

Communities of Tweeples:
How Communities Engage with Microblogging When Co-located

Edgardo Luis Vega

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Computer Science

Manuel A. Perez-Quinones (Chair)
Steve Harrison
Andrea L. Kavanaugh

April 27th, 2011
Blacksburg, VA

Keywords: Twitter, Social, Conferences, Network, Communication, Graph, Community,
Influence, Information, Microblogging

Communities of Tweeples:
How Communities Engaged with Microblogging When Co-located

Edgardo L Vega

Abstract

Most of the research done on microblogging services, such as Twitter, has focused on how the individual communicates with their community at a micro and macro level; less research has been done on how the community affects the individual. We present in this thesis some ideas about this phenomenon.

We do this by collecting data of Twitter users at a conference. We collected 21,150 tweets from approximately 400 users during a five week period and additionally collected survey data from a small subset of the tweeters.

By observing users of Twitter, before, during, after a specific event we discovered a pattern in postings. Specifically, we found that tweets increased the week of the conference and that by the end of the conference the network was strong. These findings lead us to conclude that collocation of communities, like conferences, has a substantial effect on online microblogging behaviors.

To my loving and supportive family

Acknowledgements

I would like to thank the many people who have helped me through the completion of my thesis. The first is my advisor, Dr. Manuel A. Perez-Quinones (@mapq), who never doubted that I would finish. He was always there when I needed him to guide the way or just someone to talk to. I was also lucky enough to work with two supportive and vibrant committee members; Dr. Andrea L. Kavanaugh (@akavanaugh) and Steve Harrison. I would also like to thank the Computer Science Department at Virginia Tech for providing an excellent environment of learning, exploration, and collaboration.

I would like to thank those instrumental in starting a Masters degree in the first place; Laurian Vega (@ladylaurian), Doug McMurtrie, and Dr. Manuel A. Perez-Quinones (@mapq). Laurian is my wife and the love of my life thank you for always pushing me to be the best I can be. Doug as my boss pushed me to further my education on a regular basis until I gave in. Doug you started as my boss, but I am proud to call you a friend and a mentor. Finally, Dr. Manuel A. Perez-Quinones for just saying one day “why not?”

I am truly grateful to be able to do a thesis on a product that I was passionate about. I would like to thank all the entire Twitter team (@twitter) and building something truly amazing. But I must specifically call out Taylor Singletary (@episode) who swapped numerous emails with me and provided support.

To the participants of my thesis I would like to thank the Anita Borg Institute for Women and Technology and all the people who put on the amazing Grace Hopper Celebration of Women in Computing Conference. Specifically BJ Wishinsky who showed genuine enthusiasm for the work I was doing and answer all my questions.

I did have some help along the way. I worked with two amazing student who made contributions to this work early on: Josette Torres (@girlinblack) and Ramanujam Parthasarathy (@ramanujam).

I have also had an amazing support network of family who have been there along the way. This includes, but is not limited to Laurian Vega (always being there in just the right way), Cameron Vega (you rocked me world but only for the better), Edgar and Nitza Vega (the encouragement and love to keep going), Raiza Vega (providing motivation in betting you in one more thing), Richard and Julia Hobby (for giving me my wife and being incredible supportive in laws), Rich Hobby (for a good laugh when I needed one). My Abuelo and Abluela (who taught me the value of hard work and determination).

Also priceless are the group of friends I have along the way. They were not only there when I needed to talk or hang out, but stayed with me even when I couldn't do the same for them. This includes but is not limited to Jon Howarth, Rishi Pande, Ross Goddard, Stacy Branham, Emilee Venn, Jane Jones, Justin Shirk, Laura Harty, Jason Lee, Shahtab Wahid, Tejinder Judge, Sarah Peck, Kelly Meredith, Manas Tungare, Ben Congleton, Pardha Pyla, Theresa Blanchard, the Garcoskis, and Ben Hanrahan.

I would like to thank all my coworkers who would put up with my sometimes crazy schedules, ask about my classes, and offer encouragement.

Table of Contents

Abstract	ii
Acknowledgements	iv
List of Figures.....	ix
List of Tables	xii
1 Introduction.....	1
1.1 Background.....	2
1.2 Twitter User Web Interface and Features.....	3
1.3 Types of Tweets.....	3
1.4 Friends and Followers.....	7
1.5 Twitter API	7
1.6 Motivation	8
2 Related Work.....	10
2.1 Measurement	10
2.2 Tweepie	11
2.2.1 Demographics.....	11
2.2.2 Timeline of tweeting/activity	13
2.3 Content & Description of Tweets	14
2.3.1 Number of Tweets.....	14
2.3.2 Content Analysis.....	15
2.3.3 URLs.....	16
2.4 Communities and Interaction.....	17
2.4.1 Conversations	17
2.4.2 Relationships	18
2.4.3 Influence & Centrality.....	20
2.4.4 Retweeting	21
2.4.5 Replies	21
2.5 Tools	22
2.5.1 3 rd spaces (Ambient displays)	22
2.6 Summary.....	23
3 Methodology	24

3.1	Data Collection	24
3.2	Survey	24
3.3	Quantitative Analysis.....	26
3.3.1	Timeline	26
3.3.2	Social Graph.....	26
3.3.3	Conversation.....	26
3.4	Conference Hashtags.....	26
3.5	Conference Setting	28
3.6	Tools Used to Analyze Data	29
4	Results	31
4.1	Grouping.....	31
4.2	Tweeple	32
4.2.1	Demographics.....	32
4.2.2	Timeline of tweetings/activity.....	34
4.2.3	Characteristics	35
4.2.4	Source	42
4.3	Content & Description of Tweets	43
4.3.1	Characterization of Message	43
4.3.2	Hashtags	53
4.3.3	URLs	61
4.3.4	Content Analysis	63
4.4	Communities and Interaction.....	66
4.4.1	Conversations	66
4.4.2	Mentions.....	80
4.4.3	Relationships	82
4.4.4	Influence & Centrality.....	82
4.4.5	Retweeting	97
4.4.6	Replies	101
5	Discussion	105
5.1	Tweeple	106
5.2	Activity.....	106

5.3	Community	108
5.4	Hashtags	110
6	Conclusion	111
6.1	Future Work	111
7	References.....	113
8	Appendix A: IRB	117
9	Appendix B: Survey.....	118
10	Appendix C: Code	133

List of Figures

Figure 1. A picture of the Twitter web interface with feed of tweets from those followed on left, and a panel on the left displaying recent tweet, people being followed, followers, lists, favorite tweets, trends, and advertisements.....	3
Figure 2 Tweet.....	4
Figure 3 Twitter Retweet Button.....	4
Figure 4 A Retweet	5
Figure 5 Twitter Mention	5
Figure 6 Twitter Reply Link.....	5
Figure 7 A Reply Tweet.....	6
Figure 8 Creating a Direct Message.....	6
Figure 9 A Direct Message.....	6
Figure 10 Tweet with Hashtag and URL	7
Figure 11 Representation of the different archives of tweets	13
Figure 12 Sample of Tweets Advertising Study.....	25
Figure 13 Business Card Advertising Study	25
Figure 14 Hyatt Regency Atlanta Conference Center.....	29
Figure 15 Vocation of Survey Participants	33
Figure 16 Histogram of Age of Participants.....	33
Figure 17 Number of tweets	35
Figure 18 Primary Use Of Twitter.....	35
Figure 19 Use of Direct Messaging.....	36
Figure 20 Use of Replying.....	36
Figure 21 Why do you use Twitter	37
Figure 22 Top three reason for using Twitter.....	38
Figure 23 Length of Twitter Use	38
Figure 24 Frequency of status update per day.....	39
Figure 25 Primary method of interaction with Twitter	39
Figure 26 Smart phone OS.....	40
Figure 27 Method of interaction on mobile device.....	40
Figure 28 Use of Hashtags	41
Figure 29 Recent use of hashtags.....	41
Figure 30 Creating list on Twitter	41
Figure 31 Check lists on Twitter	41
Figure 32 Follow lists on Twitter	42
Figure 33 Most than one account on Twitter.....	42
Figure 34 Number of Tweet by Source from participants tweets.....	43
Figure 35 Number of Tweet by Source from following GHC Hashtags	43
Figure 36 Types of Tweets.....	44

Figure 37 Group 1 - Experimenter Types of Tweets.....	45
Figure 38 Group 2 - Part-timer Types of Tweets.....	46
Figure 39 Group 3 - Active Types of Tweets.....	47
Figure 40 Group 4 - Immersed Types of Tweets.....	48
Figure 41 Content in Tweets.....	49
Figure 42 Group 1 - Experimenter Content.....	50
Figure 43 Group 2 - Part-timer Content.....	51
Figure 44 Group 3 - Active Content.....	52
Figure 45 Group 4 - Immersed Content.....	53
Figure 46 Number of hashtags.....	60
Figure 47 Percentages of hashtags.....	60
Figure 48 Number of URLs.....	62
Figure 49 Percentages of URLs.....	63
Figure 50 Word Cloud Text Depiction Week 1.....	64
Figure 51 Word Cloud Text Depiction Week 2.....	64
Figure 52 Word Cloud Text Depiction Week 3.....	65
Figure 53 Word Cloud Text Depiction Week 4.....	65
Figure 54 Word Cloud Text Depiction Week 5.....	66
Figure 55 Graph of all conversations.....	67
Figure 56 Piece of Conversation by Week and Group.....	68
Figure 57 Number of Conversation Pieces by Group.....	69
Figure 58 Number of Conversation Started by Group.....	69
Figure 59 Number of Conversations Started by Week.....	70
Figure 60 Group 1 - Experimenter Unique Tweets that started Conversation by Week.....	71
Figure 61 Group 2 - Part-timer Unique Tweets that started Conversation by Week.....	71
Figure 62 Group 3 - Active Unique Tweets that started Conversation by Week.....	72
Figure 63 Group 4 - Immersed Unique Tweets that started Conversation by Week.....	72
Figure 64 Conversations to Original Tweets by Group.....	73
Figure 65 Conversations to Original Tweets by Week.....	74
Figure 66 Group 1 - Experimenter Participation in Conversations.....	75
Figure 67 Group 2 - Part-timer Participation in Conversations.....	75
Figure 68 Group 3 - Active Participation in Conversations.....	76
Figure 69 Group 4 - Immersed Participation in Conversations.....	77
Figure 70 Top 25 User by Betweenness Centrality Based on Conversation Alone.....	79
Figure 71 Network Diagram of the Conversations.....	79
Figure 72 Number of Mentions.....	81
Figure 73 Percentages of Mentions.....	81
Figure 74 Friends and Followers Distribution.....	82
Figure 75 The two user left off the parallel coordinate visualization.....	86
Figure 76 Parallel Coordinates Visualization of Top 25 Users Over Time.....	87
Figure 77 Parallel Visualization by Group on the Sum of their betweenness centrality.....	88

Figure 78 Social Graph time slice at the date of 9/14	92
Figure 79 Social Graph time slice at the date 9/21	93
Figure 80 Social Graph time slice at the date 9/28	94
Figure 81 Social Graph time slice at the date 10/5	95
Figure 82 Social Graph time slice at the date 10/12	96
Figure 83 Social Graph time slice at the date 10/19	97
Figure 84 Number of Retweets	100
Figure 85 Percentages of Retweets.....	101
Figure 86 Number of Replies	103
Figure 87 Percentages of Replies	104

List of Tables

Table 1 Track/Session Hashtags	27
Table 2 Birds of a Feather (BOF) HashTags	27
Table 3 Group Distribution	32
Table 4 Top Hashtags by Week. The items in bold italics and have a grey background are hashtags listed on the Grace Hopper Conference websites. You can find this list on Table 1 and Table 2 of this paper.....	54
Table 5 Number of Tweets found with a Grace Hopper hashtag.....	58
Table 6 Percentage of all Tweets with a GHC hashtag.....	59
Table 7 Conversation Statistics.....	77
Table 8 Statistics of Social Graph Evolution Over Time.....	84
Table 9 Top 25 User Based on Betweenness Centrality Over Time Full Graph.....	84
Table 10 Top 25 User Based on Betweenness Centrality Over Time Friends Graph.....	88
Table 11 Top 25 User Based on Betweenness Centrality Over Time Followers Graph.....	90
Table 12 Unofficial Retweets from List	98
Table 13 Unofficial Retweet from GHC HashTags	99

1 Introduction

Twitter is a microblogging platform that allows users to post 140 character updates called tweets. This character limit and Twitter's straightforward architecture and interface make it appear simplistic. Although the small number of allowed characters per tweet limits users, users disseminate, share, and co-construct a new form of communication. Users address each other in the system as '@username,' where the username is replaced with the user's handle. Since its founding in 2006, Twitter has grown to be one of the largest websites on the internet. Two factors have set Twitter apart in its short but fast developmental history; (1) how users have adopted it, and (2) the unanticipated uses. At first it was considered as only a microblogging platform, but Twitter has quickly become a tool used by communities to connect events and topics. For example, Twitter has been used as a political tool (Khonsari, Nayeri et al. 2010) – as it was used in the 2008 presidential campaign (Greengard 2009)—a tool for gathering public opinions and survey results, a tool to physically and geographically organizing communities – as in flash mobs (Rott 2010), and much more.

The variety of uses is what makes Twitter so appealing yet challenging to study. For example, Twitter has been used as a 'software as a server' (SaaS) to build up other internet sites. SaaS, as described by Bennett, Layzell et al. (2000), is an application that is deployed and accessible over the internet. The provider of the software will license the application through a subscription at no charge which generates revenue through other means such as advertisements. TwitPic¹ and yFrog² are two of the hundreds of sites that use Twitter as a tie in service. These two example software facilitate Twitter users, or tweeples (people who tweet), share photos on Twitter. A user can post a picture on one of these sites and the service will automatically post the link to the photo and a description to their Twitter feed. These services, while initially not intended, are important for the microblogging platform. However, they have enveloped into the practices of online communities as a way to convey richer information. For example (Sutton, Palen et al. 2008), pictures were posted by inhabitants who lived through the Santa Barbara fires that spread in 2009. These pictures served as a dense information source that was capable of conveying the tragedy of destroyed homes. Twitter feeds were especially useful since news broadcasters and the Santa Barbara community did not have an accurate picture of the spreading flames until these photographs were released. With people forced to consider the life-changing decision of whether or not to evacuate their homes, the real-time Twitter pictures provided concrete data that then influenced decision-making and behavior.

¹ <http://twitpic.com/>

² <http://yfrog.com/>

Additionally, Twitter users use the site as a marketing channel to communicate with larger communities. For example, TechCrunch (@techcrunch³), a popular technology weblog, publishes links to all their posts on their official Twitter stream. This not only provides additional visibility for the content, but at the same time it helps in initiating conversations and comments from fellow Twitter users. Many other communities and media organizations use Twitter as a broadcasting platform, including CNN (@CNN⁴), *The New York Times* (@nytimes⁵), *The Washington Post* (@washingtonpost⁶), and Wired (@wired⁷), just to name a few. The significance of companies and organizations creating twitter channels is one of the larger indicators for the value that people are placing in online ad hoc communities, which online social services like Twitter facilitate.

The flexibility of Twitter, as discussed above, demonstrates that Twitter and other tools that tie into Twitter are valuable tools for supporting individuals and even individuals interacting with a larger community. My research focuses on how communities use Twitter when they gather in a physical space, namely a conference. In turn, I also study how these groups then co-produce online experiences that impact individuals in these communities.

I focus my research on how twittering behaviors change when large groups come together at events, such as during conferences. Conferences have specific attributes that impact my findings. First communities gather for conferences in part with a goal of networking and participation in community events. Secondly, conferences take place within a short controlled period of time, which gives me a very focused time period of data collection. Finally, people are not in their customary daily routines at conferences, which is important to understand if there is a change in behavior.

1.1 Background

Twitter.com is a popular microblogging service where users post status updates, which are often referred to as “tweets.” While Twitter does have a native web interface (see Figure 1), a good percentage of the users use third party clients and applications to post updates. The simplest description of Twitter is that it provides a timeline of 140 or less character status updates from people who the user is following. In this section, I elaborate on Twitter and provide a description of the conventions and types of tweets, and a description of the basic Twitter user web interface.

³ <http://techcrunch.com/> & <https://twitter.com/#!/TechCrunch>

⁴ <http://www.cnn.com/> & <https://twitter.com/#!/cnn>

⁵ <http://www.nytimes.com/> & <https://twitter.com/#!/nytimes>

⁶ <http://www.washingtonpost.com/> & <https://twitter.com/#!/washingtonpost>

⁷ <http://www.wired.com/> & <https://twitter.com/#!/WIRED>

1.2 Twitter User Web Interface and Features

Given the focus of this thesis on a particular microblogging service, Twitter, this section describes the different user interface features currently available for tweeters or “tweeple.”

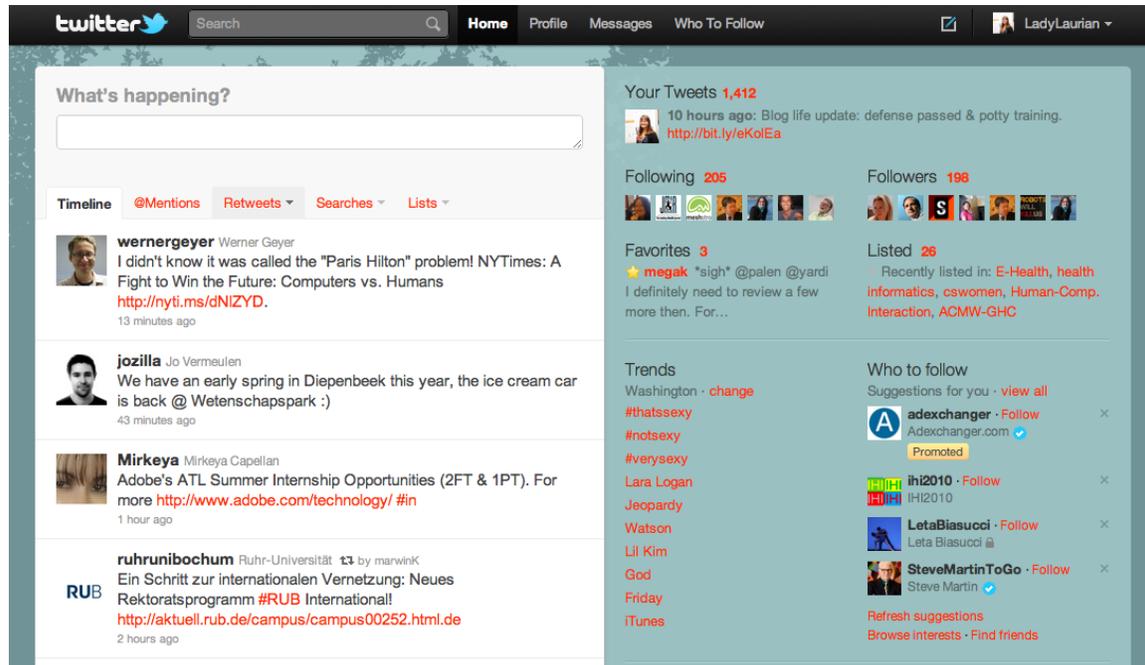


Figure 1. A picture of the Twitter web interface with feed of tweets from those followed on left, and a panel on the left displaying recent tweet, people being followed, followers, lists, favorite tweets, trends, and advertisements

1.3 Types of Tweets

Twitter is used in many different ways, not merely as a status update service. People use Twitter to share recommendations, make new friends, track events, and share pictures among other uses. The variety of uses is what makes Twitter appealing and also challenging to study. The tweets, or updates, posted on Twitter can be categorized into several different types.

The following are the different types of tweets that one could encounter on a typical Twitter timeline. These are important because they differentiate and convey the multitude of possible uses of a single tweet. These are the types that are supported by the Twitter application on both the website and through the API (see section 1.5 Twitter API).

1. Figure 2 below shows a Regular Twitter message. It shows the limited area where the message appears and a few other things that the Twitter web client provides to the users, like the count of characters used so far (27 in this example) and the location where the tweet originated (San Francisco, CA in this example).

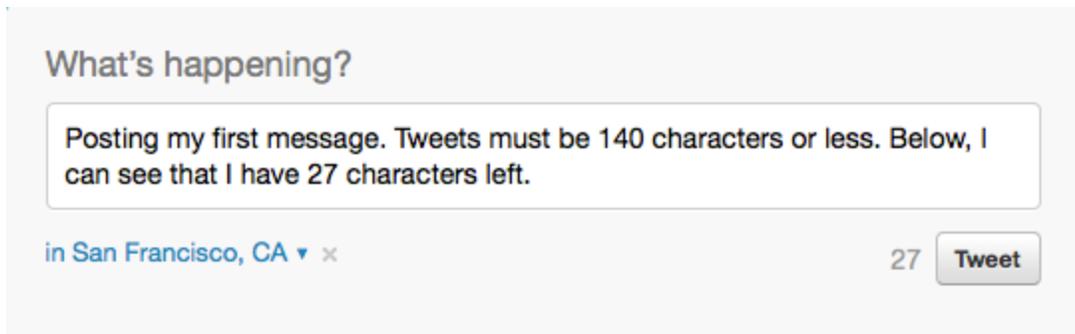


Figure 2 Tweet⁸

2. Retweets – Messages originally posted by another user and then reposted (typically denoted by “RT”). Retweets only became officially supported on Twitter in November 2009⁹. Before Twitter supported retweets, researchers (boyd, Golder et al. 2010) found many different retweet styles. They found the following retweet syntax ‘RT: @’, ‘retweeting @’, ‘retweet @’, ‘(via @)’, ‘RT(via @)’, ‘thx @’, ‘HT @’, ‘r @’, and ‘ @’. A user now must create retweets via the link displayed in their timeline in order for a retweet to be official. An example of a retweet link is in Figure 3 Twitter Retweet Button, which then would like look Figure 4 in the timeline.



Figure 3 Twitter Retweet Button¹⁰

⁸ <http://support.twitter.com/groups/31-twitter-basics/topics/109-tweets-messages/articles/15367-how-to-post-a-tweet>

⁹ <http://blog.twitter.com/2009/11/retweet-limited-rollout.html>

¹⁰ <http://support.twitter.com/forums/10711/entries/77606>



Figure 4 A Retweet¹¹

3. Mentions – As seen in Figure 5 below, mentions allow the user to refer to another Twitter user. Denoted by @ symbol and the Twitter handle located anywhere in the message.



Figure 5 Twitter Mention¹²

4. Replies –A subset of mentions, which consist of messages that are in reply to another message. The end user would have to click on the reply link for it to be valid reply. In order to reply to a tweet a user must click on the Reply button as show in Figure 6, which would then show in the timeline as Figure 7 shows.

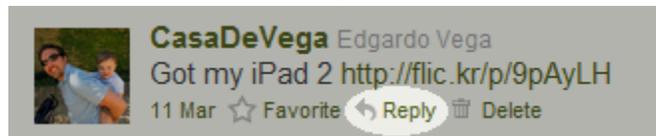


Figure 6 Twitter Reply Link¹³

Replies look very similar to mentions. The difference is the subtle but visible speech bubble in the top right of the tweet as seen in Figure 7. The Twitter API also makes a distinction between these two.

¹¹ <http://support.twitter.com/forums/10711/entries/77606>

¹² <http://support.twitter.com/groups/31-twitter-basics/topics/109-tweets-messages/articles/14023-what-are-replies-and-mentions>

¹³ <http://support.twitter.com/groups/31-twitter-basics/topics/109-tweets-messages/articles/14023-what-are-replies-and-mentions>

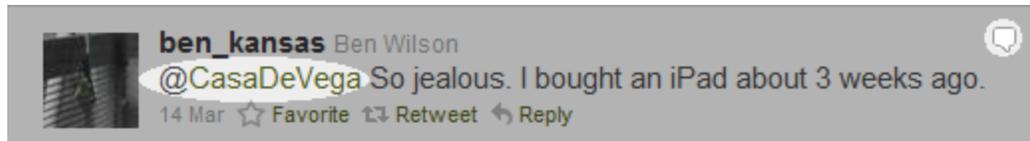


Figure 7 A Reply Tweet

5. Direct Messaging – A feature used for exchanging private messages between two users, commonly referred to as DM. Figure 8 demonstrates a user creating a direct message, which would then show up in a user’s message screen as shown in Figure 9 . Note that messages are completely separate from the timeline in order to prevent confusion between public and private messages. There is one restriction, however: in order to direct message another user you must follow that user.

