. [Online]. Available: http://www.sfaajournals.net/doi/abs/10.17730/humo.7.3.f4033344851gl053?code=apan-site&journalCode=humo

[102] D. R. Thomas, “A general inductive approach for analyzing qualitative evaluation data,” *American Journal of Evaluation*, vol. 27, no. 2, pp. 237–246, jun 2006. [Online]. Available: https://doi.org/10.1177/1098214005283748

[103] S. Berg, *Snowball Samplingâ€”I*. American Cancer Society, 2006. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/0471667196.ess2478.pub2

[104] B. Rieder, “Studying Facebook via data extraction,” *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci ’13*, pp. 346–355, 2013. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2464464.2464475

[105] H. R. Bernard, *Social Research Methods: Qualitative and Quantitative Approaches*. SAGE Publications, Inc, 2012. [Online]. Available: https://www.amazon.com/Social-Research-Methods-Qualitative-Quantitative/dp/1412978548?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=1412978548

[106] “Gephi - the open graph viz platform,” https://gephi.org/, (Accessed on 08/19/2018).

[107] “Snap: Stanford network analysis project,” http://snap.stanford.edu/, (Accessed on 08/19/2018).

[108] P. Erdos and A. Renyi, *On the evolution of random graphs*. [Online]. Available: http://snap.stanford.edu/class/cs224w-readings/erdos60random.pdf

[109] S. Milojević, “Power law distributions in information science: Making the case for logarithmic binning,” *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2417–2425. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21426

[110] M. Newman, "Power laws, pareto distributions and zipf's law," *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, 2005. [Online]. Available: https://doi.org/10.1080/00107510500052444

[111] M. Riccaboni and S. Schiavo, "Structure and growth of weighted networks," *New Journal of Physics*, vol. 12, no. 2, p. 023003, feb 2010. [Online]. Available: https://doi.org/10.1088/1367-2630/12/2/023003

[112] G. Fagiolo, J. Reyes, and S. Schiavo, "The evolution of the world trade web: a weighted-network analysis," *Journal of Evolutionary Economics*, vol. 20, no. 4, pp. 479–514, jul 2009. [Online]. Available: https://doi.org/10.1007/s00191-009-0160-x

[113] S. Schiavo, J. Reyes, and G. Fagiolo, "International trade and financial integration: a weighted network analysis," *Quantitative Finance*, vol. 10, no. 4, pp. 389–399, 2010. [Online]. Available: https://doi.org/10.1080/14697680902882420

[114] Y. Tian, T. Galery, G. Dulcinati, E. Molimpakis, and C. Sun, "Facebook sentiment: Reactions and emojis," in *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*. Association for Computational Linguistics, 2017. [Online]. Available: https://doi.org/10.18653/v1/w17-1102

Appendix A

Streaming Massive Graphs to Gephi

Due to Gephi's memory limitations, it is difficult to visualize dynamic graphs with the inbuilt operators. In order to circumvent the limitations a python script was written which dynamically adds and removes the relevant nodes and edges based on timestamp. It is built on top of an existing plugin named pygephi.

```
#!/usr/bin/python
# coding: utf-8

'''
Use this script with Gephi and Graph Streaming plugin.
1. Open Gephi, create a new project
2. Go to the tab Streaming, right-click on "Master Server", click on "
    Start"
3. Go to the tab Layout, select "Force Atlas" and click "Run"
4. Run this script
'''

import time
import pygephi
import random
import unicodecsv as csv

# Connect to Gephi
g = pygephi.GephiClient('http://localhost:8081/workspace1', autoflush=
    True)
g.clean()
node_attributes = {"size":10, 'r':1.0, 'g':0.0, 'b':0.0, 'x':1}

#create graph
f = open('dump/edgelist.csv','r')
```

```python
data = list(csv.reader(f, encoding='utf-8'))
data = sorted(data, key=lambda x: x[2])
total_interaction_count = len(data)

d = {}
known_groups = {}
edges = {}
counter = 1

startTime = int(data[0][2])

def process_user(user, g, groupName):
    # if the user is new, record the group current interaction happens
    # increment the size of the group node
    if user not in d:
        d[user] = groupName
        known_groups[groupName] += 1
        new_size = known_groups[groupName]
        g.change_node(groupName, **{"weight": new_size})

    else:
        old_group = d[user]
        if old_group == groupName:
            # self edge. Increase edge size
            pass
        else:
            # increase node size of groupName. AKA new group
            # create edge to groupName from old_group

            known_groups[groupName] += 1
            new_size = known_groups[groupName]
            g.change_node(groupName, **{"weight": new_size})

            edge_id = old_group + "_" + groupName
            print user, edge_id
            # if the edge is new, initialize the weight
            # if it's an existing edge, increment the weight
            if edge_id not in edges:
                edges[edge_id] = 1
                g.add_edge(edge_id, old_group, groupName, **{"weight":
    edges[edge_id]})
            else:
                edges[edge_id] += 1
                g.delete_edge(edge_id)
```

```
                g.add_edge(edge_id, old_group, groupName, **{"weight":
edges[edge_id]})
            edge_weight = edges[edge_id]

for dp in data:
    counter += 1
    user1 = dp[0]
    user2 = dp[1]
    nodeTime = dp[2]
    groupName = dp[3]
    node_attributes['groupName'] = groupName
    while startTime < int(nodeTime):
        startTime += 100000000
        print time.strftime('%m/%d/%Y %H:%M:%S', time.gmtime(startTime
/1000))

    # if the group is not in graph create a new node and store its size
    # size represents the number of interactions that happen in that
group
    if groupName not in known_groups:
        known_groups[groupName] = 0
        g.add_node(groupName, **{"weight":0})


    process_user(user1, g, groupName)
    process_user(user2, g, groupName)
```

Appendix B

Document Structure of the Content Coding

interaction Type	commentText	postText	Type	groupId	timestamp	link	Supportive	hateSpeech
comment	jayawewa	xxxxx		xxxxx	xxxxx	xxxxx	1	0
comment	Mulleriyawa Manasika rohalenda	xxxxx		xxxxx	xxxx	xxxxx	0	0
comment	Gon Katha kiyanna epa	xxxxx		xxxxx	xxxx	xxxxx	0	1
comment	meharakta oya widihata	xxxxx		xxxxx	xxxx	xxxxx	0	1
reaction		xxxxx	LIKE	xxxx		xxx		

Table B.1 Content coding document structure given to a single user