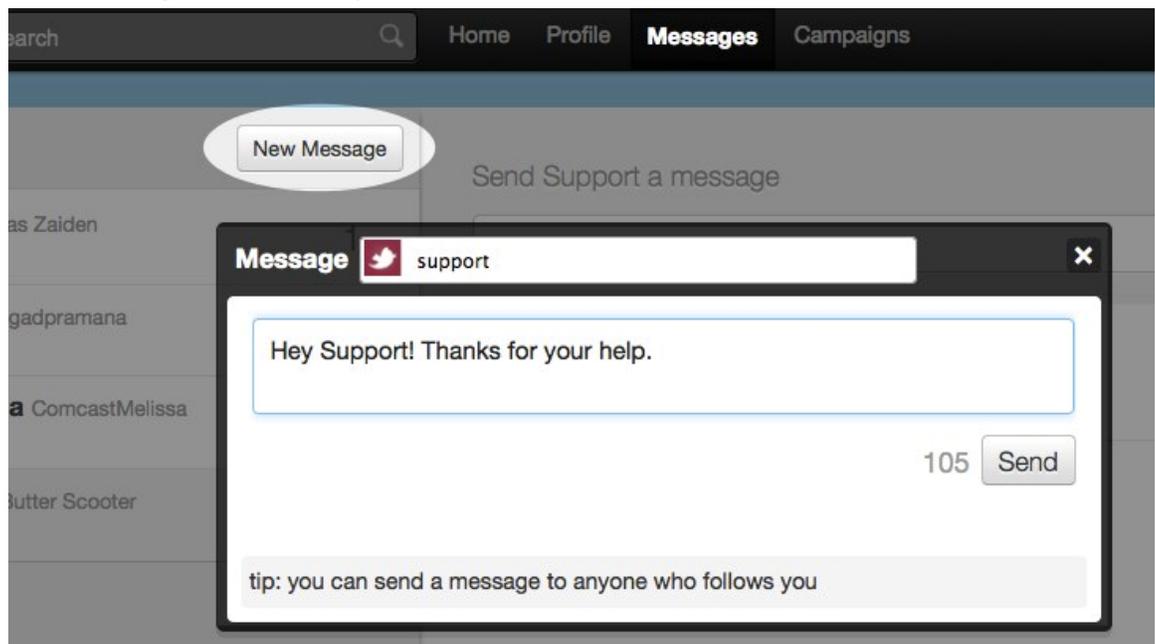


Figure 8 Creating a Direct Message¹⁴



Figure 9 A Direct Message

The following are different features a tweet could have. These features are supported by the Twitter application on both the website and through the API.

¹⁴ <http://support.twitter.com/groups/31-twitter-basics/topics/109-tweets-messages/articles/14606-what-is-a-direct-message-dm>

1. Hashtags – A convention of prefixing a # symbol at the beginning of a keyword, making it easy to filter and search tweets.
2. URLs – are the address to resources on the web. Most URLs on Twitter use a URL shortening service to save space on their tweets, such as bit.ly or tinyurl.



Figure 10 Tweet with Hashtag and URL¹⁵

Some of these features, such as retweets and hashtags (see Figure 10), were started by the user community who came up with conventions. Hashtags are an easy way for users to find other tweets on a particular topic. At first, the interface did not support linking hashtags to the search, but Twitter started linking hashtags in the user interface in July 2009¹⁶.

1.4 Friends and Followers

Twitter’s model for friending is a one way motion. This means a user (A) can follow another user (X) without the permission of that user (X) or user (X) following the first user (A) back. Friending or following someone means that: you are subscribing to the user’s tweets as a follower, the user’s updates will show in your timeline, and that person can send you direct messages, which are otherwise restricted. Twitter officially uses the term “following,” but for the purposes of this study we will say friending in order to keep the direction of the relationship clear.

Followers are people who will get your tweets in their timeline, are listed as a follower in your account, and have the ability to direct message you. Many celebrities have a large amount of followers, but they only friend a few individuals. Ashton Kutcher (@aplusk¹⁷) has 623 friends and 6,548,068 followers, Conan O’Brien (@ConanOBrien¹⁸) has 1 friend and 2,879,366 followers and Charlie Sheen (@charliesheen¹⁹) has 41 friends and 3,563,023 followers as of April 2, 2011.

1.5 Twitter API

API stands for Application Programming Interface, which is an interface between different software programs. Twitter’s API uses HTTP, the same protocol that browsers use to interact with the internet.

¹⁵ <http://support.twitter.com/groups/31-twitter-basics/topics/109-tweets-messages/articles/49309-what-are-hashtags-symbols>

¹⁶ <http://techcrunch.com/2009/07/02/twitter-makes-hashtags-more-useful/>

¹⁷ <https://twitter.com/#!/APLUSK>

¹⁸ <https://twitter.com/#!/conanobrien>

¹⁹ <https://twitter.com/#!/charliesheen>

Twitter currently has three APIs which developers can use: REST, Streaming, and Search API's. The REST API allows developers to access core Twitter data, such as blocks, profile information, timeline, etc.. The Search API allows the developer to access Twitter Search and trend data. Lastly, the Streaming API provides real-time high-value access to Twitter.

The REST and Search API are rate limited, which means that a developer can only make a limited number of calls to the API per hour. By default, an anonymous call can make 150 requests per hour and an authenticated call can make 350 requests per hour to the REST API. The Search API is still rate limited, but that limit is not public.

The Streaming API allows more access to Twitter data than the REST and Search API. A developer can use the filtered stream and, by default, track 400 keywords and follow 5000 user IDs. There are other roles available that offer higher values, but Twitter grants these three. There is also a method of getting a sample of all tweets from Twitter, but the default this is 1% of all tweets coming through the system. Twitter can grant other roles and methods to a developer that would allow access to all tweets for special purposes only.

1.6 Motivation

Conferences are assemblies of individuals who have common interests or goals. A conference provides a physical location for these people to meet and discuss these interests or goals. Interestingly, social networks augment the physical conference setting. The motivating question is: how do people change their behavior in these virtual spaces during conferences? During the remainder of the time at the conference these people can usually only connect using virtual methods such as social networks. These networks provide virtual places to discuss and disseminate information. By only looking into the change in behavior during the conference we can relate our results to other settings. The number of individuals who tweet at conferences is not a factor in this research. A user who fits into a category will follow similar habits before the next conference in which they participate.

For some individuals, OSNs are critical components to doing their job on a day-to-day basis, such as individuals in recruitment and public relations roles. Public relations professionals utilize social networking sites to connect with their customers, provide advice, answer questions, or address complaints. Others use OSNs to keep and expand their business network, which is the main purpose of the site, LinkedIn. Entertainment is another use of OSNs, with social games such as Cityville, which have millions of users per month. Cityville has reached 100 million players on Facebook in one single day²⁰.

One of the main purposes of a conference is to network with other individuals with similar interests. Networking has been one of the biggest topics when related to employment. Twitter has brought those networks into the web as another place where individuals can network and

²⁰ <http://blog.games.com/2011/01/24/has-cityville-reached-its-user-peak-falling-numbers-say-yes/>

discuss ideas. Twitter allows users to keep relationships they previously had, such as email or phone or to create new ones.

But what happens when you bring the two together, the virtual and the physical? How does having these users in the same place affect their Twitter usage? These are just two of the many questions driving the researcher. In order to find out we looked at user activity, tweet types, influence and centrality, and collaboration and conversation. We collected these measurements in order to understand people's participation, how they used Twitter, and how the community was affected.

2 Related Work

Twitter is an information repository and provides an ample data pool for us to search for varied information. It acts as an information sharing medium (Kwak, Lee et al. 2010), a medium for social search (Brooks 2010), a phatic communication medium (Makice 2009), a personal diary, and much more. Twitter search (Golovchinsky and Efron 2010) is turning out to be a valuable tool for finding information, with one of the biggest advantages being real-time results. The search stream brings in content from millions of users, and, unlike a normal Web search engine, there is no issue of indexing latency. This enables the user to get opinions on any given subject, not only from their personal network of friends, but also from public users. With location-based filtering, the additional feature brought greater relevance and context to the tweets.

In this section, I provide an overview of research on Twitter and microblogging as a social phenomenon. This includes describing the types of measurements that have been made on tweeters, using Twitter as an information-sharing tool, the archiving of tweets, Twitter as a communication tool, using Twitter as an awareness mechanism with ambient displays, the role of relationships embodied in tweets, and understanding the content of tweets. Given the relatively youthful research of this field, prior work is limited.

2.1 Measurement

There are numerous ways that Twitter is measured. However, its measurement is highly dependent on how a researcher defines it. Krishnamurthy, whose work serves as seminal research on how to measure social networks, defines an Online Social Network (OSN) as “a network consisting of real users who communicate with each other in an online setting in diverse ways” (Krishnamurthy 2009). OSNs became popular with the rise Web 2.0 (Cormode and Krishnamurthy 2008). Web 2.0 facilitated user-centered design through interface customization, crowdsourcing through mass contribution to online websites, the adoption of ‘web as platform’ through the use of online services that mimic desktop applications, collaboration through mass adoption, power decentralization through the democratization of services and content, and dynamic content with users providing real-time updates (Sharma 2008). These features facilitated the emergence of OSNs like Twitter that created a whole new space and phenomenon that impacted online and offline interaction.

Just as in the physical world, Krishnamurthy found that relationships in OSNs can be fragile or solid and as such there are different means to communicate in both worlds depending on the relationships. Krishnamurthy (2009) lists properties that should be studied in OSNs and discusses the difficulties in researching them. He outlines four axes that research should be concerned about: basic characterization, dynamic interactions, network-level issues, and social issues. I discuss each of these below.

Network-level and dynamic interaction issues are ones that require access to the user's computer and need to have some kind of collection software between the user and the OSN in order to measure properly (Krishnamurthy 2009). Network-level measurement gathers computer communication protocols, bytes to be transferred based on user actions, and overlay networks. Dynamic interaction measurements gather session durations?, external application usages, and sub sessions. This ability is outside the purview of my thesis, but can be valuable for future data collection efforts.

Krishnamurthy (2009) also discusses basic characterization and measurement challenges in relation to OSNs. Basic characterizations include the number of users, friend counts, personal settings, sub-communities, communication methods, content, frequency of updates, cultural norms, languages, social graph diameter, edge to node ratio SCC, backward links as aspects of OSN's behavior that should be studied. He states that crawling a user's social network is hard in OSNs, because they are closed networks, APIs are rate limited, and there are no general-purpose tools. Also, current measurements are of small snapshots and sub-communities that require membership. Finally, cultural effects might have more of an impact on the social network; this, however, is difficult to characterize.

Twitter does have features that address Krishnamurthy's concerns on studying OSNs. For example, Twitter supports the social issues of anonymity and privacy. Twitter allows its users to stipulate whether their accounts are publicly or privately viewable, but a large majority of user accounts are public. This allows a researcher to examine Twitter accounts with the knowledge that he or she may need to receive permission first from users with private accounts or otherwise to get a list of usernames for followers and users through individual contact. If the account is public, a researcher is able to see the user's social graph and is able to take a snapshot of the graph whenever needed, thus allowing some measurements of social networks, as in the research of Knoke and Yang (2008). If an account is public, most of Krishnamurthy's basic characterizations are available for a researcher to retrieve, such as I have done in this study.

2.2 Tweepie

2.2.1 Demographics

There are three different types of Twitter users as described by Java et al., (Java, Song et al. 2007) namely the information sources, friends, and information seekers. The "information sources" might be automated newsfeeds and tend to have a large follower base, while "friends" is a more general category that includes all common users. "Information seekers" are those users who generally do not post updates but tend to keep track of updates posted by other users. While some of the types were present natively, designers added a good number of these features to the service, due to popularity among users. While these categories are useful, researchers created them with all the users of Twitter in mind and not for distinguishing user behaviors at a conference.

Rejaie, Torkjazi et al. (2010) built upon Krishnamurthy's work and created a measurement technique to infer the growth of Online Social Networks (OSNs) in a fine grained manner. They were able to do this by determining the number format of the user's IDs and the OSN's allocation strategy. From this evidence, they were able to determine the growth in total and the number of valid accounts remaining. They then were able to separate the valid accounts into active users of the system. Based on this measurement scheme, the researchers estimated that Twitter has 144 million accounts. Of these accounts 76.7% are valid with about 37.9% being moderately active and 18.5% being very active. They found that 45% of users have not sent any tweets and are therefore non-active. Of the accounts that were sampled, 23.3% users had deleted them, 70.7% were public, and 6% were private. Of the deleted accounts, 25% were removed by Twitter due to suspicious behavior.

Huberman, Romero et al. (2009) studied 309,740 users and retrieved user's tweets to try and understand different tweeting behaviors (and in the process categorize microbloggers). In this study, each user posted on average 255 tweets, had 85 followers, and followed 80 others. The paper did not indicate the timeframe in which they retrieved the data. Only 211,024 users posted at least twice during the collection timeframe. These users were 'active' in the study and were active for 206 days on average. In this study, the researchers discovered that there is a network within a network. The first network is the full Twitter graph of friends and followers and the second lighter network is composed of 'actual friends.' The determination of 'actual friend' is when a user directly mentions another user at least twice in tweets. These groups of people were the most influential in the user's network when defined by interaction on a microblogging platform. What is interesting and relevant for the study presented in this paper is that users with many 'actual friends' tended to update Twitter more often than those users with less 'actual friends.' Also, users with many friends and followers posted less than those with fewer friends and followers. How tweeting behavior changes when the number of friend versus 'actual friend' is an interesting question.

In other research on communication in Twitter, researchers used the communication behavior to categorize different types of users. For example, Liere derived three patterns to understand information flow in Twitter (Liere 2010). The first pattern, the baseline pattern, is the 'random pattern,' because there is uniformed distribution in reference to whom the information is being communicated. The second pattern is the 'local pattern' where the user follows those that are geographically close to them, meaning that the user follows people who they are friends with in the 'off-line' world. The final pattern is the 'information broker pattern' as is based on individuals who have shared interests, but who are not necessarily friends in the 'off-line' world. An information broker is a user who acts as a filter for other users. They consume the large amounts of data and filter it down to other users who might not have the time to do so. From the results of the study, the information broker pattern is the best description on how information flows through Twitter. These groups are important for my research because people may be information brokers prior to the event studied in my dissertation. However, afterwards

they may follow a more ‘local pattern’ of communication behavior. Thus, events such as the one I am studying may have a large impact on online behavior rather than offline behavior being a large indicator of tweeter type.

2.2.2 Timeline of tweeting/activity

Each individual Twitter user sees updates from the people whom he or she follows on their personal timeline. The number of updates depends on the number of people the user follows; if it is on the higher side, the number of tweets that appear on the personal timeline could be overwhelming and could also cause information overload. For an individual who has specific tasks and information needs, the process of re-finding can and will be cumbersome given the apparent (or lack thereof) affordances present in the interface. To manage this issue the Web interface of Twitter not only archives personal tweets, but also archives tweets from friends in the network. In April 2010, the U.S. Library of Congress additionally acquired the entire public Twitter archive for preservation and research purposes (2010), which demonstrates the value being placed on this form of information.

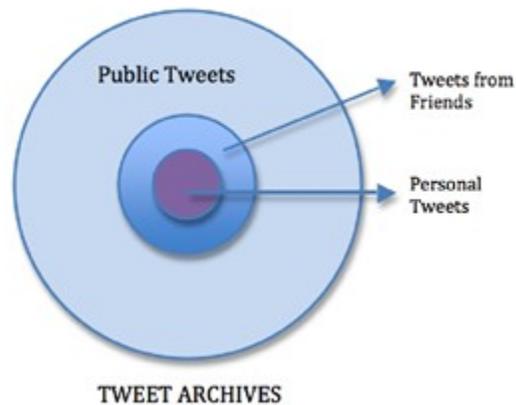


Figure 11 Representation of the different archives of tweets

An individual can view the Twitter archive from the three different perspectives discussed in this section, see Figure 11 Representation of the different archives of tweets. The first is the larger set that contains billions of tweets from all the users of the service. The next level is the archive of updates from friends in the network. The final one is the personal archive, or, in other words, the tweets published by the individual user over a period of time. The amount of personal information is local to personal tweets. The relationship in spectrum between personal and public tweets is unique and leads to questions about tweet content such as: what consists of relevant information, what is the value of time in accessing information in relation to a topic, and what consists as the user’s social network that would be valuable for a user to access.

There are a few drawbacks in the methods of accessing public tweet archives, or, in other words, re-finding one's own tweets. The current Twitter search model only allows the average Twitter user to access data from tweets that were within the past seven days. Twitter also limited the number a tweets a user can have to 3000 tweets. Creating new tweets will cause Twitter to discards older tweets. Even though the Twitter profile page has all the past 3000 tweets, there is no organizational methodology; it is tedious to look into past tweets. The Twitter API provides access to a greater number of tweets. There are also external tools (e.g., Tweetake²¹, Tweetscan²², BackupMyTweets²³, TweetBackup²⁴, etc.) and plug-ins that will help to build a browseable and searchable archive of tweets. Also, Twitter's direct messaging capability stores private conversations between users, which are accessible only to the two parties involved in the conversation.

The archiving of tweets is important in this design space primarily due to the difficulties previously discussed, such as the limits in search and tweet numbers. Users are encouraged to archive their own digital online profiles (Perez 2009) through online tools like TweetBackup²⁵, and larger companies scrape the Twitter timeline by storing their own copies of the data on external servers (Krazit 2010). The ability to get access to tweets is inherently difficult because the system values impermanence.

2.3 Content & Description of Tweets

2.3.1 Number of Tweets

In 2010, the users of Twitter generated over 50 million tweets in a single day, which accounted for over 600 tweets per second (Weil 2011). The sheer amount of data created by users of Twitter demonstrates the raw magnitude of its power as an information-sharing tool. One large task, which might lie in the hands of the user when faced with this tool, is to find the information that he or she is looking for amidst thousands of tweets in the public timeline. But without thinking about the micro-level actions, undoubtedly Twitter is a collective source of intelligence that helps users to obtain opinions, ideas, and other relevant information with ease (Cheong and Lee 2009).

Lee and Sumiya (2010) find that it is possible to detect unusual events that are happening based on Twitter. First, they find the patterns of users in a geographic region, and then they compare that as a baseline to find unusual events. This method can be important for using the measurements of Twitter behavior to detect important events that the news media has yet to broadcast.

²¹ <http://tweetake.com/>

²² <http://tweetscan.com/>

²³ <http://backupmytweets.com/>

²⁴ <http://tweetbackup.com/>

²⁵ TweetBackUp (<http://tweetbackup.com>) allows users to sign up and store their own profile on an additional server on a daily basis.

2.3.2 Content Analysis

A large portion of research on Twitter focuses on classifying behaviors based on the numerical data. However, as will be discussed in this section, there is also some research that examines the content of the tweets.

Westman and Freund analyzed tweets using genre analysis (Westman and Freund 2010). They started with the work of Yoshioka et al. (Yoshioka, Herman et al. 2001) which used a granular genre classification system. The system presented in their paper was based on the elements of “who, what, where, when, why, and how.” For Twitter, Westman and Freund used the following to code each tweets: information sharing, conversation, information seeking for why or purpose, personal information or commentary, news and public information, business and promotion, issues related to Twitter for what or content, popular or experienced for who or participants, daytime (8AM to 8PM) for when, Twitter.com for where, and addressing other users, linking, hashtags, retweets, or emoticons for how. Based on the analysis of the “who, what, where, when, why, and how,” they identified five comment genres: personal updates, direct dialogue, real-time sharing, business broadcasting, and information seeking. They discovered that the numbers in their results for frequency of events were higher than in previous work. They assert that there appears to be a trend that communication practices are changing. Other findings were that most of the messages in this study fell into the information sharing category, followed by conversation. Personal content was the most common followed by news and business tweets. Additionally, public information sharers have a smaller friends-to-follower ratio than information sharers. These finding extend the work of Naaman, Boase et al. (2010).

Michelson and Macskassy (2010) discuss that previous work in analyzing Twitter in real time focuses on topics such breaking news, natural disasters, etc.. Instead, what they propose is to understand what topics a user would be interested in based on the text contained in the user’s timeline. The authors suggest that a way to do this is to use machine-learning techniques due to the large number of tweets and user profile types. In their prototype, called Twopics, the researchers attempt to do this. First, the system takes all the capital words and categorizes them as nouns. Then the system uses Wikipedia and retrieves a list of candidate topics, and further siphons the candidates by disambiguating them by using the rest of the words in the tweets. Twopics then takes the Wikipedia categorization tree for each word and combines them to create a tree. Other methods such as bag-of-words or latent Dirichlet allocation are suggested by the authors, but the authors state that both these methods have drawbacks and were able to overcome some of these issues by using Wikipedia as a collective intelligence knowledge base. This research suggests that it can support analyzing the content of a tweet into topic areas, even if it is not at the level of detail or accuracy that previous literature suggested was necessary to be useful.

As an extension on the work of Java, Song et al. (2007), Naaman, Boase et al. (2010) address gaps in previous work by analyzing Twitter both qualitatively and quantitatively. The

researchers created a list of ten content categories to code a tweet in. They are: Information Sharing, Self Promotion, Opinions/complaints, Statements and Random Through, Me Now, Question to Followers, Presence Maintenance, Anecdote (me), Anecdote (others). They used a grounded approach to arrive at these categories and code them. They conclude that the majority of users focus on the “self” but that the more popular and active users focus on information sharing. They coin the term “Meformers” as a way to describe the first group of users. In their analysis, 80% of users were “Meformers” and the remaining 20% were Informers. This work is a useful analysis of different types of users, but more importantly it provides a valuable method for analyzing the content of tweets.

Qiu, Feng et al. (2010) explore Twitter in identifying if it is possible to determine cell phone network issues. In order to have a comparison, the research used the ticket-based customer support system where people make calls to the service center against the data that was being collecting from Twitter. The researchers found that the issues on Twitter were in fact complimentary to the customer care calls. Users report short term, or minor, issues on Twitter. However, the customer care reports did not record these same issues. Finally Twitter was faster to report problems when compared to customers calling in by a margin of ten minutes. Twitter was better at identifying the timing of events (90.1%) in comparison to the customer care cases (23.7%). This paper suggests that the content of tweets has varying and useful applications outside of those originally conceived.

Oulasvirta et al. (Oulasvirta, Lehtonen et al. 2010) built on the work of McNely to further discuss the communication genre in microblogging and how it makes ordinary things visible to others. The study involved content analysis and qualitative analysis using a dataset that contained 400,000 posts. They focused on categorizing responses for mundane and non-mundane responses and emphasized that microblogs are for individuality and serve as a medium for letting the public know about daily happenings in the user’s life. Conversely, microblogs are being used in many different scenarios in the past two years, particularly in ways in which it was not originally designed to be used, such as campaigning, customer interaction, and education.

2.3.3 URLS

Galuba, Aberer et al. (2010) tracked 15 million URLs over a 300 hour period from Twitter. They allowed 300 hours for system training using the first 150 hours. And then, after training, the system could be used to analyze the second 150 hours. This study found that a system could roughly predict how a URL propagated across tools like retweeting in Twitter. The model predicted 55% of the URL mentions with at most 15% of the URLs being false positives. They found that predicting who will propagate a tweet further than “two hops” (meaning across two networks) is currently computationally impossible. For example, they found that the diffusion of a URL throughout the network is log-normal distribution. Additionally, the graph consisting of only follow data for Twitter is an enormous connected component and has a mean shortest path of 3.61. However, the largest point made by the authors is that URLs are useful as a way to contextualize online information to communities who participate in microblogging

2.4 Communities and Interaction

Reinhardt, Ebner et al. (2009) conducted a study of conferences using online surveys. Based on data from five separate conferences, workshops, and meetings, the researchers received 41 responses. They concluded that Twitter is another method of discussing presentation materials and exchanging additional information. It was found that communicating and sharing resources are the main reasons the respondents used Twitter. At conferences users would attend a second session virtually that they would have otherwise missed completely. Most of the content is either text or links to sites. Limitations with this paper include the limited number of responses and conference types studied. In my study, I had 78 completed surveys at one conference. While a valuable contribution to studying co-located communities, this article only studies Twitter and conferences through surveys and does not measure data coming directly from Twitter, as I present in this thesis.

Barnes and Böhringer (2009) studied why Twitter has grown in size, in spite of its numerous criticisms, such as repetitive down time. In order to study this, they created a survey and used the Twitter API to collect data on their participants. They collected 131 survey responses and then used the API collected user tweets for a two-week period previous to the survey. With this information they were able to measure seven constructs: perceived usefulness, confirmation, satisfaction, continuance intention, habit, comprehensiveness of users, frequency of past behavior, and social network size. The results indicate that habit, satisfaction, and perceived usefulness strongly influence continuance behavior. Social network size influenced perceived critical mass, comprehensiveness of usages, and frequency of past behavior. These three constructs impacted habit along with satisfaction. Confirmation and perceived usefulness influenced satisfaction. The researchers state that all of these factors demonstrate that Twitter use is actually quite complex.

Johnson and Yang (2009) discovered two factors for both gratification sought and gratification obtained. These were social motives and information motives. They surveyed 242 users of Twitter and asked questions about Twitter users, gratifications sought, and gratifications obtained. The researchers asked 15 questions on each of the last two categories and measured satisfaction by the difference in the obtained pairs of gratification sought and gratification obtained. While individuals use Twitter for many things, users were most satisfied with Twitter in the way it helps the users interconnect, communicate with others, share thoughts in a public space, and track what other users are talking about. Finally, the researcher found that social gratification did not affect Twitter use, but that information gratification had large effects on it.

2.4.1 Conversations

Twitter serves as a conversation tool and allows users to make new connections; it is a social tool, which allows and facilitates social interactions. In this section, I highlight research projects that are valuable when examining conversations and connections through Twitter.

In the first example, Honeycutt and Herring (2009) found through their analysis that the collaboration and conversational aspects of Twitter are important when focusing on user replies and the coherence of exchanges. In particular, the @ sign used in the tweets facilitated conversations to a great extent. This allowed the users to respond to topics directly with one another instead of users addressing comments to their general social network.

Participation in conversations and communication does not have to be dependent on the model of a user contributing and an individual participating. Research has engaged in how ambient intimacy can leverage Twitter as a phatic communication medium (Makice 2009). In this context, phatic means that the expression does not convey information, but only performs a social task. Ambient intimacy is being able to keep in touch with individuals one normally would not be able to keep in touch with due to geography and time. Kevin Makice studied this aspect by examining the previous work on “TwitterSpaces” (Hazlewood, Ryan et al. 2008), which are large displays of streaming tweets from a closely-knit community. TwitterSpaces provided a glimpse of the daily happenings of the members and creates a community feeling. This particular finding is important for my research given that a similar setting is very common in many of the conferences. Users at these conferences attend sessions where organizers stream their tweets live on a large screen. This facilitates a persistent and mobilizing backchannel environment. Experiences with back channeling, Twitter, and conferences are conveyed systematically in (boyd 2010) and see 2.5.1 3rd spaces (Ambient displays)

In particular, McNely (2009) discusses the need for a framework for understanding the role of backchannel environments in collaborative meaning-making capacity. Part of the reason for Twitter’s success is due to factors such as the lower barriers to participation, SMS functionality, and real-time activity, which all enhance the backchannel. The co-construction of information allows people to feel as though they are collaborating for a low cost, thereby increasing participation.

2.4.2 Relationships

Early research into internet communities stated that relationships facilitated through digital means were limited and less valuable than offline ones such as phone calls or face-to-face meetings for developing and maintaining relationships (Cummings, Butler et al. 2002). Later, weblogs evolved and laid the foundation for how relationships and communication took on new meanings in an online context: quick and easy publishing, permalinks, comments, and trackbacks were all features that evolved from blogs. These features allowed for the connection of weblogs, writers, and readers to establish a community (Blood 2004). The researchers have observed that strong virtual relationships were key indicators that the user would continue to be active in the social network platform. Therefore, strong social interaction is significant to user interaction and retention in these systems (Lento, Welser et al. 2006). Both the research of Blood and Lento (Blood 2004; Lento, Welser et al. 2006) prove that relationships are important in online and physical worlds and in order to achieve a strong social network both are necessary

– especially with social networking websites increasing in adoption across the world (e.g., Facebook) states that they have 500 million active users as of 2011).

Kumar, Novak et al. (2006) define the idea of a time graph for social network analysis that serves as a visualization of relationships. The work explains the use of time graphs as a way to study changing graphs, which depict changing relationships over time. The nodes and edges of the graph are only created in a fixed moment in time and therefore can only exist in certain intervals, however these graphs can serve as a useful representation of the social graph.

Building on both their previous work (Leskovec, Kleinberg et al. 2005) and that of Kumar, Novak et al. (2006), Leskovec, Kleinberg et al. (2007) focus on how the graph of social relationships changes over time. Previously the assumption was that a single static graph could be analyzed and that the average node degree would be constant and that the diameter was a gradually growing function of the network size. After their analysis, the researchers concluded that the assumption that the degree remains constant was false. They discovered that graphs have out-degrees that grow over time and call this the ‘densification power law.’ Also the diameter of a graph may actually decrease gradually as the network grows over time. This is the graph power-law degree distribution. To model these phenomenon, the researchers created the “Forest Fire Model” that allows forecasting in computer and social networks and realistic graph generators. To incorporate this research in my own thesis, I have created a dynamic visualization for understanding how the social relationships change over time. Section 4.4.4 Influence & Centrality contains snapshots of the visualization created.

Researchers evaluate relationships in social networks by densification. Densification follows the same unexpected pattern in a rapidly growing, declining, and slow but steady growth network (Kumar, Novak et al. 2006). The researchers classified members into three groups; singletons, the giant components, and the middle region. Singletons are degree-zero nodes, which are users who join the service but have not made any connections. Giant components represent the large group of people who are connected. The middle region includes all others who are users who connect with small unconnected communities. Often, half the mass of a network is outside the giant component, which is composed of stars. The researchers created a model with three types of users; inviters, linkers, passive users. Linkers are full participants in the OSN. Inviters are moving from an offline network into an online social one and are actively inviting their friends. Lastly, passive users join the network out of curiosity but never engage in a meaningful way. Using those users allows researchers to generate a simulation that they can use to understand how social relationships change. While these classifications of users are helpful for understanding general tweeting behavior and relationships, they are not specific enough for the behaviors that I observed in my study.

Twitter serves as a medium that enables sharing of information with likeminded people. For example, Kwak et al. (Kwak, Lee et al. 2010) studied the influence of Twitter as an information sharing medium by crawling the entire Twitter user base and obtaining a dataset containing 41.7

million user profiles and 106 million tweets, among other information in the first comprehensive quantitative study of the entire Twittersphere. After ranking Twitter users and analyzing the trending topics and retweets, they have observed that the follower-following distribution in Twitter has a deviation from that of a human social network. In particular, they found that non-power-law follower distribution, a short effective diameter, and low reciprocity apparent in relationships indicate that other factors are influencing tweeting behavior in information seeking. This research demonstrates that when considering all users on Twitter as part of a community, the large majority is trending news topics and a few key tweets are shared to 1000s on the network through re-tweeting behavior. This is important to my study given the consideration on factors that permeate the community behavior, however, my study instead focuses on information sharing within a smaller sub-section of people congregated at a conference.

The researchers Mark and Shneiderman worked with NodeXL, which can be used across many social networking applications for visualizing and measuring online networks (Smith, Shneiderman et al. 2009; Hansen, Shneiderman et al. 2010). They have found that Social Network interactions reveals organizational structure (Smith, Hansen et al. 2009), social context (Smith, Barash et al. 2008), and understanding social connections (Swift 2011).

In general, prior research has found that by crawling social networks, researchers can generate the statistics. These statistics are in relation to the number and type of connections that people have along with the behavior of the user such as the number of updates the user provides.

Researchers Bakshy, Hofman et al. looked at 1.6 million Twitter users over a two month period and found 74 million diffusion events (Bakshy, Hofman et al. 2011). This finding is consistent with previous work (Cha, Haddadi et al. 2010; Kwak, Lee et al. 2010; Weng, Lim et al. 2010) which showed that the distribution of followers and friends was lopsided. Using Amazon's Mechanical Turk, the researchers had individuals rate 1000 URLs and then had each URL rated three times. Based on this, they claim that if the content is more interesting it will produce longer cascades of information sharing on average.

2.4.3 Influence & Centrality

The researchers Cha, Haddadi et al. use the Twitter API to get the tweets of 54 million users whom in total had 1.7 billion tweets and 2 billion follow links (Cha, Haddadi et al. 2010). This paper discussed how three measures captured different perspectives. Indegree, or the number of people who follow a user, is a valid measure of a user's popularity but does not relate to how much influence this user has. Driven by the value of a tweet, Retweeting and Mentions are indicators of the value of the user. Thus, those who have a high indegree are not those who have the most influence. The researchers state that in order to become more influential, a user must concentrate effort on their topic of interest or post tweets that are of high value to the other users. Also of note is that most influential users have significant influence over many topic areas.

To measure and understand influence in Twitter there has been previous work that has tried to understand how influential certain individuals are (McPherson, Lovin et al. 2001), (Kempe, Kleinberg et al. 2003). However, this work does not evaluate the importance of followers and the user followed on Twitter to understand influence. Weng et al. (2010), found that 72.4% of users in Twitter follow more than 80% of their followers, and that about 80% of Twitter users have 80% of their followers following them back. This finding indicates that reciprocity is a large factor in influence – that who you know will influence you, and that you will influence those you know. This is important for my research because as people interact on and offline, the reciprocity of following may change.

Rodriguez, Leskovec et al. (2010) created an algorithm for tracing paths of diffusion and influence and validated it using over 172 million webpages from 1 million online sources over a period of one year. They called this algorithm NETINF. By creating and studying this algorithm, they discovered that there is a core-periphery structure that mass media influences highly. Communities of sites form across topics such as politics, sports, business, etc.. From these sites, few have the influence across clusters to diffuse information between them.

2.4.4 Retweeting

Apart from individual microblogs, there has been additional research that has examined the effects of different forms of communication in Twitter, such as retweeting. Liere (2010) found in his study that retweets were used as a way to determine which user's information flows through, and when, on Twitter. As a way to measure retweets, he defines the geographical distance traveled by a tweet as "the distance (in meters) between the sender of tweet and the receivers who retweets that message." He finds that even though Twitter makes the world smaller it does not make it that much smaller. For example, a retweet is like a vote for that tweet, just like linking to a webpage of a vote for the webpage (Brin and Page 1998). Liere further explains the three reasons why a person would retweet: vying for attention of the originator, gaining influence in the network by acting as a filter and a vote for the tweet, and transferring information from one group to another.

2.4.5 Replies

Other work has demonstrated how Twitter facilitates the way in which people use the status message update feature to ask questions. Often referred to as "social search" (Brooks 2010), this behavior involves finding answers with the help of friends in the social network or even with unknown human resources. There are many motivations like trust, subjective nature of questions, and human expertise for using social search (Brooks 2010).

Over time, Twitter has shifted from a personal status sharing medium to an information sharing and conversational medium, which is important when considering how large groups of tweeters use it. The way we share and find information has changed with the introduction of microblogging networks, such as Twitter, to the Web. Brooks and Churchill (Brooks 2010) recorded how users search local news, find shopping deals and use Twitter for getting

recommendations from members of their social network. This service may seem similar to tools like Aadvark²⁶, which allows users to tap into their friends' social networks for information. However, the researchers mention important differences: Twitter responses are not anonymous nor are they generally decentralized from the user's social network. Additionally, the study found that Twitter users use it for time sensitive and location-based information.

2.5 Tools

2.5.1 3rd spaces (Ambient displays)

Ambient displays are relevant for my research given that they are used regularly at conferences by conference organizers for the purpose of maintaining a pulse on the online community, by the audience for providing comments about a talk in back channels (boyd 2010), and by attendees to orient towards what others are saying. In this section, I discuss the research on ambient displays using Twitter.

The first work done on ambient displays and Twitter was Hazlewood, et al. and was named Twitterspace (Hazlewood, Ryan et al. 2008). Twitterspace was a way to blend the virtual and physical spaces through the use of a public display of published tweets, giving the sense of *community-at-a-glance*. Once a user followed the account, the tool added them to the visualization stream for Twitterspace. One of the goals of this project was to open Twitter up to a wider audience including those without Twitter. By publicly displaying the stream in the community center, Twitterspace was able to reach that goal. One of the outcomes was a verification of the work of Java, Song et al. (2007) in identifying four common user intentions for Twitter users – that of sharing information more broadly. Hazlewood also observed that Twitter users form their own experience in Twitter arranging their virtual space as they see fit. They also observed Twitterspace as a way to comment in the back channel of Twitter on speakers, classes etc .

After the work of Twitterspace, Buzeck and Müller created TwitterSigns (Buzeck and Müller 2010). This research is an extension of the work on ambient displays comparing two different interfaces TwitterSigns and iDisplays (Müller, Exeler et al. 2009). iDisplays would show the information from the user group on the left and then on the right it would show data about the surrounding environment. TwitterSigns would show one tweet at a time displayed on map with the location and the photos of the writer. The researchers concluded that even though the user spent more time observing TwitterSigns over iDisplays (though not significantly), passersby would only remember tweets from users they knew or could affect them. The researchers concluded that the user considered the rest of the tweet trash.

²⁶ See vark.com

2.6 Summary

In this chapter, I have presented the broad spectrum of research on the use of Twitter. This research has demonstrated that people use Twitter for a diverse array of purposes, and that the information that is being collected on Twitter can be used as first ground-truth source of information but also for other applications such as detecting cell phone issues (Qiu, Feng et al. 2010). The measurements of the actions of Twitter users also span the qualitative and quantitative, with the majority of the research on the later. From these measurements we know that there are influencers in star-like networks of users, but that most users have an equal follow-to-follower ratio. It was also reported that these interactions are to change as people communicate on and offline. The content of tweets has generally focused on the topics covered inside Tweets with little ability to examine these further given the broad array of topics that are covered.

3 Methodology

My study focused on finding how participants at a conference used Twitter. The first step was to choose the conference where we would run the study. I made the decision to choose The Grace Hopper Celebration of Women in Computing 2010 Conference. The most significant reason that I chose this conference was because each attendee, during conference registration, was asked to share their Twitter username. If the attendee chose to do this their username, the conference organizer added it to a public Twitter list. This was a large benefit to studying Twitter at conferences. Other work in this space would only be able to analyze hashtags from the conference and assume that users using the hashtags were indeed attendees. By using this list we had a pool of users who at least registered for the conference and quite possibly attended the same. By the time the conference started, there were 399 users on the Twitter list. The conference also pre published a list of hashtags for attendees to use (see Table 1).

3.1 Data Collection

Once I selected the conference, work on a custom data collection application began. The program was a custom C# .NET application that had a SQL Express database as a datastore. The application used Twitterizer²⁷ to talk to the Twitter API. The program would run a sequence of data collection every hour on the hour. It would go the Twitter list of all attendees and retrieve the users from the list getting basic information about the users, such as name, number of follower/friends, location etc., and insert the data into the database. After this was complete the application would go through each user and get all their current friends and followers. Then the application would search Twitter for any tweets that contained any of the hashtags that the conference pre-published (see Table 1 Track/Session Hashtags). I started the application two weeks before the conference began and then ran it for a total of five weeks.

After the conference was completed, I started a secondary collection phase that went through all the users from the database and retrieved all the tweets from their timelines during the five week period that the study was running (see 3.3 Quantitative Analysis).

3.2 Survey

We also created a survey. The survey had 57 questions and asked about the user's demographics, regular Twitter usage, and conference Twitter username?. I distributed the survey, by tweeting messages and handing out business cards, see Figure 12 and Figure 13 for examples. I would tweet out one of the following tweets about 5 times in the course of 24 hours, starting after the first full day of the conference. I also would hand out business card to people I met at the conference. There were 78 completed surveys. I hosted the survey on

²⁷ <http://www.twitterizer.net/>

SurveyMonkey²⁸ (@SurveyMonkey). I upgraded to a paid account in order to get SSL and HTTPS security as well as remove the limit on the number of questions that a free account has. The questions asked in this survey are located in Appendix B: Survey.



Figure 12 Sample of Tweets Advertising Study



Figure 13 Business Card Advertising Study

²⁸ <http://www.surveymonkey.com/>

3.3 Quantitative Analysis

3.3.1 Timeline

In order to explore Twitter users at conferences we collected and analyzed two distinct timelines: the users' timelines and a timeline consisting of all the tweets that contained the conference hashtags (see Table 1 Track/Session). For both of the timelines, it was important to analyze several factors over the five week period of the study and how these factors changed over time and focus in on the behavior of users before, during, and after the conference. The factors that we took into account were: number of tweets, hashtags, links, mentions, retweets, replies.

In particular to the user timelines, we also analyzed how often the tweeter used a conference related hashtag during all phases of the study. For the conference hashtags specifically, I constructed a timeline of events by using the conference program and then matched against conferences hashtags in order to analyses the engagement per event.

3.3.2 Social Graph

The social graph in Twitter is the global network of individuals linked together by users following each other. Every social network has their own social graph and is a cornerstone of these services. One of the purposes of conferences is to network with other individuals with similar interests. It was important to analyze several factors of the social graph over the five week period of the study and how these factors changed over time and focusing in on the behavior of users before, during, and after the conference. The areas of focus for this study are the number of friends and followers in total, but also the number of added and removed friend and followers across the study. It was also important to examine if the participants where connecting with other participants in order to see if users were mimicking networking behavior in the virtual world.

3.3.3 Conversation

One of Twitter's main uses is communication, either one-on-one or conversational. For the purposes of this study, we focused on the length of the conference, span of conversations, the number of participants in the conversations, the ratio of participants on each conversation to the number of status updates in the conversation, and the ratio of the conference participants to external users.

3.4 Conference Hashtags

The Grace Hopper Conference published a list of hashtags that they wanted the conference attendees to use during the conference²⁹. We used this list to pull all tweets regardless of who created the status update and added this data to the database.

²⁹ <http://gracehopper.org/2010/community/ghc-on-twitter/>

Table 1 Track/Session Hashtags

TRACK/SESSION HASHTAGS (use with #ghc10)

Track/Session	Hashtag(s)
Academic	#acad
Award Winners	#award
Birds of a Feather (BOF) Sessions	#bof[1..14] Table 2
Career	#career
Informational Career Development	#info
Practical Career Development	#pract
CRA-W Undergrad	#ugrad
CRA-W Grad	#grad
CRA-W Early Career Researchers	#early
Industry	#indst
Invited Technical Speakers	#invtech
K-12 Teachers in Computing Workshop	#k12
Leadership Workshop	#ldrshp
New Investigators 1-3	#newinv[1..3]
PhD Forum 1-4	#phd[1..4]
Science	#sci
Steering	#steer
Students	#stdnts
Technical Executive Forum	#exec
Technical Theme: HCI	#hci
Technical Theme: Open Source	#osrc
Technical	#tech
Theme: Collaborating Across Boundaries	#theme
Women Build	#build

Table 2 Birds of a Feather (BOF) HashTags

BIRDS OF A FEATHER (BOF) HASHTAGS

BOF Session	Hashtag
Mobile Computing: The Internet is in Your Hand	#bof1
Girls, Games, and Getting to the First Day	#bof2
Are You a Salmon Too?	#bof3
Barbie has a Pink Laptop: Redefining How the World Views a Computer Scientist	#bof4
Running an Outreach Program to High School Girls	#bof5

Conference Networking Across Boundaries	#bof6
Mommies Know Best: Pregnancy, Graduate School, & Beyond	#bof7
Minorities without Borders: Giving Back to Developing Countries	#bof8
Confidential Advice Session on How to Succeed as a Junior Faculty Member	#bof9
Women Thriving in Leadership	#bof10
Speed Mentoring for Latinas in Computing	#bof11
Education Across International Borders	#bof12
Women in Tech Employee Resource Groups: Challenges and Success Stories	#bof13
Influencing Without Authority: Collaboration from Idea to Implementation	#bof14

3.5 Conference Setting

The conference, as described by The Grace Hopper Celebration of Women in Computing website, was “Co-founded by Dr. Anita Borg and Dr. Telle Whitney in 1994 and inspired by the legacy of Admiral Grace Murray Hopper. The Institute’s Grace Hopper Celebration (GHC) Of Women in Computing Conference is designed to bring the research and career interests of women in computing to the forefront. It is the largest technical conference for women in computing and results in collaborative proposals, networking and mentoring for junior women and increased visibility for the contributions of women in computing. Conference presenters are leaders in their respective fields, representing industry, academia and government. Top researchers present their work while special sessions focus on the role of women in today’s technology fields.” (2010)

The 2010 Conference was in Atlanta, Georgia from September 28th to October 2nd. The conference was located at the Hyatt Regency Atlanta and sold out at 2,149 attendees. This was a 37% increase in growth over the previous year (2010) . The conference provided free Wi-Fi in all public spaces and in meeting rooms. The conference area was located across three different floors and two different towers that had a passageway at the bottom level and through the lobbies of the two towers as seen in Figure 14. Throughout the conference space and in the lobby area were seats for people sit throughout the day and have conversations. The hotel also had a two restaurants, two bars, and a 24/7 coffee and snack bar where attendees could gather to connect.

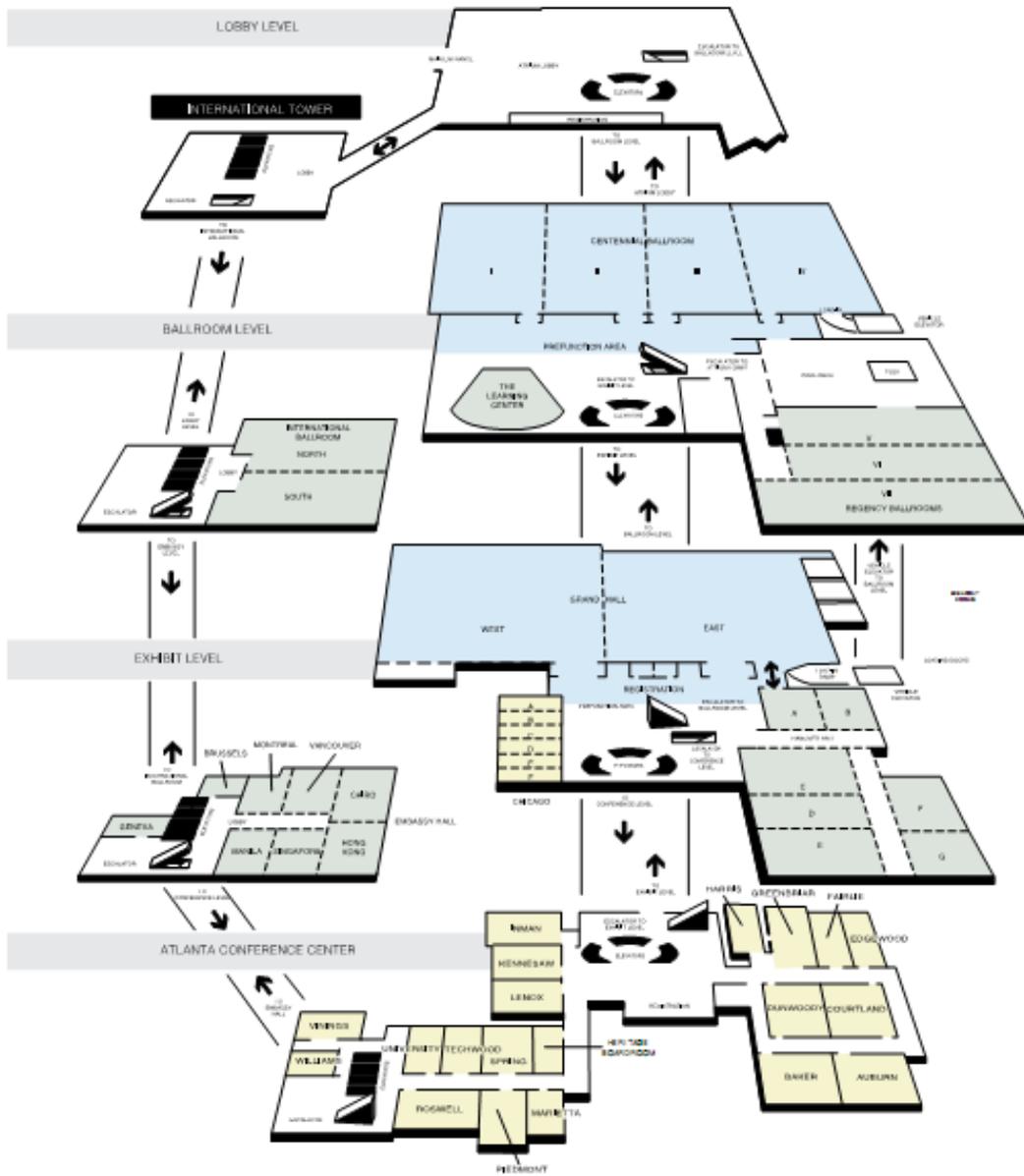


Figure 14 Hyatt Regency Atlanta Conference Center

3.6 Tools Used to Analyze Data

Once I collected all the data by putting it into a database, additional data manipulation was required. Each table that has a date column had extra columns created. I created a day, hour, minute, and week column then filled out each based on the date in each row in the table. Based

on the number of tweets during the five week period, I created a new table assigning a user to a group. After the data manipulations and additions, I imported all the data into Microsoft Excel.

Excel has a feature called the Pivot Table tool. Pivot tables allow you to quickly and easily summarize large amounts of data for analysis. A pivot table can automatically count, sum, average, min, and max. It also allows for sorting and filtering of the data. Excel does this by taking the raw data and summarizing dynamically on the field you choose by dragging and dropping graphically onto the worksheet. The ability to rotate or pivot the table gives the model its name.

After using Excel to summarize the data, I copied the data into a secondary worksheet to perform extra calculations, such as percentages, differences, or multiples. In some instances the only method available was to execute SQL queries directly against the database and report of the findings from there. In order to keep records, I created Excel a separate worksheet with the results of these queries.

I generated word clouds based on the text of the tweets per week for the entire group of participants. Wordle, the application, removed common English words and I removed the word RT. All other Twitter features have remained such as mentions and hashtags. The tool automatically removed the # character from hashtags in the analysis, but the words still appear. Only the top 150 words appear in the Word Cloud.

Most social graph statistics and graphs were generated used Gephi³⁰. Gephi is open source software for network analysis and visualization. I chose it for this study for its ability to render large graphs quickly. It also had the features to change layouts, generate metrics, dynamic network analysis, export with customizations, clustering, and filter. From the database I would generate .dot or .gexf files to represent my graph and then import them into Gephi for network analysis. I generated the .gexf files to support dynamic networks such as conversations to get better visualizations of the data. Once I imported the data I ran the metrics tools against the data and then used the visualization tools to color and size nodes and lines appropriately. The Network Centralization Indexes were generated using UCINET³¹. From the database I had to generate a Data Language file representing the graph. I would tell UCINET to import and run Freeman Node Betweenness, and it would generate the Network Centralization Index numbers for the graph.

³⁰ <http://gephi.org/>

³¹ <http://www.analytictech.com/ucinet/>

4 Results

4.1 Grouping

In this study, we were able to classify all users, 399 people that were part of the conference twitter list, into six categories. This categorization makes comparisons easier. Below I define these six categories: inactive, lurker, experimenter, part-time, active, and immersed. I have listed the groups below from least to most active users.

- a) Inactive Users are users who have accounts but did not have any activity during the monitoring process. There were 46 inactive users during this five week period. Inactive users could indicate that during the conference they were inactive, or that, in general, they had stopped participating in Twitter.
- b) Lurker Users are users who have Twitter accounts, but who did not participate in the conversation on Twitter during the five week period. They did, however, add and remove friends and ended up with a change in followers. The only way we know they exist in this study is because Grace Hopper organizers added them to the public Twitter list when they registered for the conference. If it was not for this fact, they would have been invisible. There were 53 lurkers of at the conference. 37 of these lurkers added followers, but no friends. 1 lurker added friends, but did not connect with any followers. 15 lurkers made friends and followers.
- c) Experimenter Users only interacted with a minimum number of users and only participated in the conversation occasionally. They tended to follow more users than users followed them back. These users microblogged sporadically during the five week period.
- d) Part-time Users use twitter irregularly. As a whole, this group has more people following them than they follow other people. What is interesting about this group is that they mention other users in a larger percentage of tweets than the rest of the groups.
- e) Active Users use twitter often, but they are in the average range for all stats. This group of users had the biggest gain in terms of friends and follower increases during every week in the collection period and the biggest gain in the week of the conference.
- f) Immersed Users are users committed to using Twitter every day. These users tend to have large numbers of friends and followers. They tweet an inordinate amount, 4.6 times the next highest group and 74 times the lowest group and use hashtags and reply to other users more than any other group.

Inactive	No tweets and no new friendships
Lurkers	No tweets
Group 1 - Experimenter	1-9 tweets
Group 2 - Part-timer	10-33 tweets
Group 3 - Active	35-98 tweets
Group 4 - Immersed	102-1231 tweets

Table 3 Group Distribution

We were not able to include all the users in our data collection. Two reasons for this are:

- a) Private Users are users who have accounts but have set their accounts to be private. We could not collect their data and therefore we have had to exclude them from our study. This does not mean that they were not participating with people at the conference, but that their interactions were not observable. There were 53 private users.
- b) Data collection errors. A second group of users had Twitter accounts, but due to technical issues with Twitter, we could not retrieve the users' timelines. Therefore we have excluded these users from data collection. There were 16 of these users.

4.2 Tweeples

4.2.1 Demographics

78 users completed the survey. Out of the 78 users, 58 users choose to give their Twitter username. All 78 survey takers were woman. The survey takers are highly educated as Figure 15 shows. We can assume that the some college means they are still attending a college. Some post graduate means they are a masters or PhD student. Post Graduate Degree most likely means they are professors, and college graduate means they are professionals. 84.6% resided in the United States, 7.7% in Canada, 2.6 Western Europe, 1.3% Africa, and 3.8% other. 6.4% identified themselves as Hispanic, Latino, or Spanish. 82.1% acknowledged themselves as White, 9% as Asian, 3.8% as Black or African American, and 5.1 other. Figure 16 depicts that most of the participants fell into the age range of between 21 and 40.

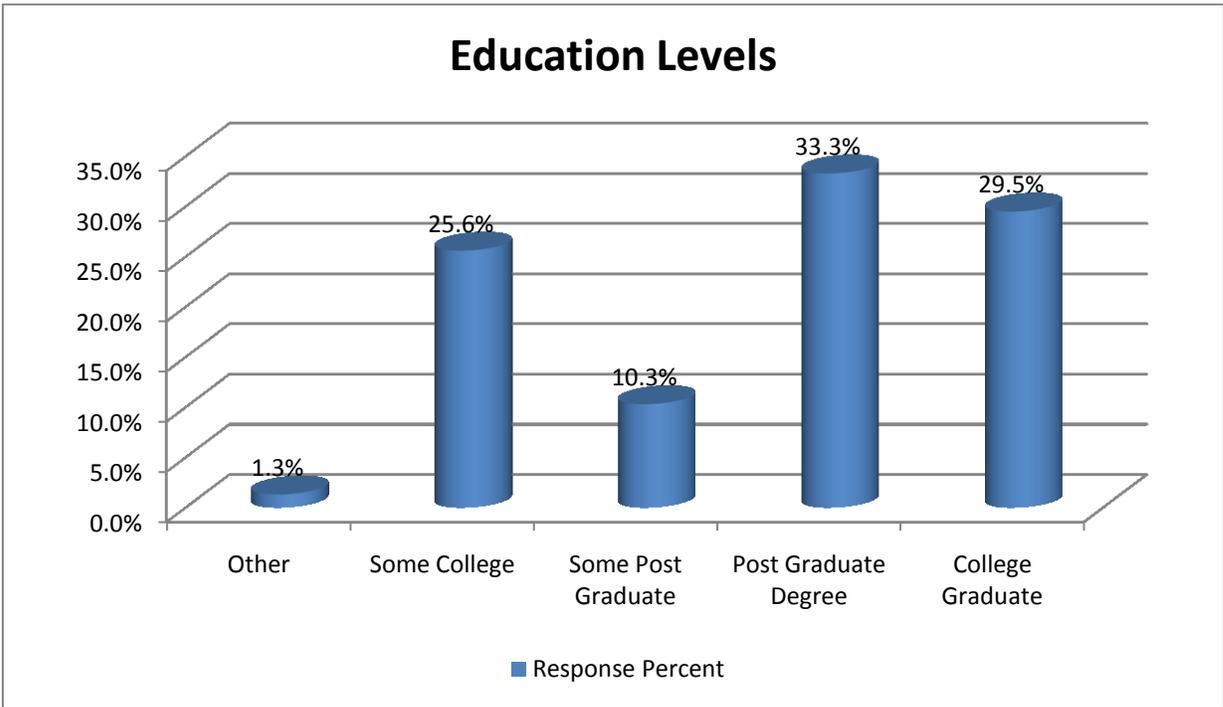


Figure 15 Vocation of Survey Participants

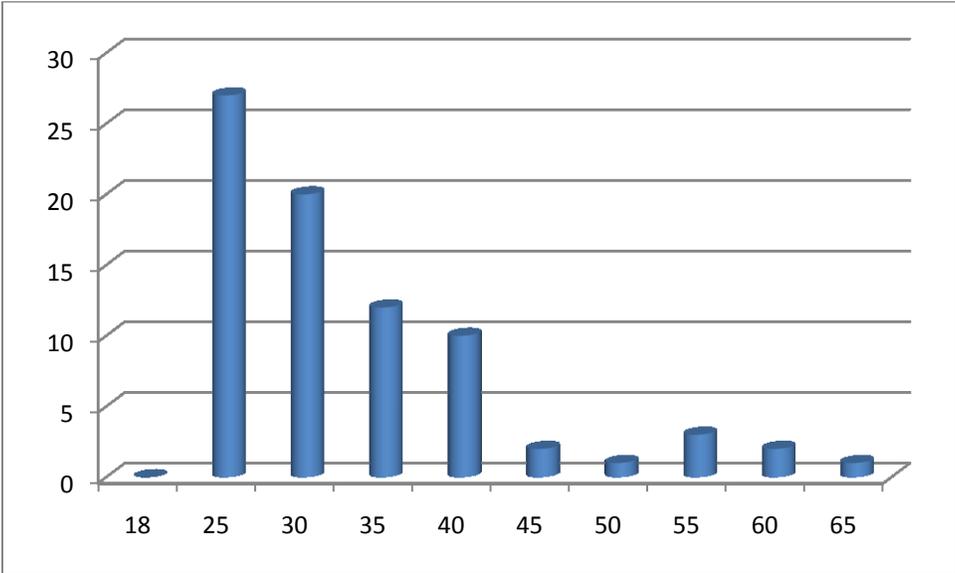


Figure 16 Histogram of Age of Participants

4.2.2 Timeline of tweetings/activity

When a user creates a status update responding to the question “What’s happening?” in Twitter, it is called a tweet. Each tweet or message must be less than or equal to 140 characters. These graphs show how active a user is on Twitter.

4.2.2.1 Group 1- Experimenter

Group 1 - Experimenter had a total of 221 tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Tweets	29	25	97	39	31	221	<i>total</i>
Percent Change		-14%	288%	-60%	-21%		

4.2.2.2 Group 2- Part-timer

Group 2 - Part-timer had a total of 1077 tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Tweets	107	159	521	158	132	1077	<i>total</i>
Percent Change		49%	228%	-70%	-16%		

4.2.2.3 Group 3- Active

Group 4 - Immersed had a total of 3532 tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Tweets	509	618	1408	449	548	3532	<i>total</i>
Percent Change		21%	128%	-68%	22%		

4.2.2.4 Group 4- Immersed

Group 4 - Immersed had a total of 16320 tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Tweets	2296	3018	5506	2821	2679	16320	<i>total</i>
Percent Change		31%	82%	-49%	22%		

4.2.2.5 Summary

In all there were a total of 21,150 tweets made by the Twitter users of this study. As demonstrated in Figure 17, Group 4 (Immersed) had the most tweets. However, all groups had an increase in the number of tweets during the week of the conference. This means that for the week of the conference, participants approximately doubled the number of times they tweeted. For example, Group 3 went from 618 tweets in Week 2 to 1408 tweets in Week 3 (the week of the conference).

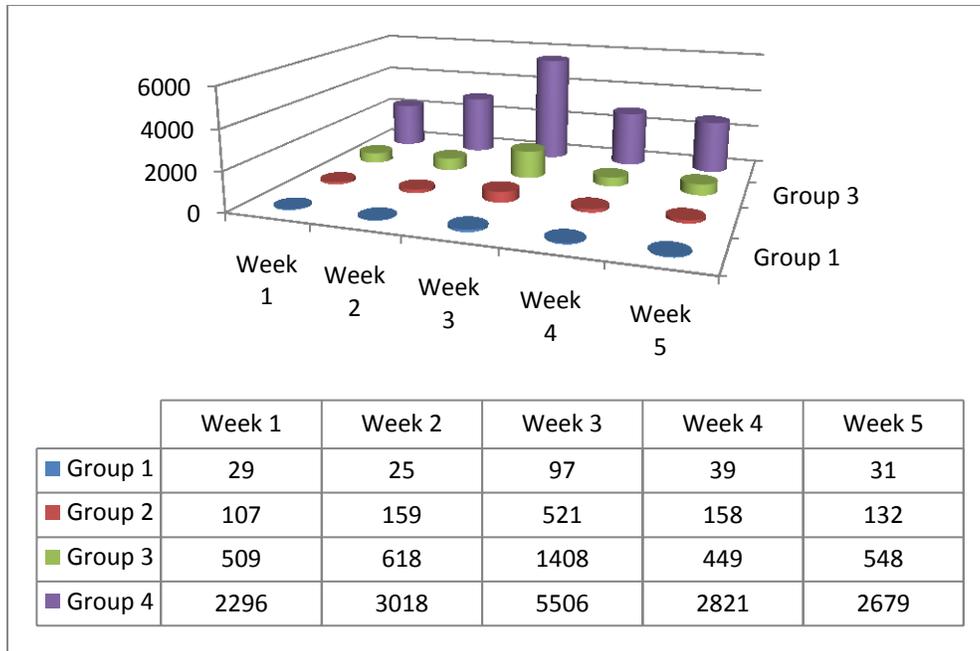


Figure 17 Number of tweets

4.2.3 Characteristics

The participants of this study also had other relevant characteristics. First, 60.3% of tweeters used Twitter for both personal and professional purposes, as demonstrated in Figure 18. Of the remaining participants, 25.6% used Twitter for personal purposes, and 14.1% used Twitter for professional purposes.

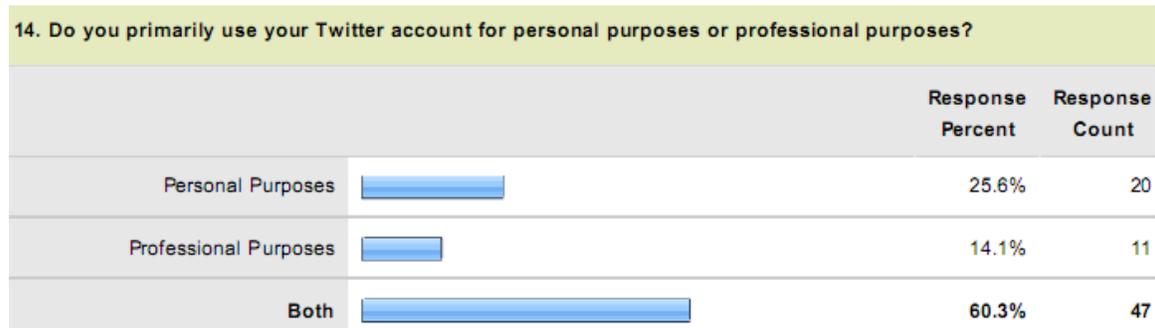


Figure 18 Primary Use Of Twitter

Second, 67.9% of the tweeters had used Twitter to discuss topics with other users by sending direct messages, as demonstrated in Figure 19. This means that the remaining 32.1% of users always have public tweets. This is evidence that participants use Twitter for direct and person-to-person communication.

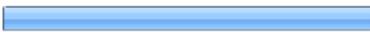
17. Have you discussed topics with other users of Twitter (Tweeple) by sending direct messages?			
		Response Percent	Response Count
Yes		67.9%	53
No		32.1%	25

Figure 19 Use of Direct Messaging

Third, similar to the use of direct messages, 91% of the tweeters had discussed topics with other users by replying to tweets, as demonstrated in Figure 20. This means that most of the people in this study had public conversations on Twitter and is evidence again that participants use Twitter for direct and person-to-person communication.

18. Have you discussed topics with other users of Twitter (Tweeple) by replying to tweets?			
		Response Percent	Response Count
Yes		91.0%	71
No		9.0%	7

Figure 20 Use of Replying

Fourth, in general, people in this study use Twitter for diverse reasons, as demonstrated in Figure 21. The most frequent reasons provided by participants was for networking (76.9%), as a news source (69.2%), to establish an online presence (65.4%), for personal communication (65.4%), and to participate in discussion (64.1%). The least cited reasons for using Twitter was to jot down notes (25.6%), and to pose organization questions (15.4%).

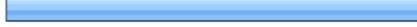
19. Why do you use Twitter (Select All the Apply)?			
		Response Percent	Response Count
Sharing Resources		55.1%	43
Personal Communications		65.4%	51
Business Communications		47.4%	37
Participate in Discussion		64.1%	50
Jot down Notes		25.6%	20
Networking (overall)		76.9%	60
Establish Online Presence		65.4%	51
News Source		69.2%	54
Pose Organization Questions		15.4%	12

Figure 21 Why do you use Twitter

When asking participants to then rank which of the uses of Twitter in Question 19 were most important, the results from the previous question shift into relevance. As demonstrated in Figure 22, participants ranked networking, personal communication, and news source as the first reason to use Twitter. Participants then ranked participating in discussion, networking, and news source as the second reason. Last, participants ranked establishing an online presence, news source, participation in discussion, and sharing resources as the third reason.

20. What are your top three reasons for using Twitter(Rank them in order)?					
	1st	2nd	3rd	Rating Average	Response Count
Sharing Resources	12.5% (3)	45.8% (11)	41.7% (10)	1.00	24
Personal Communications	63.3% (19)	23.3% (7)	13.3% (4)	1.00	30
Business Communications	56.3% (9)	12.5% (2)	31.3% (5)	1.00	16
Participate in Discussion	16.7% (5)	50.0% (15)	33.3% (10)	1.00	30
Jot down Notes	0.0% (0)	30.0% (3)	70.0% (7)	1.00	10
Networking (overall)	46.7% (21)	33.3% (15)	20.0% (9)	1.00	45
Establish Online Presence	22.0% (9)	26.8% (11)	51.2% (21)	1.00	41
News Source	31.4% (11)	40.0% (14)	28.6% (10)	1.00	35
Pose Organization Questions	100.0% (1)	0.0% (0)	0.0% (0)	1.00	1

Figure 22 Top three reason for using Twitter

Fifth, the participants in this study had a range of the amount of time they had a Twitter account, as demonstrated in Figure 23. 37% of users had accounts for 1-2 years; 25.6% of users had accounts for 2-3 years; 3.8% of users had an account for more than three years; the remaining users had accounts for less than a year. This means that approximately a third of the people who tweeted at the conference were relative novices to Twitter. However, the remaining users are relative experts given the relative age of Twitter.

21. How long have you had a Twitter account?			Response Percent	Response Count
Less than a month		3.8%	3	
One month to six months		14.1%	11	
six months to a year		15.4%	12	
one year to two years		37.2%	29	
two years to three years		25.6%	20	
more than three years		3.8%	3	

Figure 23 Length of Twitter Use

Participants generally limited their daily tweeting, as demonstrated in Figure 24. Most participants (92.3%) tweeted between 0-10 times a day. The remaining 7.6% tweeted between 11-30 times daily. No one reported tweeting more than 30 times a day.

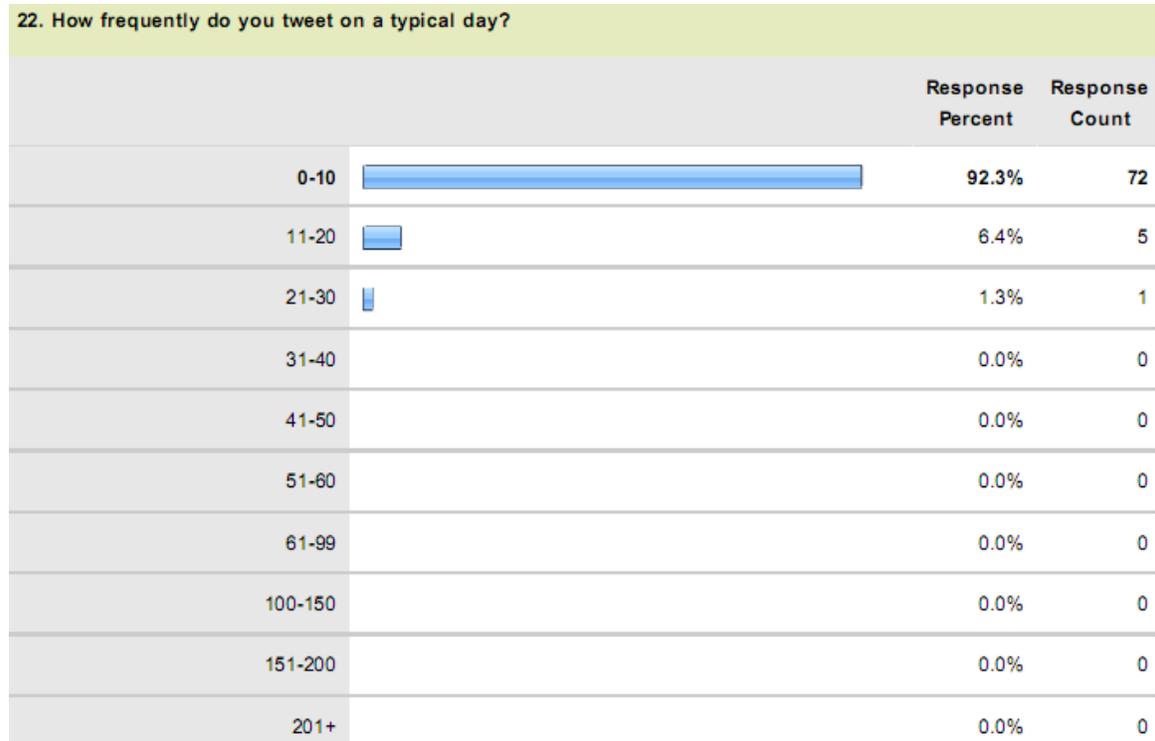


Figure 24 Frequency of status update per day

Given the necessity of mobility for people attending a conference, only 39.7% of them reported using a mobile device application, as demonstrated in Figure 25. This means that participants had to use other platforms to tweet. The participants in this study reported using a desktop client (20.5%), and Twitter.com (39.7%) to manage the remainder of their tweets.

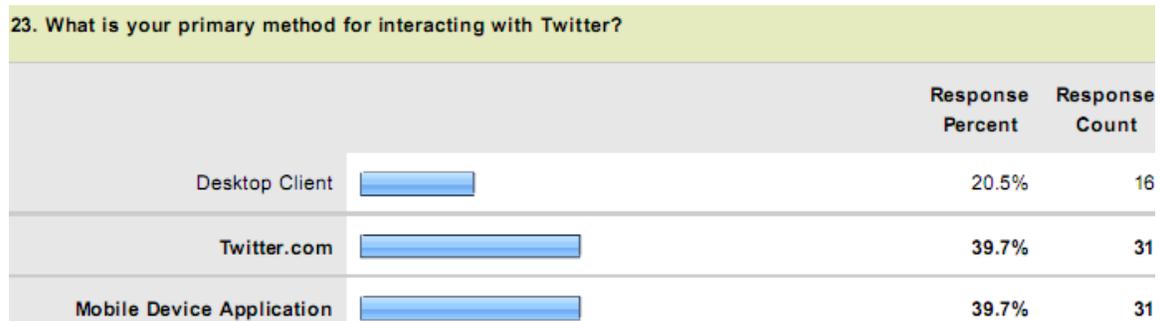


Figure 25 Primary method of interaction with Twitter

To probe the tools that people use to tweet further, researchers asked participants about their operating systems on their smartphones and their mobile devices. Participants evenly divided

between iPhone OS and Android OS, as demonstrated in Figure 26. However, there were 24.4% of users who reported not having a smartphone. For those users who did use a mobile device to interact with Twitter, 71.6% of participants used an application, as demonstrated in Figure 27. The remaining participants used text messaging (17.9%) and m.twitter.com (10.4%).

The number of people who reported using text messaging and m.twitter.com is similar to the number of participants who do not have a smartphone. This correlation may be because normal cell phones are still capable of text messaging and limited internet browsing.

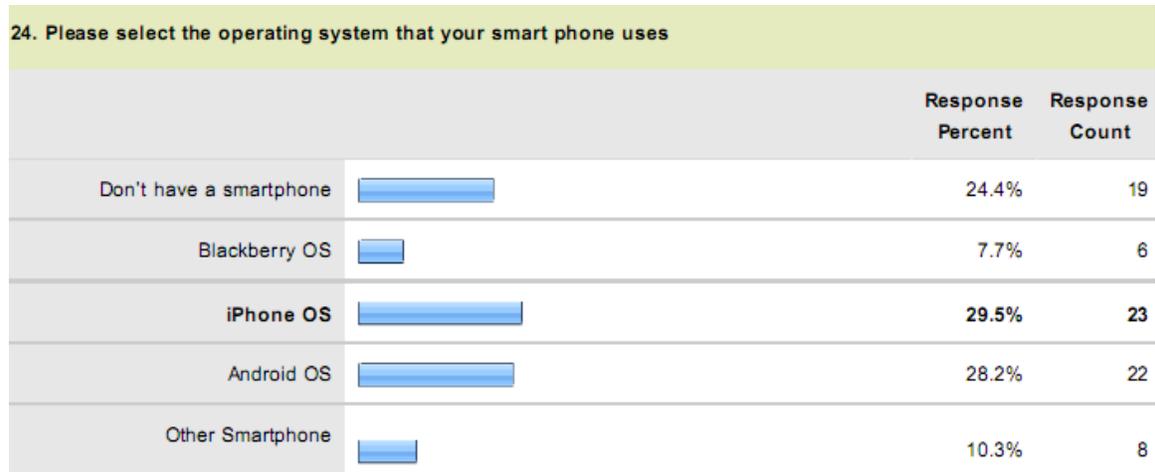


Figure 26 Smart phone OS



Figure 27 Method of interaction on mobile device

The participants in this study used hashtags frequently. Most of the people (92.3%) had previously used a hashtag while tweeting, as demonstrated in Figure 28. Participants also used hashtags recently (91%), as demonstrated in Figure 29. The shift from 92.3% to 91% indicates that 1.3% of participants have used hashtags, but had not used them recently.

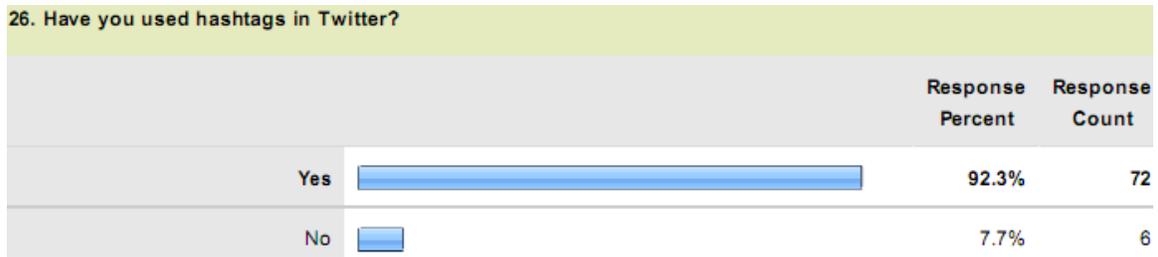


Figure 28 Use of Hashtags

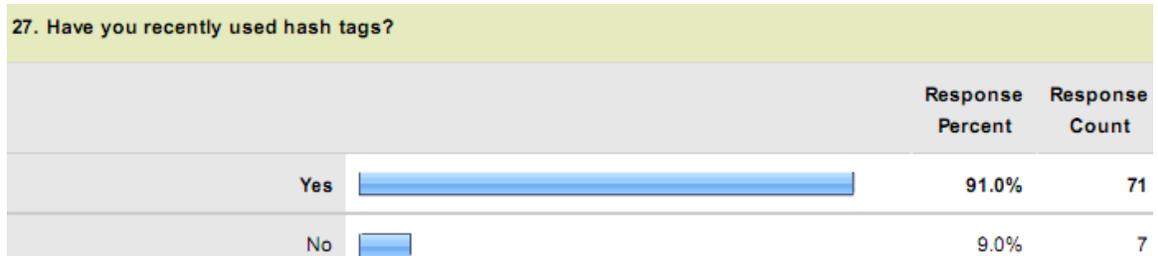


Figure 29 Recent use of hashtags

An additional feature Twitter provides is allowing its users to create lists of people. 55% of users had not created a list of users on Twitter, as demonstrated in Figure 30. Yet, only 45.5% of users reported officially following lists on Twitter, as demonstrated in Figure 32. Additionally, 20% of users reported periodically checking on lists without officially following them, as seen in Figure 31. 80% of participants reported not checking on a list of tweeters without officially following them.

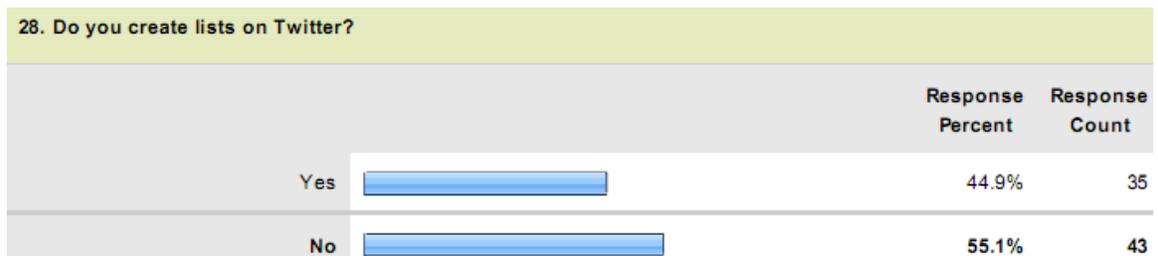


Figure 30 Creating list on Twitter

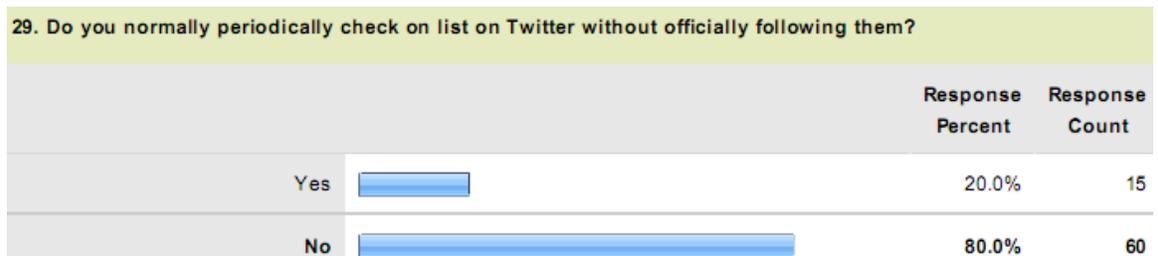


Figure 31 Check lists on Twitter

30. Do you officially follow lists on Twitter?			
		Response Percent	Response Count
Yes		45.5%	35
No		54.5%	42

Figure 32 Follow lists on Twitter

Whether or not Twitter allows users to have more than one Twitter account, 37.2% of the participants in this study reported having multiple twitter accounts, as demonstrated in Figure 33.

31. Do you have more than one Twitter account?			
		Response Percent	Response Count
Yes		37.2%	29
No		62.8%	49

Figure 33 Most than one account on Twitter

4.2.4 Source

From the data sources that I collected, I discovered how users published their tweets.

53% of tweets from users followed from the list came from Twitter official clients, as demonstrated by Figure 34. From the tweets found following the participants in the study that used the GHC hashtag, 49% came from Twitter official clients, as demonstrated in Figure 35.

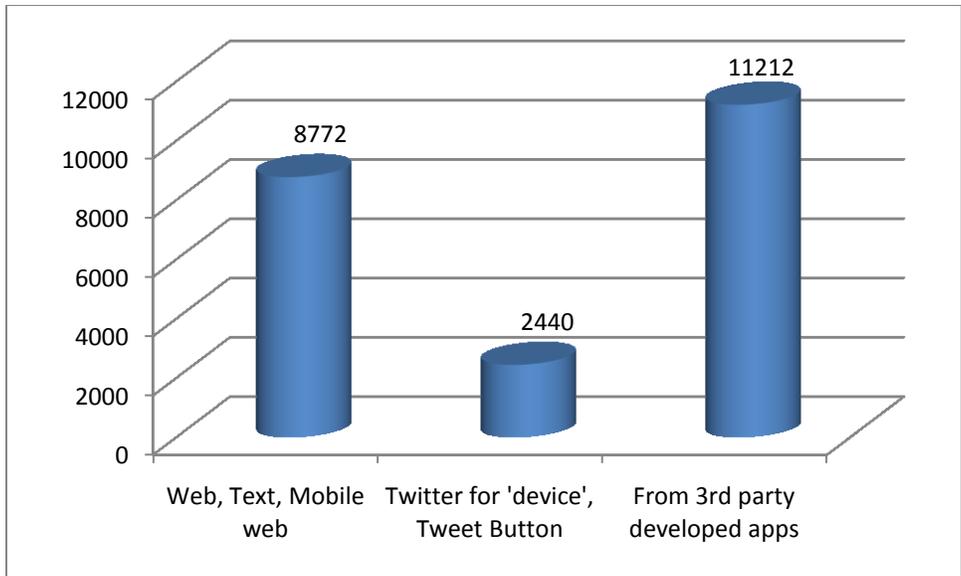


Figure 34 Number of Tweet by Source from participants tweets

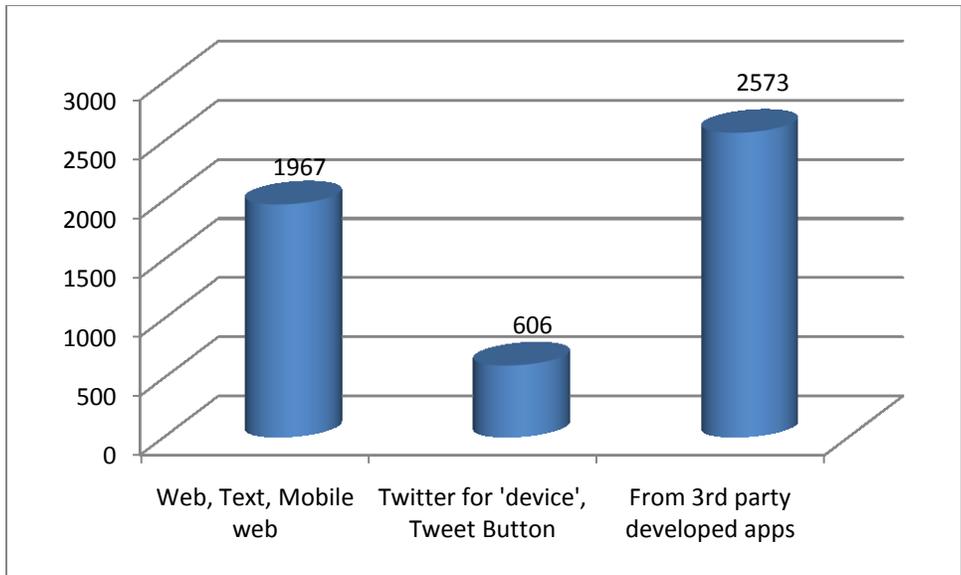


Figure 35 Number of Tweet by Source from following GHC Hashtags

4.3 Content & Description of Tweets

4.3.1 Characterization of Message

When examining the types of tweets, 60% were a type of conversation, as demonstrated in Figure 36. To break this down further, 16% were mentions, 13% were retweets, and 31% were replies. The remaining 40% were basic tweets.

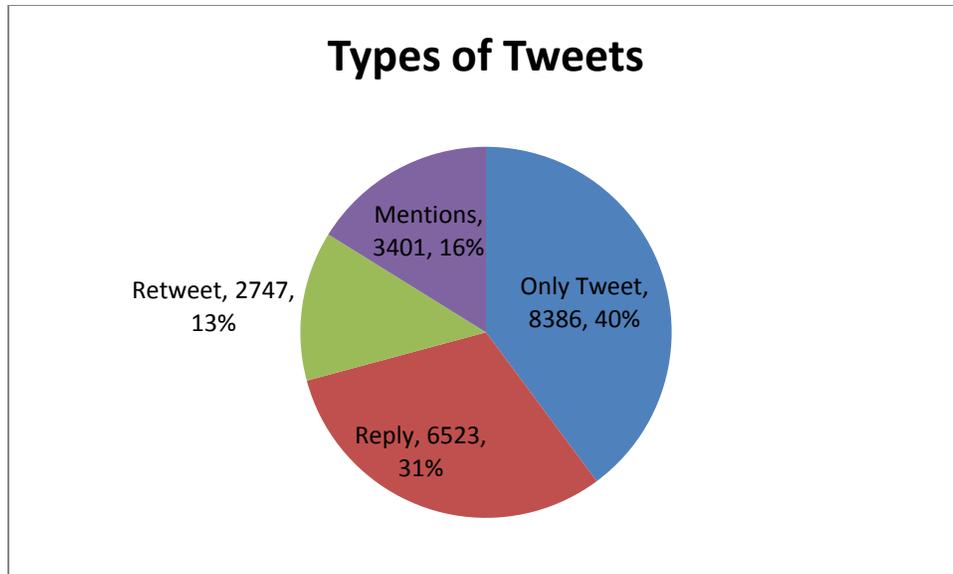


Figure 36 Types of Tweets

Again, as seen by examining the different groups in Figures 39, 40, 41, and 42, there is an increase in the number of tweets the week of the conference. However, different user groups had different types of tweets.

Group 1 - Experimenter, had a steep increase for all types of tweets during the week of the conference, but then types of tweets remained higher up to two weeks after the conference, as demonstrated in Figure 37. For example, only tweets remained a 9 to 10 tweet increase in Weeks 4 and 5, respectively. Additionally, Retweets went from two tweets in Week 2 to nine tweets in tweet 4. Replies and mention tweets remained relatively the same before and after the conference.

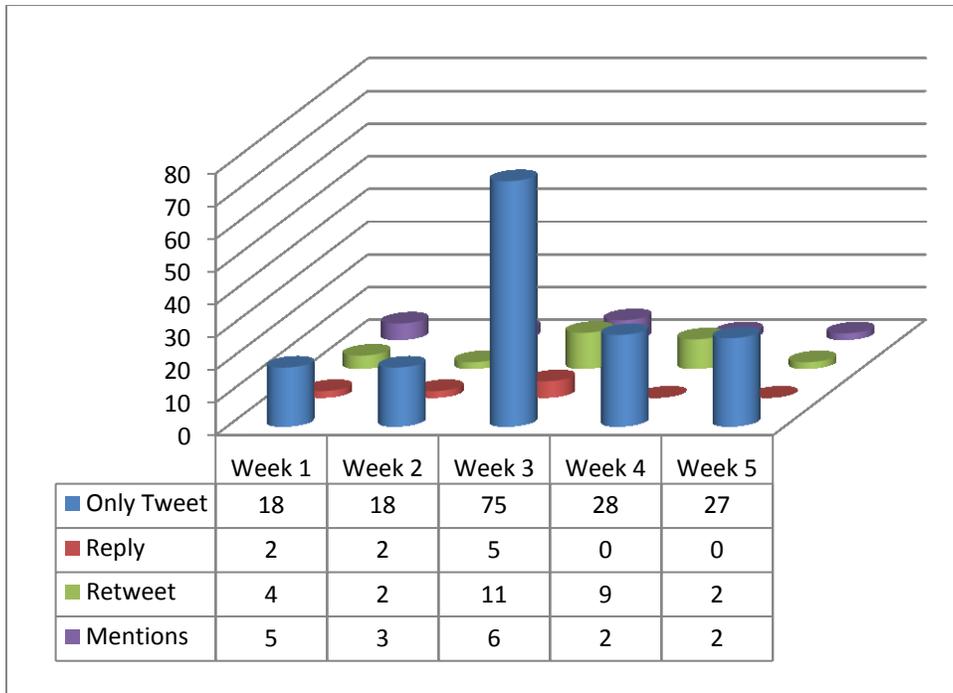


Figure 37 Group 1 - Experimenter Types of Tweets

For Group 2 - Part-timers, again there is a large increase the week of the conference, as demonstrated in Figure 38. However, the week after the conference there is a marked lull in all types of tweets. For instance, from the week of the conference to the week after the conference, the number of tweets decreased from 375 to only 28. The tweeters then appear to recover by week 5 with the number of tweets across all groups being similar to Week 2.

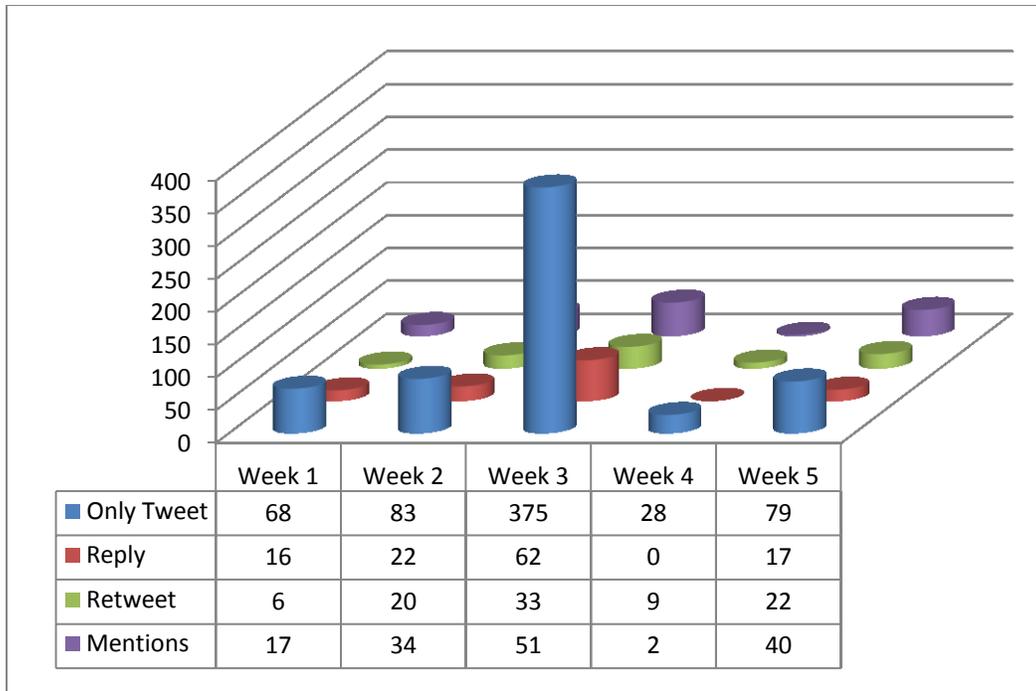


Figure 38 Group 2 - Part-timer Types of Tweets

For Group 3 - Active, there is a sharp increase in all types of tweets during the week of the conference, as demonstrated in Figure 39. However, for this group, the users appear to not have a recovery period after the conference. The tweets prior and after the conference are relatively the same. Yet, mentions decrease by about 30 average tweets the two weeks after a conference.

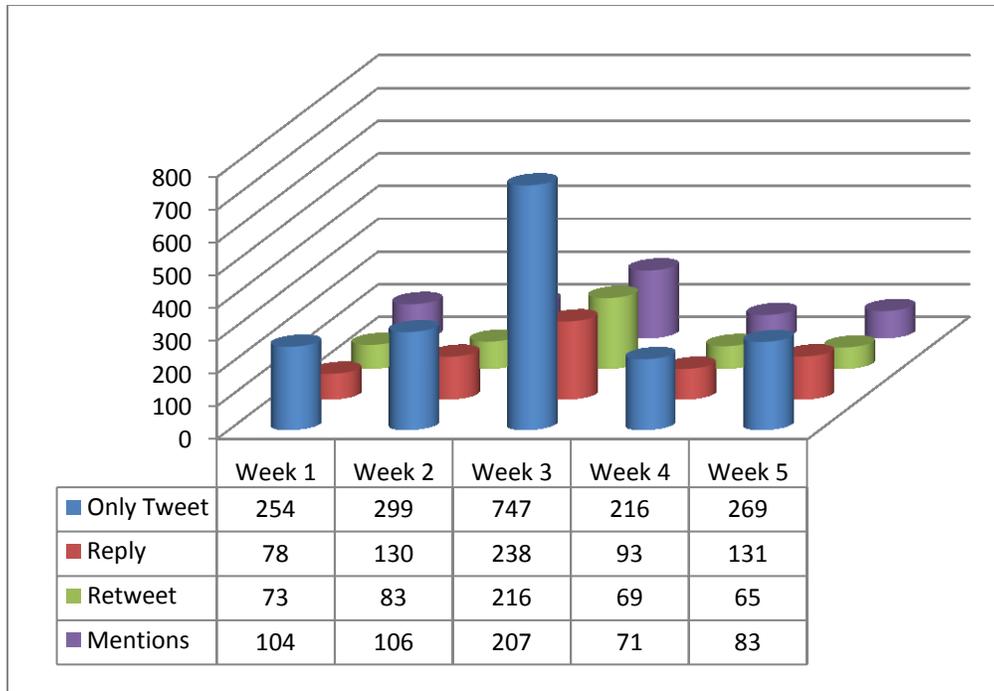


Figure 39 Group 3 - Active Types of Tweets

For Group 4 - Immersed, there is again a large increase the week of the conference for all tweet types, as demonstrated in Figure 40. However, there appears to be an increase for replies the week prior to the conference that remains high (200 tweet increase) for the two weeks after the conference. The remaining types of tweets appear to return to normal the two weeks after the conference.

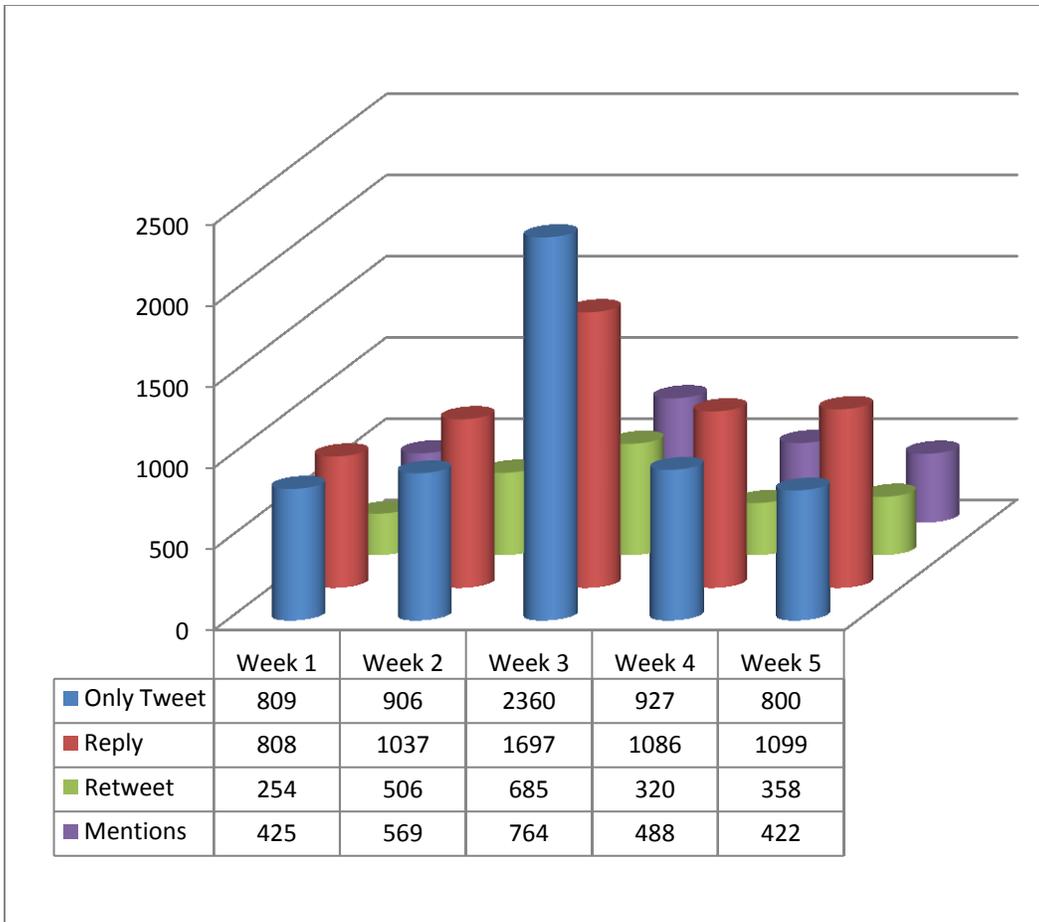


Figure 40 Group 4 - Immersed Types of Tweets

The examination of the content of tweets indicates that 45% of tweets included a URL, a hashtag, or both, as demonstrated in Figure 41. To break this phenomenon down further, 24% of tweets had a hashtag, 15% had URLs, and 6% had both. The remaining 55% did not use either a hashtag or a URL.

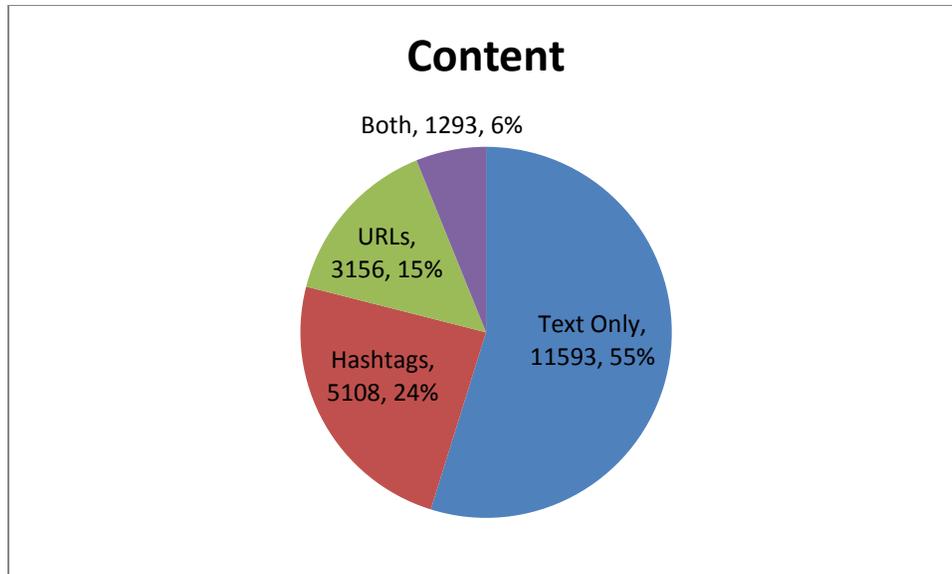


Figure 41 Content in Tweets

The content of tweets can be broken down by the different user groups and across weekly behavior. For all groups, there is a marked increase in all types of tweets the week of the conference. Additionally, for all groups, the largest ratio of change occurred for the use of hashtags. This may be due to the inclusion of the GHC hashtag to mark participation in the conference.

For Group 1 - Experimenter, there was a marked increase in all types of tweets the week of the conference, as demonstrated in Figure 42. During the week of the conference, the use of hashtags was the largest category of tweets. For text only tweets, there is a marked increase of tweets the week after the conference as compared to Weeks 1, 2, and 5. Apart from the peak in the middle and the increase of text only tweets, the tweets remain similar before and after the conference.

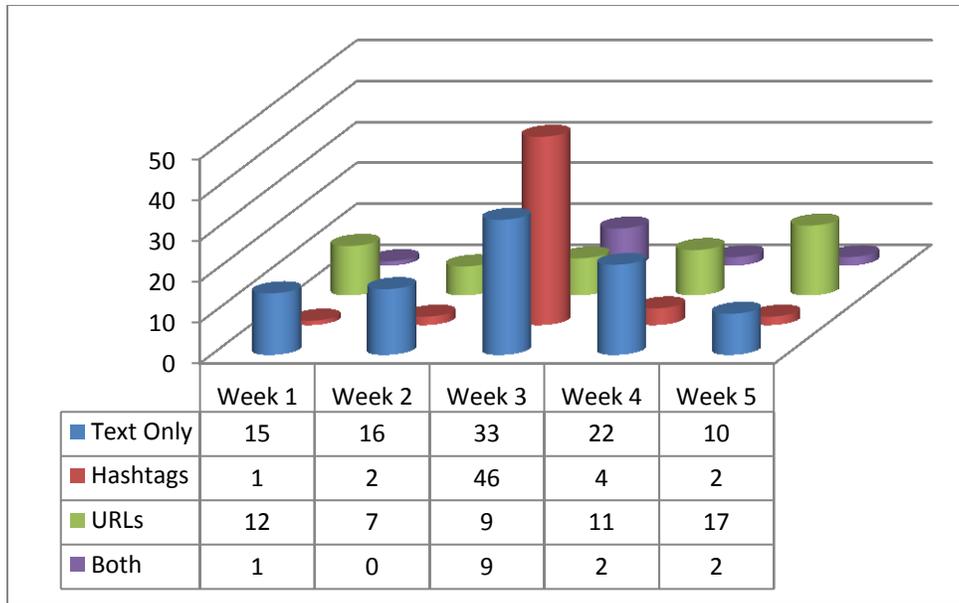


Figure 42 Group 1 - Experimenter Content

For Group 2 - Part-timer, there is a marked increase the week of the conference except in tweets that only use URLs, as demonstrated in Figure 43. The week of the conference the largest category of tweets was the use of hashtags.

For text-only tweets, there was a rise from Week 1 to 2, and then a large increase the week of the conference. Then the week after the conference, the number of text-only tweets is similar to that of the week before. By Week 5 the number of text-only tweets decreases again. This trend mimics a rise and fall.

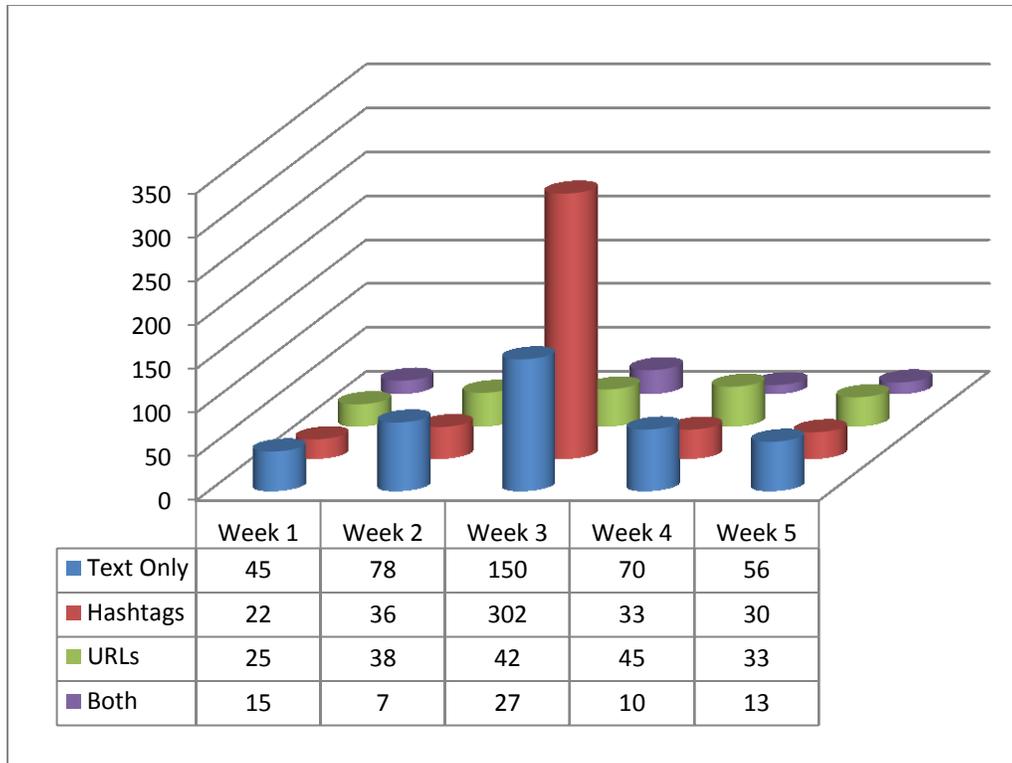


Figure 43 Group 2 - Part-timer Content

For Group 3 - Active, the week of the conference marks a large increase in all types of tweets, as demonstrated in Figure 45. The week of the conference, the largest category of tweets was the use of hashtags. After the conference, for all but one category of tweets, tweeting content returns to normal. For this group, the use of hashtags appears to need a recovery period after the conference; the number of tweets with a hashtag decreased from 655 to 45. However, by Week 5, the use of hashtags had again returned to 101 tweets.

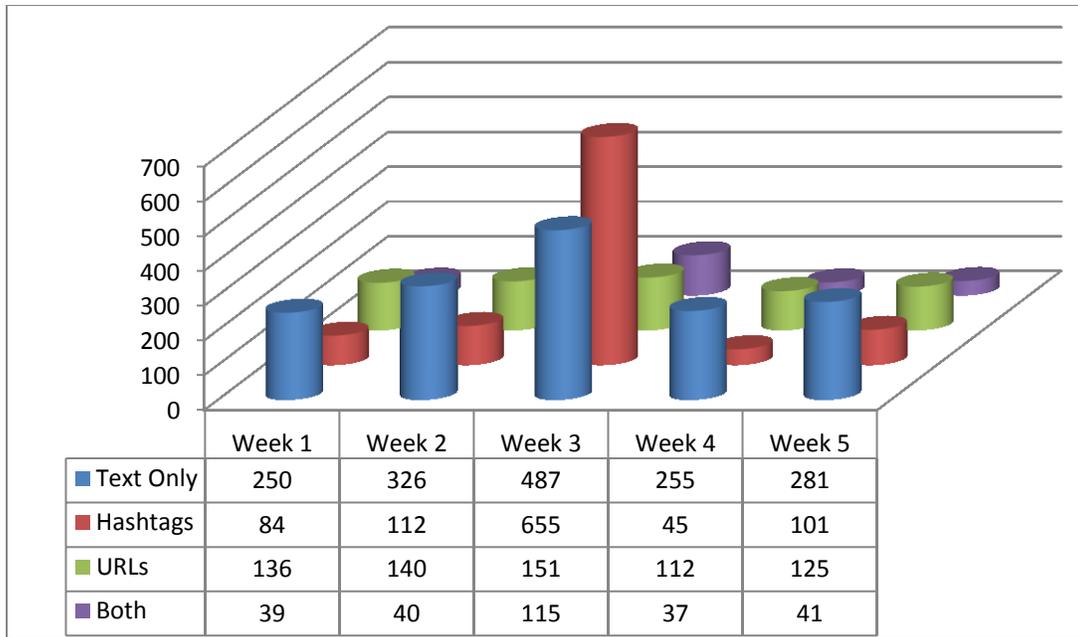


Figure 44 Group 3 - Active Content

For Group 4 - Immersed, there was an increase in all types of tweets the week of the conference, as demonstrated in Figure 45. However, for this group, during the week of the conference the largest category of tweets was the use of text only tweets. Close behind was the use of hashtags, which has the highest ratio of change the week of the conference. Another difference between this group and the others was the general ratio of text-only tweets across all weeks to any other category (with the exception of Week 3). In general, this group had more text-only tweets than any other group.

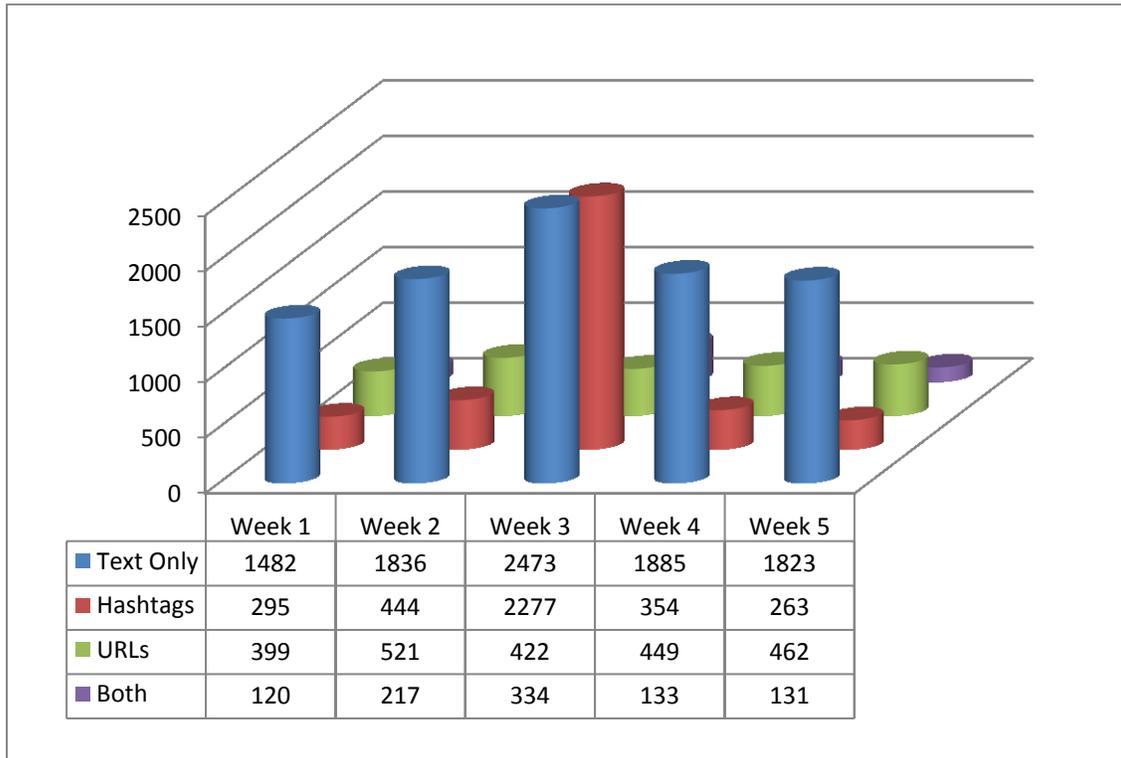


Figure 45 Group 4 - Immersed Content

4.3.2 Hashtags

Hashtag is the word for the symbol “#” in Twitter. Tweeters use it to denote keywords or topics in a tweet. Although organically created by the Twitter community as a way to find conversations about a thread, Twitter does not currently fully support it.

Currently, when clicking on a hashtag, Twitter will show you all the other tweets with that hashtag. If a hashtag becomes popular it will show up in the Trending Topics section of the website.

4.3.2.1 Group 1 - Experimenter

Group 1 - Experimenter used 80 hashtags or used hashtags in 32% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Hashtags	3	5	60	8	4	80	<i>total</i>
# Tweets	29	25	97	39	31	221	<i>total</i>
% of Total	59%	28%	27%	33%	13%	32%	<i>average</i>

4.3.2.2 Group 2 - Part-timer

Group 2 - Part-timer used 671 hashtags or used hashtags in 44% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Hashtags	83	66	419	55	48	671	<i>total</i>
# Tweets	107	159	521	158	132	1077	<i>total</i>
% of Total	43%	57%	34%	55%	31%	44%	<i>average</i>

4.3.2.3 Group 3- Active

Group 3 - Active used 1655 hashtags or used hashtags in 62% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Hashtags	174	214	994	101	172	1655	<i>total</i>
# Tweets	509	618	1408	449	548	3532	<i>total</i>
% of Total	65%	63%	56%	61%	63%	62%	<i>average</i>

4.3.2.4 Group 4- Immersed

Group 4 - Immersed used 6148 hashtags or used hashtags in 84% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Hashtags	553	883	3610	605	497	6148	<i>total</i>
# Tweets	2296	3018	5506	2821	2679	16320	<i>total</i>
% of Total	85%	94%	68%	83%	88%	84%	<i>average</i>

4.3.2.5 Top Hashtags

There were numerous hashtags used by the participants in this study during the 5 week period. As demonstrated by examining the content of Table 4, participants used hashtags to highlight events (e.g., “#chi2011” is the hashtag for the conference “CHI” held in 2011). The most used hashtag across all weeks was “#ghc10,” which was the hashtag to demark tweets relevant to the conference studied.

A second important point to call out from Table 4 is the use of official hashtags. The official hashtags have a background that is grey in the table. There were times when participants used the official hashtags less that emergent hashtags, e.g., “#techwomen.”

Table 4 Top Hashtags by Week. The items in bold italics and have a grey background are hashtags listed on the Grace Hopper Conference websites. You can find this list on Table 1 and Table 2 of this paper

All Weeks		Week1		Week2	
#ghc10	3532	#ghc10	143	#ghc10	247
#GHC10	206	#akoyaku	24	#promo	40
#career	115	#chi2011	12	#07VIP	32
#fb	90	#fb	12	#typeamom	29
#ldrshp	77	#countdown	11	#countdown	26

#osrc	68	#w00t	11	#GHC10	26
#hci	68	#jw10	10	#chi2011	20
#aiide	55	#techwomen	9	#fb	17
#promo	52	#IntuitCloudJam	8	#FF	15
#ghc	51	#wwt	8	#sorteio	12
#k12	48	#cio	7	#w00t	10
#tech	46	#cainc	7	#promoÃ§Ã£o	10
#acad	45	#NewTwitter	7	#Sorteio	9
#indst	44	#WWT	7	#PROMO	8
#chi2011	38	#SOBCon	7	#Intuit	8
#countdown	37	#ukrc	7	#HP	7
#FF	34	#HPPrint	7	#glee	6
#sorteio	33	#boulderfire	6	#Twitter	6
#award	33	#GHC10	6	#1LivroPorDia	5
#careers	32	#phd	5	#SORTEIO	5
#07VIP	32	#BIF6	5	#switc10	5
#grad	31	#cloud	5	#ghc	5
#techwomen	29	#UWMadEngr	5	#ultrafinoalways	5
#ghc11	29	#sobcon	5	#jobs	4
#typeamom	29	#aosocal	4	#concorra	4

Week3		Week4		Week5	
#ghc10	3032	#ghc10	87	#aiide	28
#GHC10	172	#kokuminID	28	#fb	24
#career	114	#aiide	27	#innowe10	24
#ldrshp	76	#mountainrb	19	#ghc10	23
#hci	68	#fb	16	#int3v2	23
#osrc	62	#fmgt	14	#wcre	22
#acad	45	#SFTH	13	#cialetras10mil	18
#k12	45	#snwusa	12	#FF	10
#tech	44	#liesyoushouldntfallfor	10	#Rescue	9
#indst	44	#blogherfood	9	#PequenaAbelha	8
#ghc	43	#NewTwitter	8	#HP	7
#award	33	#uist2010	7	#Chilean	7
#grad	31	#techwomen	6	#sorteio	6
#careers	26	#BlogHerFood	6	#Kuat	6
#ghc11	25	#in	5	#rescatemineros	5
#keynote	25	#theyhatin	5	#nyctweet	5
#awards	24	#Blogherfood	5	#osrc	5
#academic	23	#hol	5	#SextaModerna	5
#exec	22	#bhf10	5	#WCRE	5
#fb	21	#ImagineCup	4	#AATYC	5
#invtech	21	#gradschool	4	#glee	4

#phd3	19	#Yelp	4	#in	4
#bof3	18	#basec	4	#beatcancer	4
#ghc2010	15	#plan	4	#BeatCancer	4
#sorteio	13	#hp	4	#techwomen	4

I found the following hashtags to have specific meaning. The meanings were found by searching Twitter and then the internet.

- #fb - Tweets ending in #fb are imported to Facebook if the user has installed the Selective Twitter Update application on Facebook. All other tweets will not be imported.
- #ff - Used on Fridays to suggest people to follow to your followers and to whoever is tuned in to the hashtag. Shortened from #followfriday.
- #ghc – Grace Hopper Conference
- #chi2011 - The 2011 ACM CHI Conference on Human Factors in Computing Systems is the premier international conference of human-computer interaction³².
- #techwomen - TechWomen is a professional mentorship and exchange program that will identify 38 women who are emerging leaders in technical fields in Muslim communities, and bring them to the United States in summer 2011 for a five-week, project-based mentoring program at leading technology companies in Silicon Valley³³.
- #ghc11 – Grace Hopper Conference 2011
- #jw10 – Jive World 2010
- #IntuitCloudJam - A 3-day competition for student developers to create and showcase innovative new products using any technology they like. Judges will select the project that creates the most compelling use of the Intuit Partner Platform and Intuit Data Services³⁴.
- #wwt - Women Who Tech TeleSummit
- #typeamom - Type-A Parent Conference, in its third year, is designed to help social media/blogging parents and the companies who want to connect with them take it to the next level.³⁵
- #aiide - AI and Interactive Digital Entertainment Conference
- #cainc - CA Technologies
- #HPPrint – HP ePrint
- #NewTwitter – The revamped Twitter website, announced September 15th 2010.
- #SOBCon - The premier conference for successful and outstanding bloggers hosted by Liz Strauss and Terry Starbucker.³⁶

³² <http://www.chi2011.org/>

³³ <http://www.techwomen.org/faq/>

³⁴ <http://www.facebook.com/event.php?eid=141885569182919>

³⁵ <http://typeaconference.com/about/>

- #ukrc - The UKRC is the UK Government's lead organisation for the provision of advice, services and policy consultation regarding the under-representation of women in science, engineering, technology and the built environment (SET).
- #boulderfire - A wildfire that burned outside of Boulder, Colorado.
- #aosocal - Agile Open Southern California 2010
- #Intuit - Intuit
- #HP -Hewlett-Packard
- #glee – Glee is a musical comedy-drama television series that airs on Fox in the United States.³⁷
- #Twitter – Twitter
- #switc10 - Indiana University 2010 Statewide IT Conference
- #ghc2010 – Unofficial Grace Hopper Conference 2010 hashtag
- #careers - likely meaning the 'career' track, but divergent from that hashtag by the inclusion of the 's'
- #keynote - likely applying to the keynote address that took place between 8:45AM and 9:45AM on September 30th
- #awards - likely applying to the awards ceremony that took place at 7PM on September 30th
- #academic - likely meaning the 'acad' track, but divergent from that hashtag by the inclusion of the 'emic'
- #sorteio - portuguese for drawing
- #mountainrb - A single track conference for people who program in ruby
- #fmgt - likely applying to the HCI field trip to Georgia Tech titled the "Future of Media"
- #snwusa - a hashtag for the SNW conference which is on storage, data center, and IT infrastructure
- #bloggerfood - Blogger is a website for women who blog, this tag is for a conference for bloggers who write about food
- #uist2010 - An ACM Symposium on User Interface Software and Technology, in particular in the year 2010
- #theyhatein - likely meaning that people are being mean or disrespectful
- #bhf10 - possibly referring to a music event in Brooklyn
- #ImagineCup - an event hosted by Microsoft for students studying technology
- #gradschool - likely meaning the 'grad' track, but divergent from that hashtag by the inclusion of the 'school'
- #Yelp - a website that provides user generated reviews of local establishments

³⁶ <http://www.sobevent.com/colorado-2010/>

³⁷ http://en.wikipedia.org/wiki/Glee_%28TV_series%29

- #BeatCancer - an social networking event in October 2011 to make the hashtag 'beatcancer' the most widespread social networking message run through Livestrong
- #WCRE - likely meaning the working conference on reverse engineering
- #AATYC - likely meaning the Arkansas Association of Two Year Colleges
- #Chilean -- likely referring to the miners who were trapped underground for most of October 2010
- #rescatemineros - likely referring to the miners who were trapped underground for most of October 2010
- #Rescue - likely referring to the miners who were trapped underground for most of October 2010
- #PequenaAbelha - Spanish for "little bee" a book by Chris Cleave
- #int3v2 - possibly referring to some event on gaming on October 14th
- #innowe10 - likely referring to an event called Innovation Weekend

4.3.2.6 GHC Hashtags

The use of the “#ghc10” hashtag varied by group and week, as demonstrated by Table 5 and Table 6. First, the greatest increase occurred in Week 3. This change is an order of magnitude difference for all groups. This is obvious given that this hashtag is the indicator hashtag for this conference. Second, Group 4 – Immersed used the hashtag more than any other group across all weeks.

The last related finding is that there were 1,444 tweets from users who were not in the list created by the Grace Hopper Tweeter. We found 485 users who tweeted with a Grace Hopper official hashtags that were not part of the Twitter list. These could be people that did not add themselves to the official list or were people that could have been tweeting about the conference remotely.

Table 5 Number of Tweets found with a Grace Hopper hashtag

	Week 1	Week 2	Week 3	Week 4	Week 5	Total
Group 1 - Experimenter	1	2	43	0	0	46
Group 2- Part-timer	3	16	248	1	0	268
Group 3- Active	17	68	685	15	2	787
Group 4- Immersed	117	202	2240	72	24	2655
Unknown	43	108	1230	44	19	1444
Grand Total	181	396	4446	132	45	5200

Table 6 Percentage of all Tweets with a GHC hashtag

	Week 1	Week 2	Week 3	Week 4	Week 5	Total
Group 1- Experimenter	3%	8%	44%	0%	0%	21%
Group 2- Part-timer	3%	10%	48%	1%	0%	25%
Group 3- Active	3%	11%	49%	3%	0%	22%
Group 4- Immersed	5%	7%	41%	3%	1%	16%

4.3.2.7 Summary

In all there were a total of 8,554 tweets with hashtags made by the Twitter users of this study. Hashtags were used in 40% of all tweets in this study.

As seen in Figure 46, there is an increase for all groups of users the week of the conference. However, Group 4 had the largest increase in the use of hashtags. Figure 47 demonstrates the same phenomenon. It also demonstrates that in Week 3 there is a marked dip in percentages. This is due to the fact that the number of updates, in general, increased during this week.

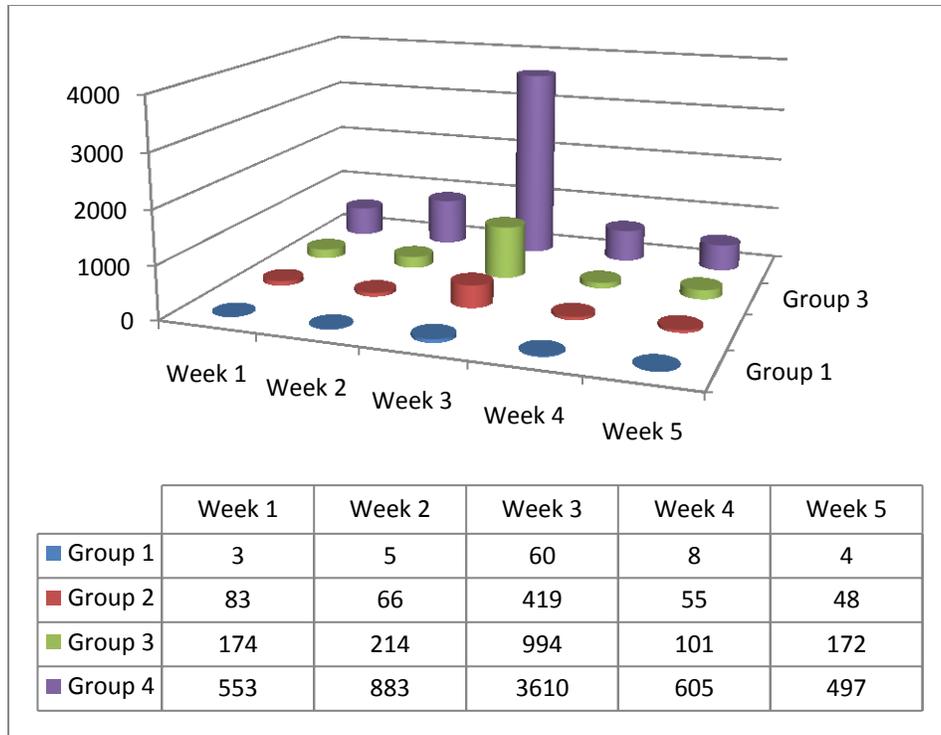


Figure 46 Number of hashtags

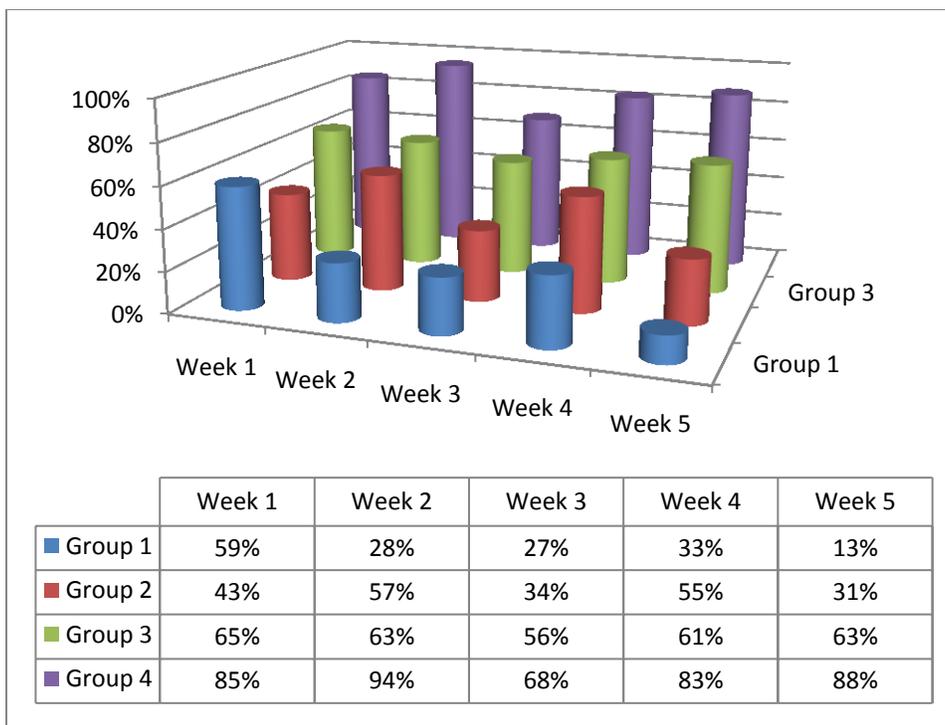


Figure 47 Percentages of hashtags

4.3.3 URLs

A tweet can contain any number of URLs. For the purposes of this study we count all URLs even if there is more than one per status update. Many of the URLs in Twitter are shortened using a URL shortening service.

4.3.3.1 Group 1- Experimenter

Group 1 - Experimenter used URLs 71 times or used URLs in 37% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# URLs	13	7	19	13	19	71	<i>total</i>
# Tweets	29	25	97	39	31	221	<i>total</i>
% of Total	45%	28%	20%	33%	61%	37%	<i>average</i>

4.3.3.2 Group 2- Part-timer

Group 2 - Part-timer used URLs 262 times or used URLs in 30% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# URLs	41	47	72	56	46	262	<i>total</i>
# Tweets	107	159	521	158	132	1077	<i>total</i>
% of Total	38%	30%	14%	35%	35%	30%	<i>average</i>

4.3.3.3 Group 3- Active

Group 3 - Active used URLs 958 times or used URLs in 30% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# URLs	180	185	275	151	167	958	<i>total</i>
# Tweets	509	618	1408	449	548	3532	<i>total</i>
% of Total	35%	30%	20%	34%	30%	30%	<i>average</i>

4.3.3.4 Group 4- Immersed

Group 4 - Immersed used URLs 3268 times or used URLs in 21% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# URLs	534	755	785	591	603	3268	<i>total</i>
# Tweets	2296	3018	5506	2821	2679	16320	<i>total</i>
% of Total	23%	25%	14%	21%	23%	21%	<i>average</i>

4.3.3.5 Summary

In all there were a total of 4,559 URLs used by the Twitter users of this study. URLs were used in 22% of all tweets in this study.

As seen in Figure 48, there is an increase for all groups of users the week of the conference. However, Group 4 had the largest increase in the use of URLs. Group 3 used URLs the second most, followed by Groups 2 and 1.

Figure 49 demonstrates the same phenomenon. It also demonstrates that in Week 3 there is a marked dip in percentages. This is due to the fact that the number of updates, in general, increased during this week.

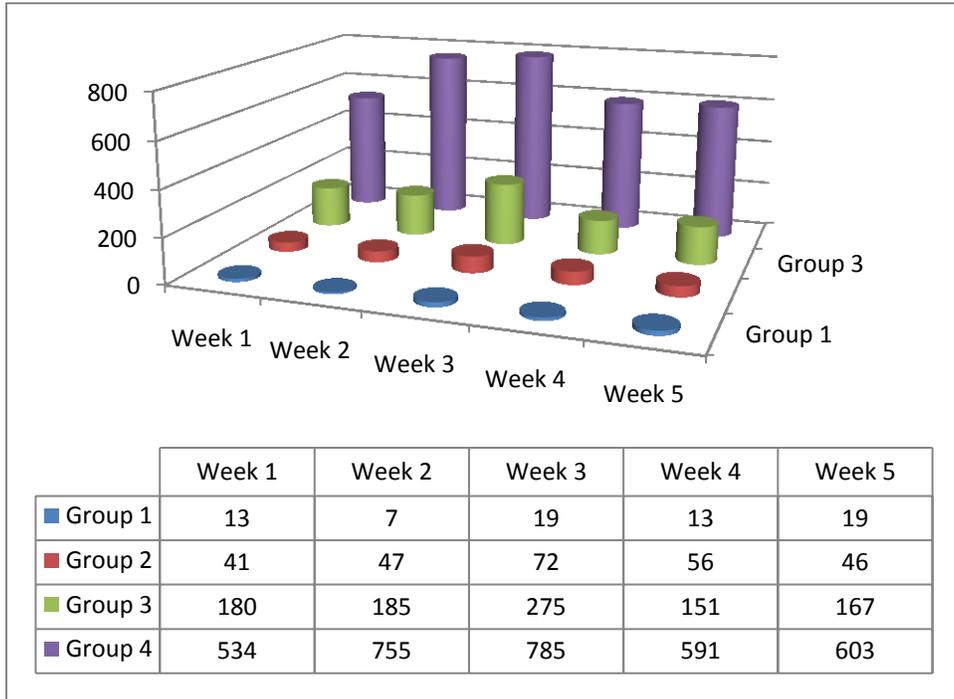


Figure 48 Number of URLs

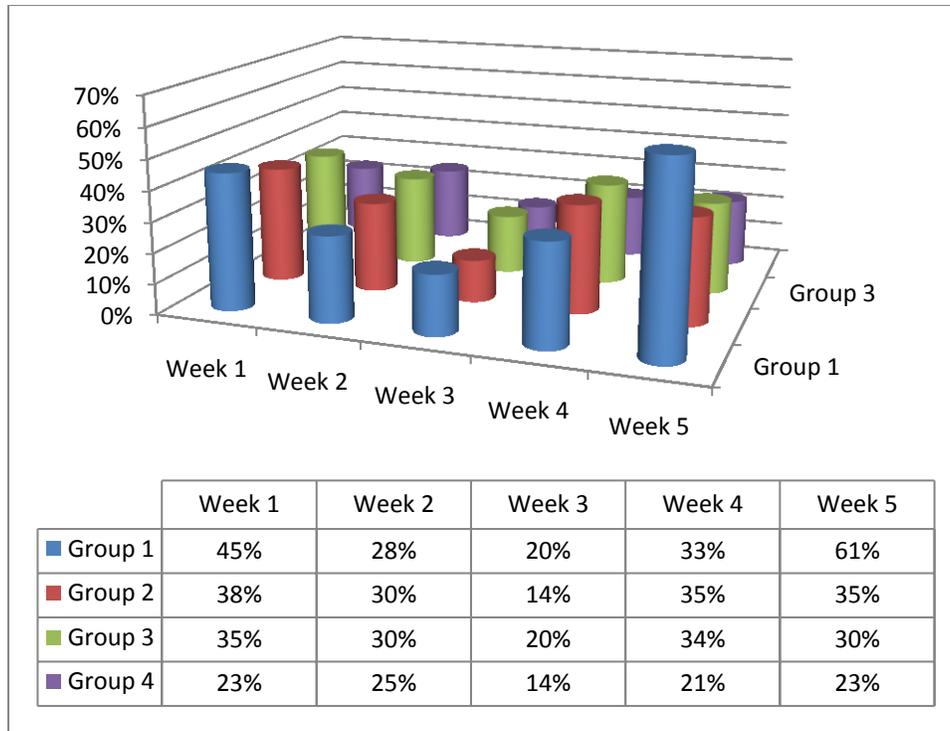


Figure 49 Percentages of URLs

4.3.4 Content Analysis

4.3.4.1 Word clouds

The researchers generated the following word clouds based on the text of the tweets per week for the entire group of participants. The application removed common English words and the researchers removed the word “RT.” All other Twitter text? has remained (e.g., as mentions and hashtags). The tool automatically removed the # character from hashtags in the analysis, but the words still appear. Only the top 150 words appear in the Word Cloud.

Each week has its own word cloud: Week 1 is Figure 50; Week 2 is Figure 51; Week 3 is Figure 52; Week 4 is Figure 53; and Week 5 is Figure 54.

There are a few key characteristics and trends across the word clouds. The first is that they all generally have the same shape. The second is the prominence of key words like “ghc10.” The third is the size of the words. Apart from Week 3, the size of the words tends to be fairly uniform. In Week 3 the word “GHC10” took over most of the figures real estate. Finally, there is a trend in the content of the word clouds. Words related to the conference start to appear and pick up prominence as the conference date approached (e.g., “grace,” “women,” and “@ghc”). However, by Week 5 almost all conference-related words are no longer present in the word cloud.

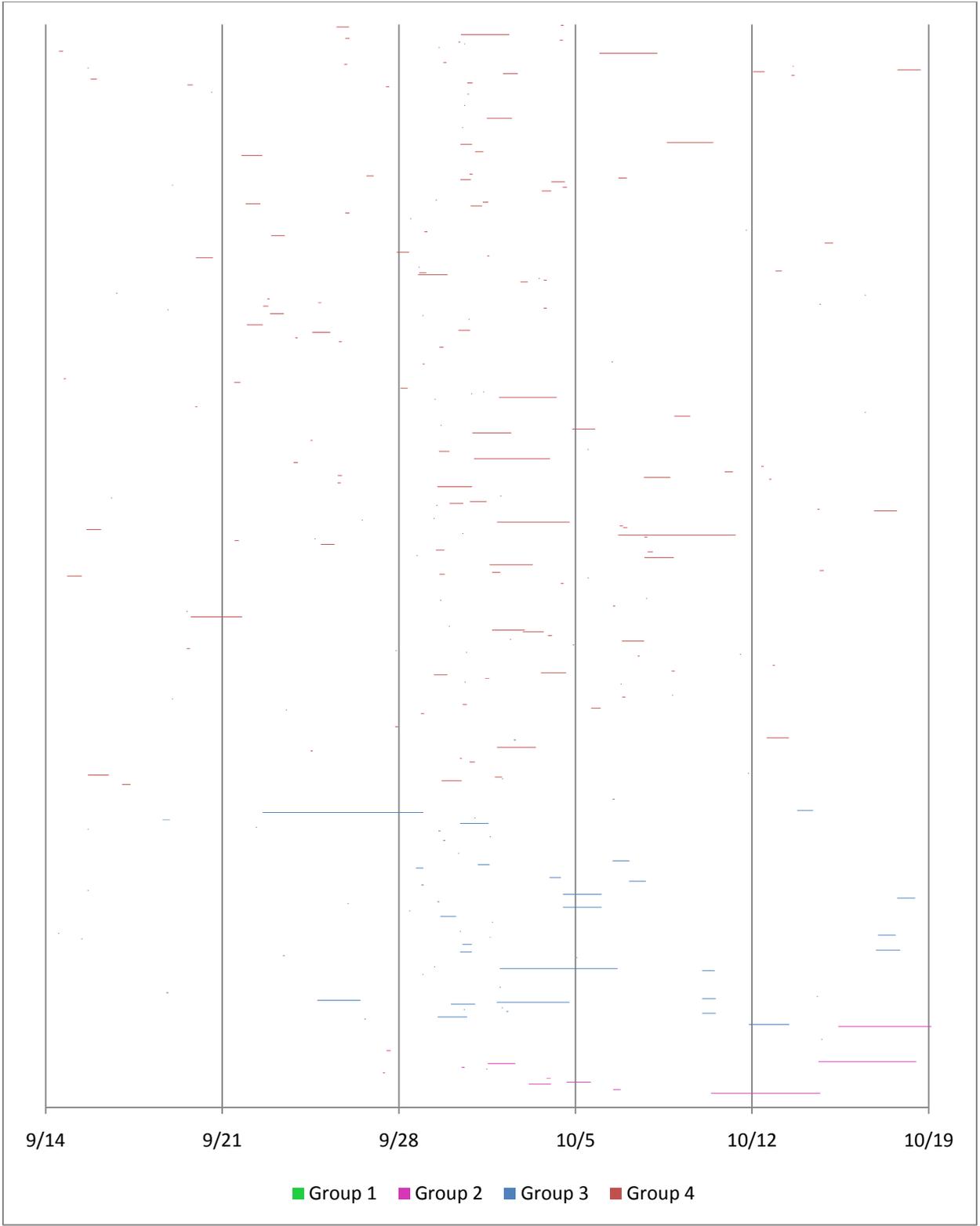


Figure 55 Graph of all conversations

These trends in the conversations are additionally presented in Figure 56, which breaks them down by Group and Week. This table demonstrates that Group 4 was the most active in conversations across all weeks. Groups 2 and 4 participated in conversations primarily in Week 3. Group 1 only participated in conversations during Week 3.

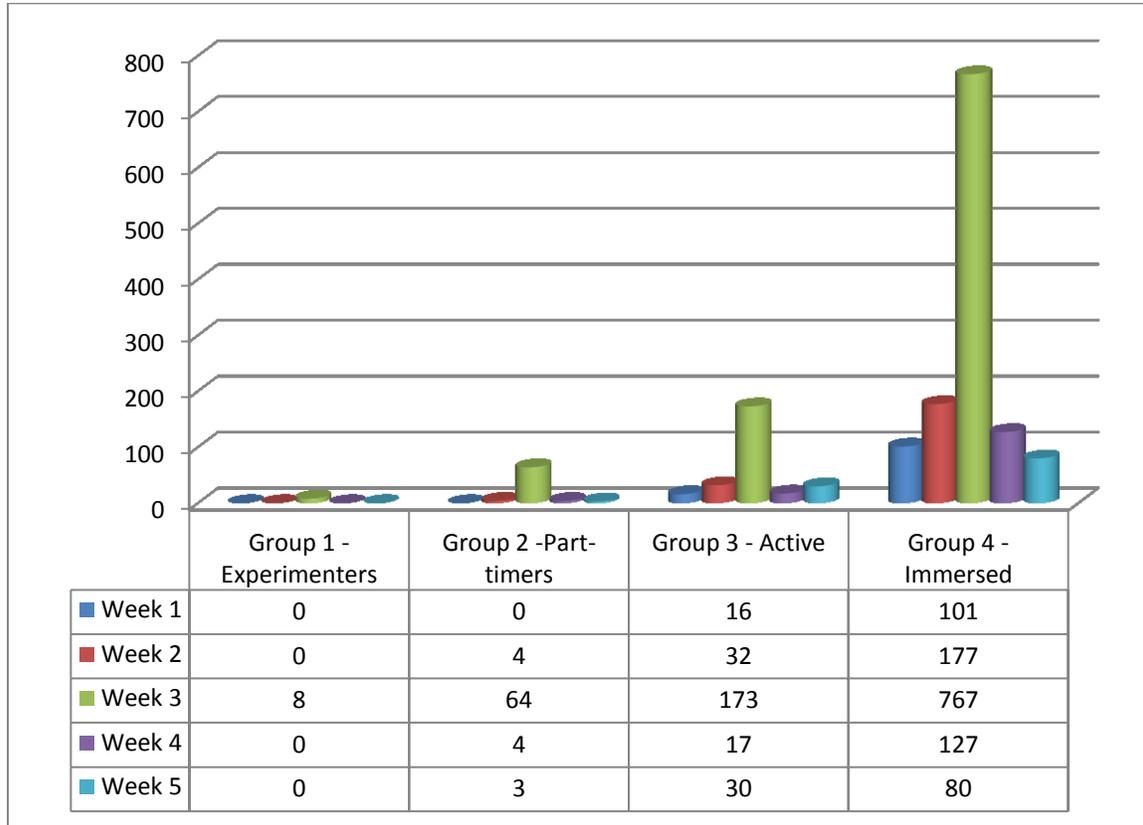


Figure 56 Piece of Conversation by Week and Group

The bulk of conversations occurred within participants in Group 4, as saliently demonstrated in Figure 57, Figure 58, and Figure 59. There were 540 tweets that started conversations but that spawned 582 conversation threads.

78% of conversations took place between participants in Group 4. Additionally, participants in Group 4 started 71% of all conversations.

Conversations occurred between people, and not always people who belonged in the same group. For example, participants in Group 3 participated in more conversations than they started.

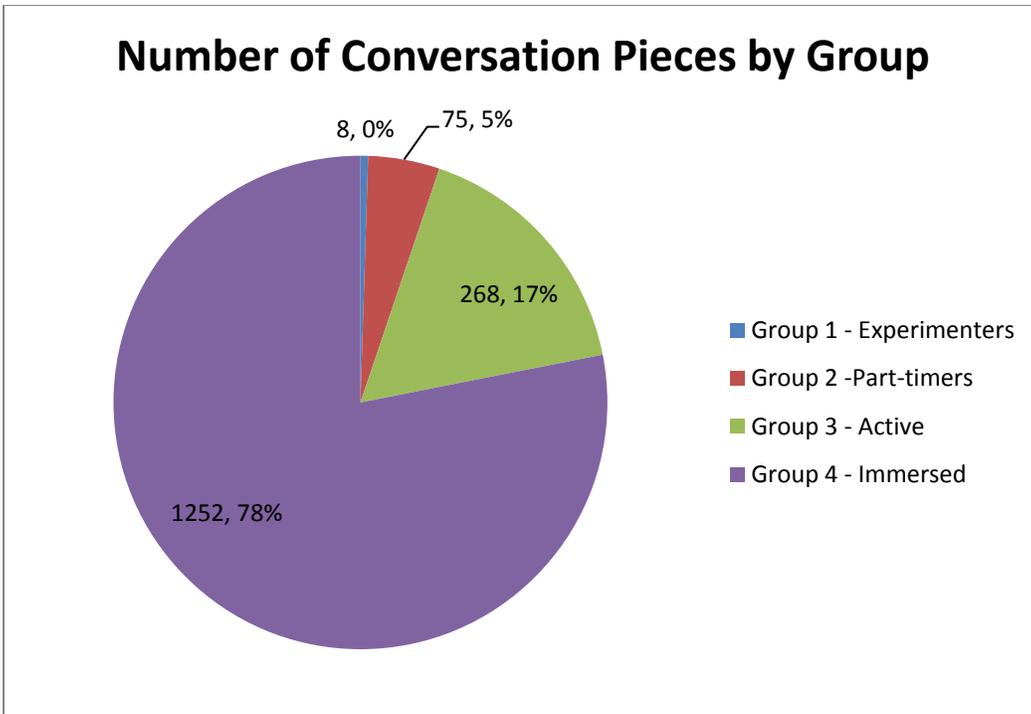


Figure 57 Number of Conversation Pieces by Group

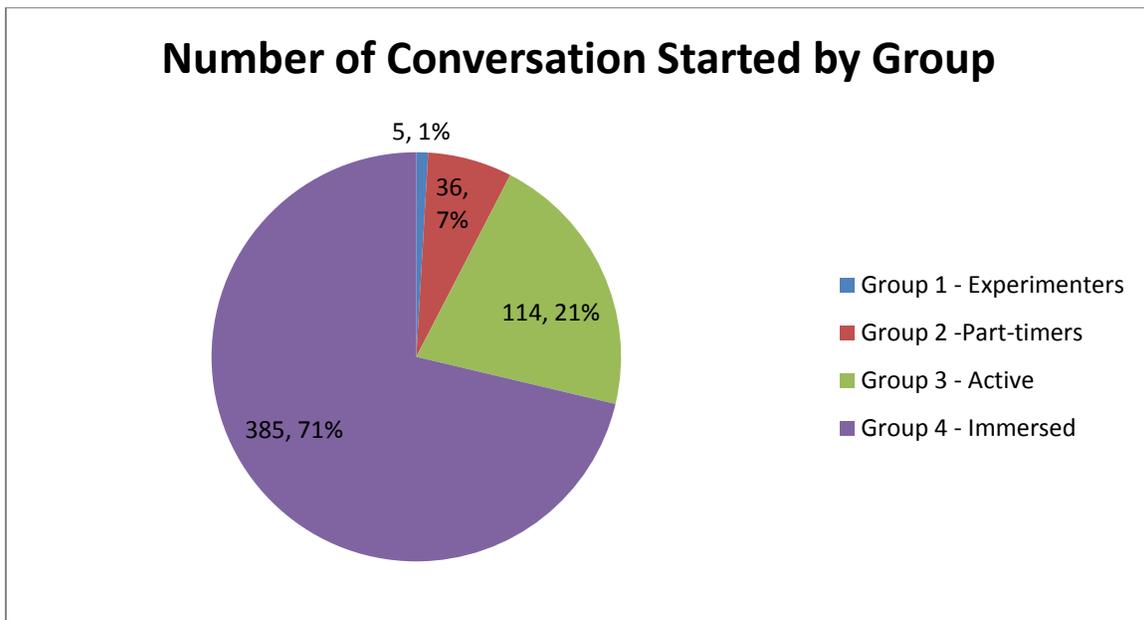


Figure 58 Number of Conversation Started by Group

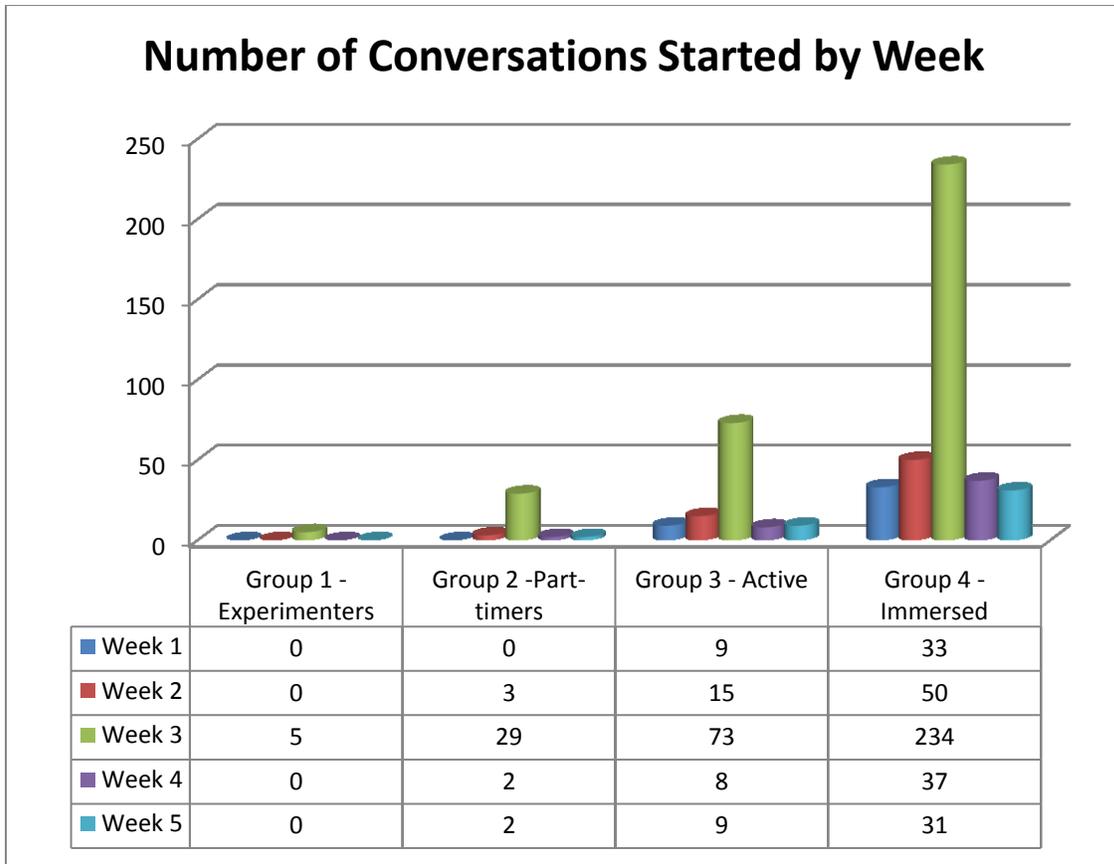


Figure 59 Number of Conversations Started by Week

There were only 540 unique tweets that started conversations. As demonstrated in Figure 60, Figure 61, Figure 62, and Figure 63, activity was high during the week of the conference. For Group 1, they only tweeted during week 3. For Group 2, participants did not use; participants used hashtags the most, and participants used mentions the most of the weeks. For Group 3, participants used hashtags the most; and participants used retweets to start a conversation once. For Group 4, participants used hashtags the most; and, the use of mentions and hashtags by participants was the reason for most of the conversations started in the weeks prior to the conference.

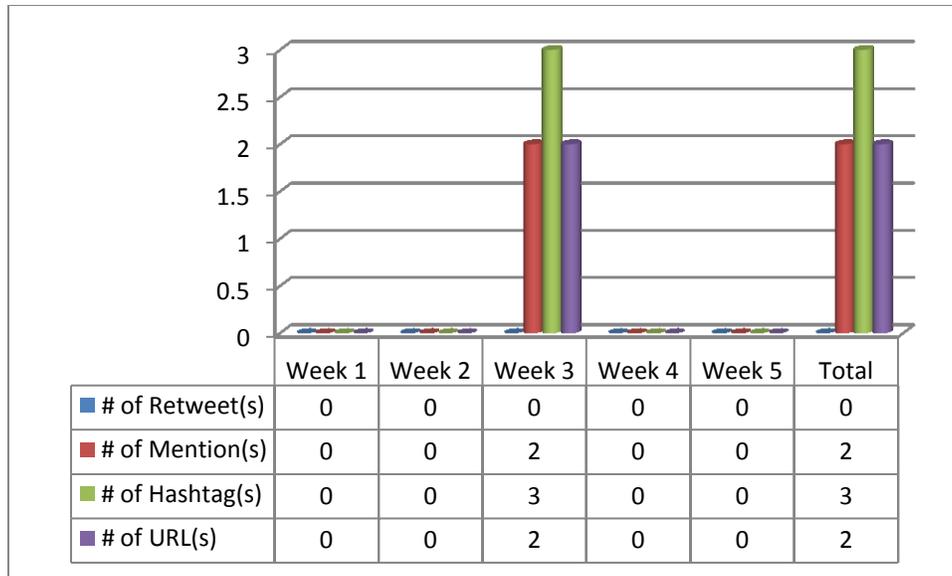


Figure 60 Group 1 - Experimenter Unique Tweets that started Conversation by Week

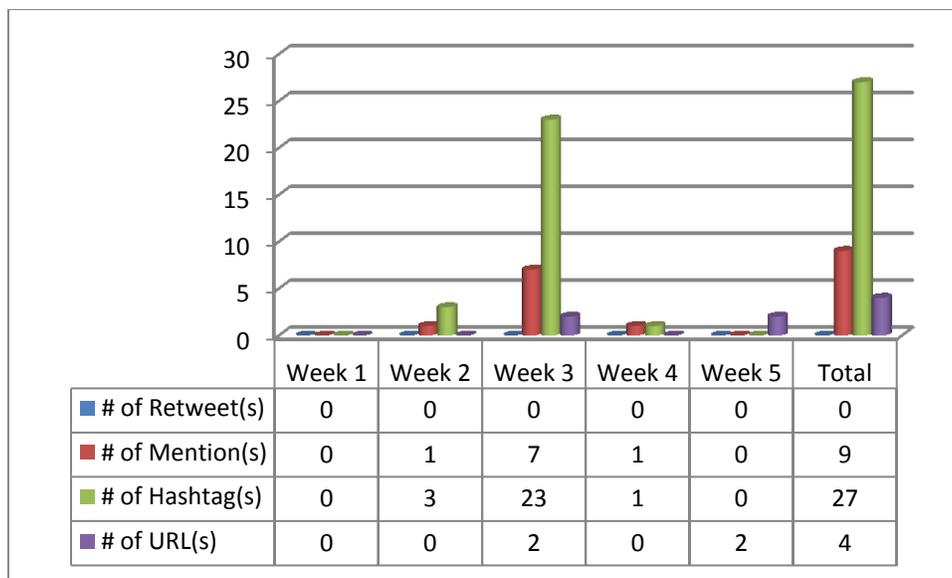


Figure 61 Group 2 - Part-timer Unique Tweets that started Conversation by Week

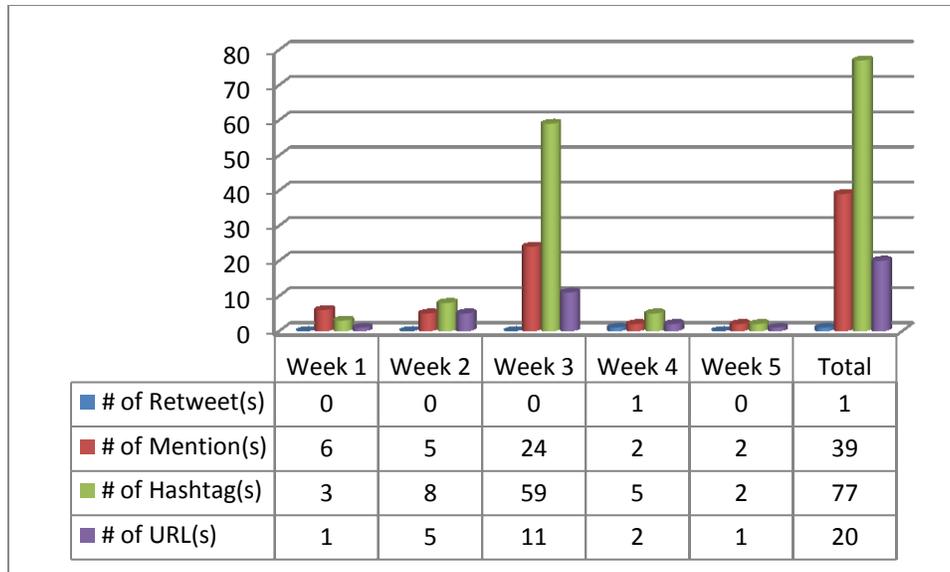


Figure 62 Group 3 - Active Unique Tweets that started Conversation by Week

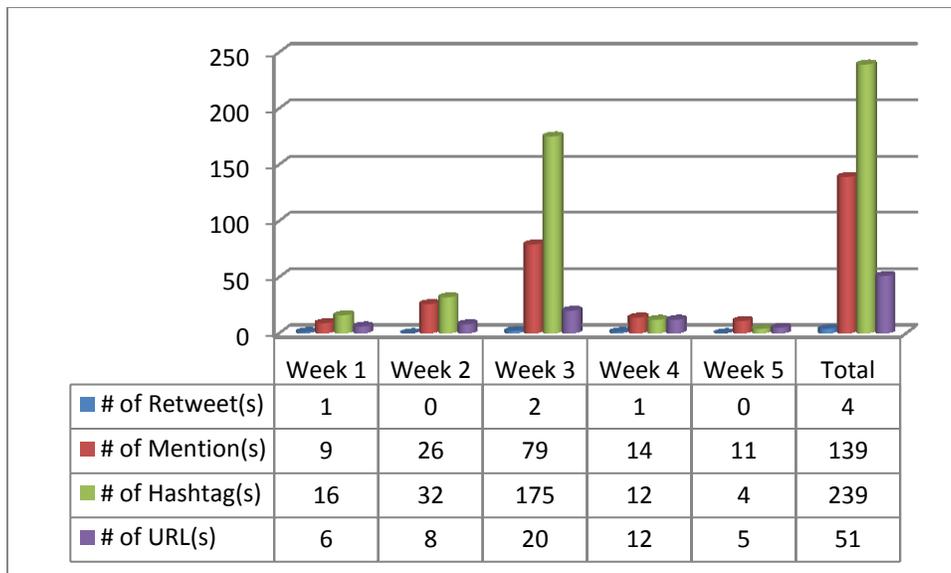


Figure 63 Group 4 - Immersed Unique Tweets that started Conversation by Week

There were 1,063 tweets that were part of conversations to original tweets. As demonstrated in Figure 64, Figure 65, Figure 66, Figure 67, Figure 68, and Figure 69, Group 4 had the most conversation tweets and conversations occurred the most during the week of the conference.

Figure 64 demonstrates that Group 4 made 82% of all conversations replies. Group three made 15% of all replies. Group 2 made three percent of the replies. Finally, Group 1 made so few replies during any week of the conference that it appears as though no one in Group 1 tweeted.

Conversations to Original Tweets by Group

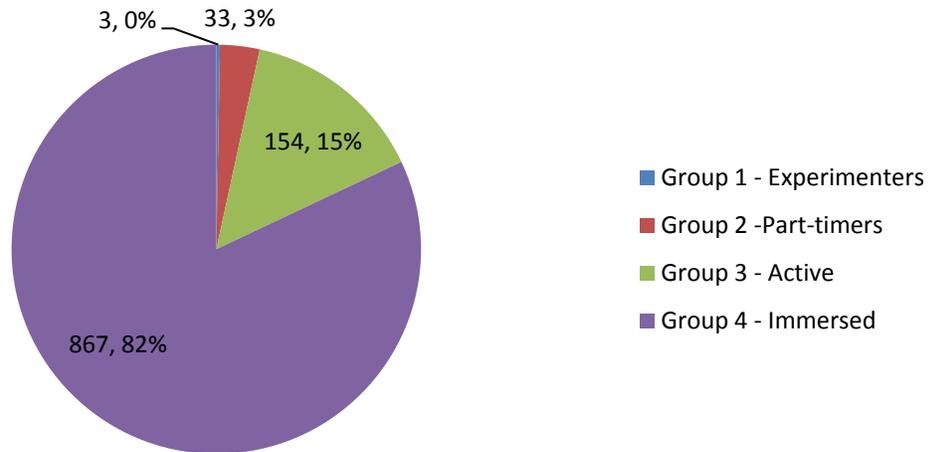


Figure 64 Conversations to Original Tweets by Group

Figure 65 demonstrates two interesting factors. The first is that, again, Group 4 had the most conversations. Second, Week 3 has more conversations than any other week.

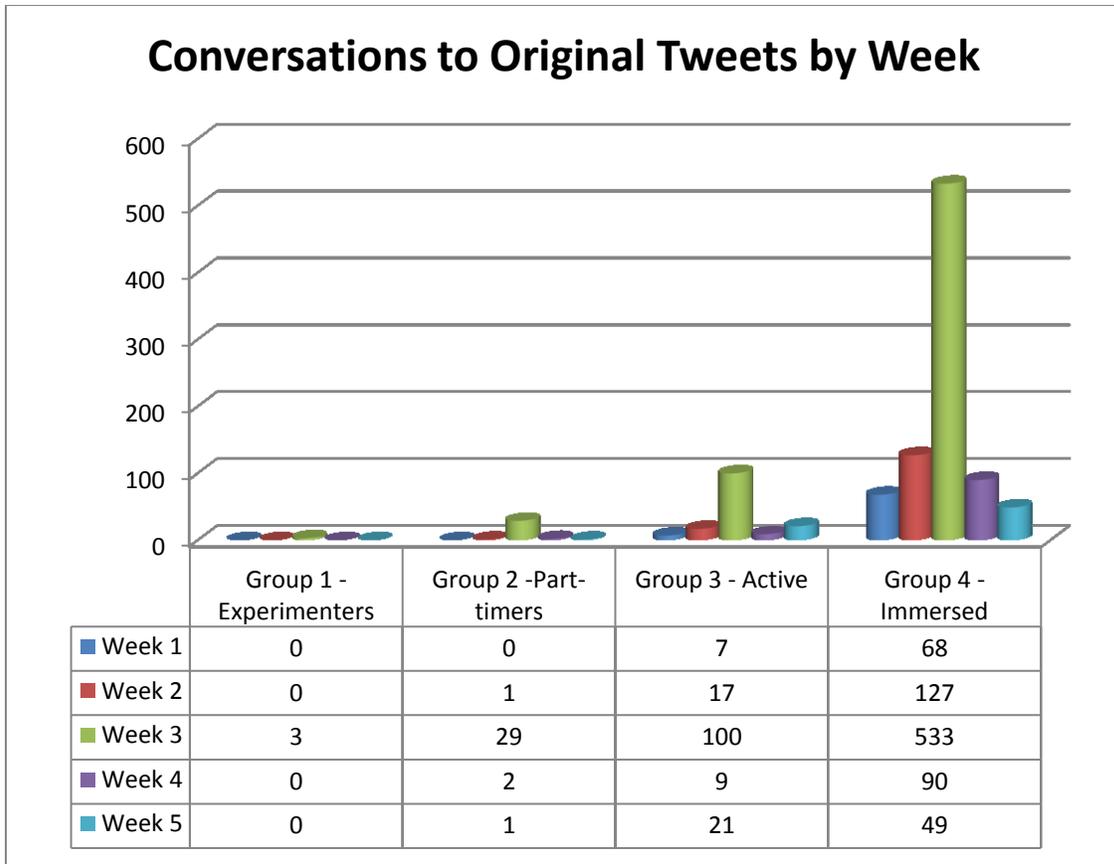


Figure 65 Conversations to Original Tweets by Week

When concentrating only on Group 1 – Experimenter, Figure 66 demonstrates that there were only three conversations during the entire span of the study. These conversations were mentions. This could mean that these tweets were responses to other tweets.

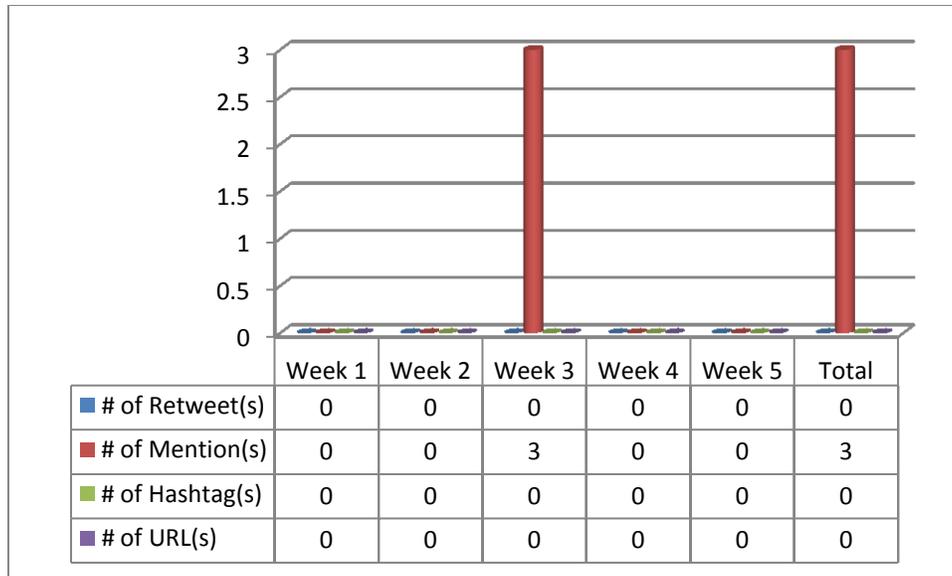


Figure 66 Group 1 - Experimenter Participation in Conversations

When concentrating only on Group 2 – Part-timer, Figure 67 demonstrates there were thirty-seven conversations. Most of these occurred in Week 3 and used hashtags. The low number of mentions in this group could imply that hashtags were the preferred way of continuing a conversation. A last point in reference to this figure is that URLs only occurred twice, and this group did not use any retweets.

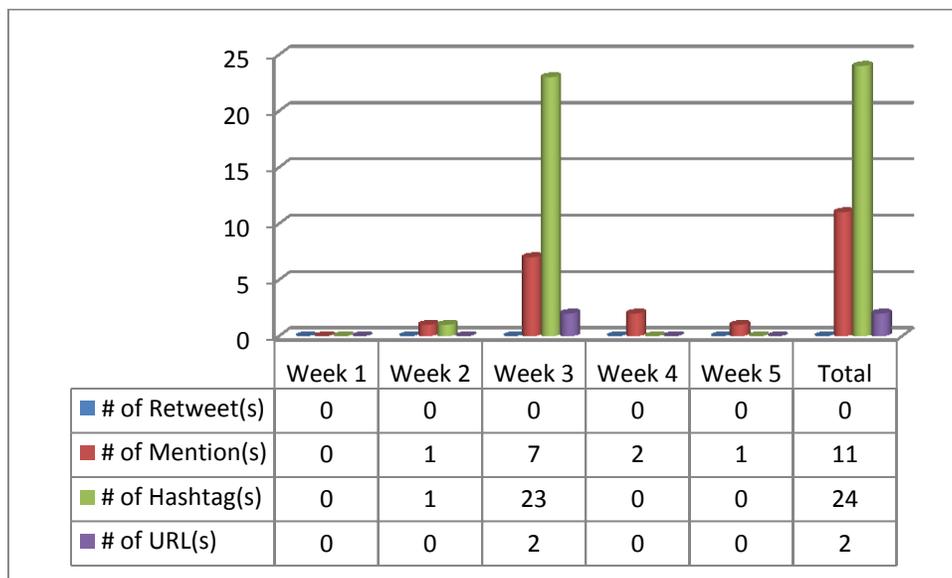


Figure 67 Group 2 - Part-timer Participation in Conversations

When concentrating only on Group 3 – Active, Figure 68 demonstrates that there were 184 conversations by members of this group. For this group there was again an increase in Week 3.

However, across all weeks there was a marked increase in the use of mentions over any other conversation mechanism. This implies that the users were directly tweeting back and used the shared context to answer questions or continue conversations. Finally, this group, again, did not use any retweets.

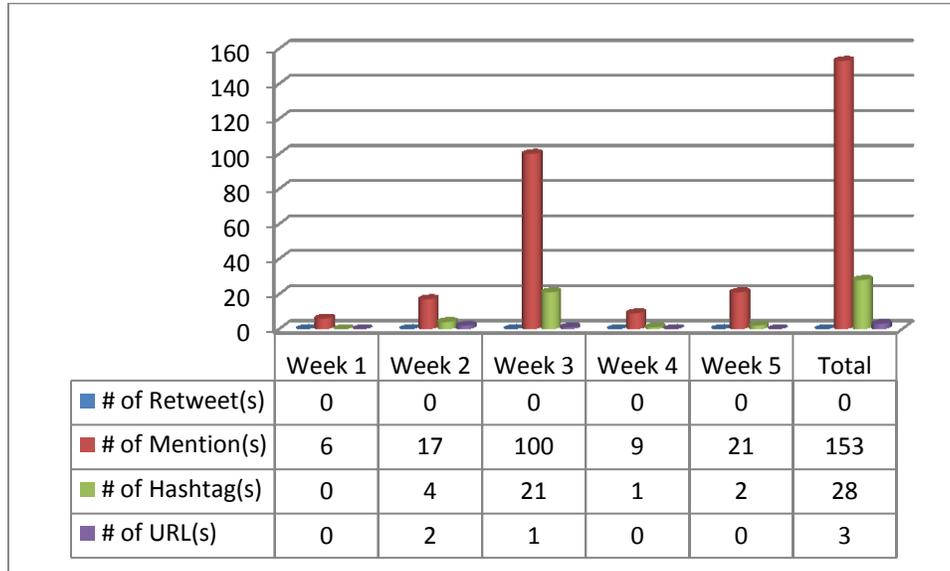


Figure 68 Group 3 - Active Participation in Conversations

When concentrating only on Group 4 – Immersed, Figure 69 demonstrates that there were 1075 conversations by members of this group. For this group there was again an increase in Week 3. However, across all weeks there was a marked increase in the use of mentions over any other conversation mechanism. This implies that the users were directly tweeting back and used the shared context to answer questions or continue conversations. Finally, this group, again, did not use any retweets.

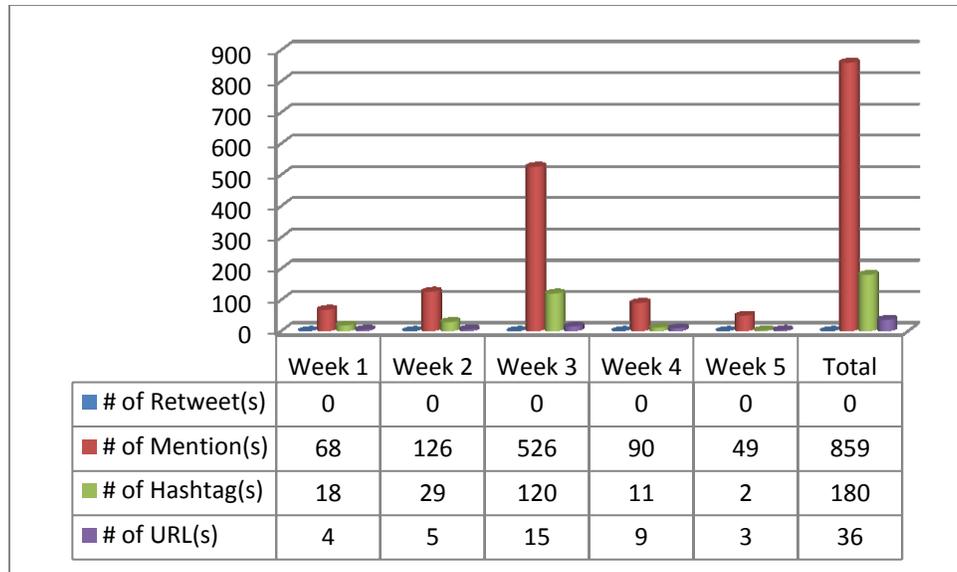


Figure 69 Group 4 - Immersed Participation in Conversations

Table 7 demonstrates a number of trends that explain the conversations of the participants. First, the average length of a conversation was 2 statements and then a reply. Second, participants completed their conversation from beginning to end tended in less than 7 hours. This trend held until Weeks 4 and 5. There is an anomaly on this table. For Group 2, the average conversation length jumped from the single digits to 136 hours. The type of user, who infrequently checked their messages, might explain this.

Table 7 Conversation Statistics

Group		Week 1	Week 2	Week 3	Week 4	Week 5	Total
1 - Experimenter		0	0	5	0	0	5
	Avg Conversation Length (hours)	0	0	0.5	0	0	0.5
	Max Conversation Length (hours)	0	0	0.9	0	0	0.9
	Min Conversation Length (minutes)	0	0	3	0	0	3.0
	Avg Conversation Length	0	0	2			
2 - Part-timer		0	3	31	1	4	39.0
	Avg Conversation Length (hours)	0	2.4	2.9	7.7	136.9	149.9
	Max Conversation Length (hours)	0	4.5	26.6	7.7	348.8	387.6
	Min Conversation Length (minutes)	0	32	1	463	79	575.0
	Avg Conversation Length	0	2	2	2	2	8
3 - Active		11	15	78	11	10	125.0
	Avg Conversation Length (hours)	1.6	3.5	6.7	13.9	11.6	37.3
	Max Conversation Length (hours)	7.1	41.7	153.6	37.2	38.9	278.5
	Min Conversation Length (minutes)	7	1	1	5	20	34.0

	Avg Conversation Length	2	2	2	2	2	10
4 - Immersed		32	57	248	42	34	413.0
	Avg Conversation Length (hours)	5.1	3.6	3.7	9.5	3.6	25.5
	Max Conversation Length (hours)	49.7	20.7	72.4	112.8	22.6	278.2
	Min Conversation Length (minutes)	1	0	0	1	1	3.0
	Avg Conversation Length	2	2	2	3	2	
Total		43	75	362	54	48	582

Figure 70, provided below, lists the top 25 users by their betweenness centrality scores. These scores are based on conversations alone. This table demonstrates that conversations do change the betweenness rankings of individuals at conferences. It also demonstrates that 36% of the group is the same when examining friends and followers.

Username	Betweenness Centrality
aengleson	636
anitaborg_org	728
bubbva	761
casadevega	340
corey_latislaw	1925
creepyed	1133
gailcarmichael	1147
ghc	5191
gillianmsmith	292
jennifer_el	285
kpanciera	451
latinasinc	330
lexyholloway	640
miss_amantha	372
mitsuba3	305
mystitat	262
nancymar	337
robert_a_walker	281
selenamarie	290
sgcxb	248
terriko	273
travelingsw	752
valeriej	331
wordymusic	3728
Yardi	585

4.4.2 Mentions

A mention is any tweet that contains @username in the body of the status update. Replies are a type of mention, but the reverse is not true. For the purposes of this study we count all mentions even if there is more than one per status update.

4.4.2.1 Group 1- Experimenter

Group 1 - Experimenter used 67 mentions or used mentions in 10% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Mentions	17	7	26	13	4	67	<i>total</i>
# Tweets	29	25	97	39	31	221	<i>total</i>
% of Total	10%	20%	62%	21%	13%	10%	<i>average</i>

4.4.2.2 Group 2- Part-timer

Group 2 - Part-timer used 443 mentions or used mentions in 54% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Mentions	46	90	179	87	41	443	<i>total</i>
# Tweets	107	159	521	158	132	1077	<i>total</i>
% of Total	78%	42%	80%	35%	36%	54%	<i>average</i>

4.4.2.3 Group 3- Active

Group 3 - Active used 2127 mentions or used mentions in 39% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Mentions	331	387	790	276	343	2127	<i>total</i>
# Tweets	509	618	1408	449	548	3532	<i>total</i>
% of Total	34%	35%	71%	22%	31%	39%	<i>average</i>

4.4.2.4 Group 4- Immersed

Group 4 - Immersed used 13272 mentions or used mentions in 32% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Mentions	1954	2851	3771	2347	2349	13272	<i>total</i>
# Tweets	2296	3018	5506	2821	2679	16320	<i>total</i>
% of Total	24%	29%	66%	21%	19%	32%	<i>average</i>

4.4.2.5 Summary

In all there were a total of 15,909 tweets with mentions made by the Twitter users of this study. This means that participants used mentions in 75% of all tweets in this study. Figure 72 and Figure 73 demonstrate this fact pictorially. Specifically, they show that Group 4 had the largest use of mentions. This could be because this is the group that starts and replies in conversations the most.

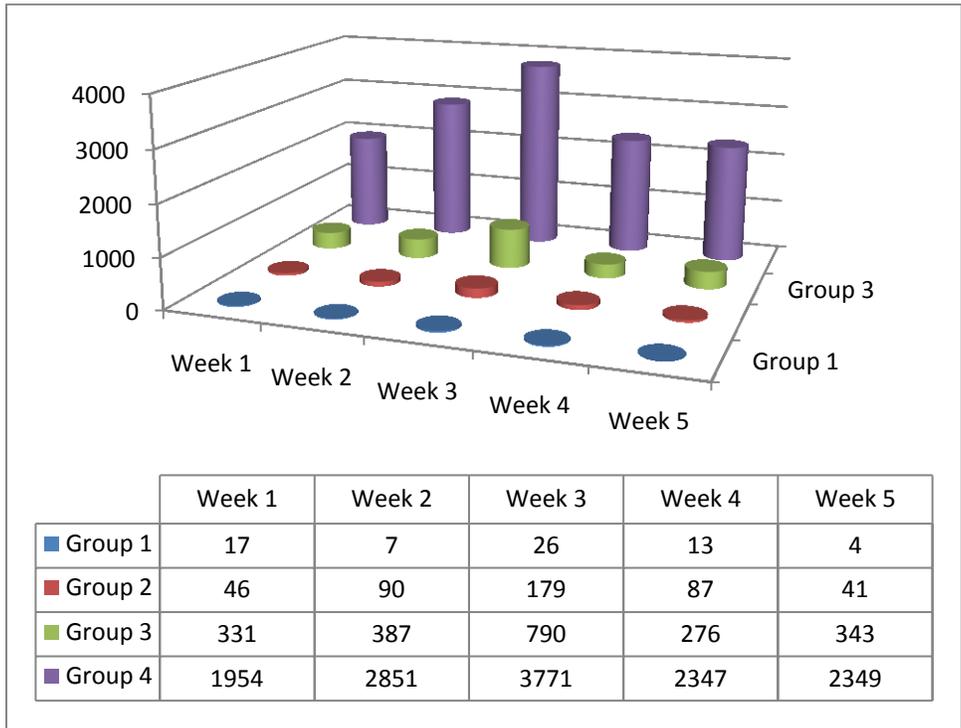


Figure 72 Number of Mentions

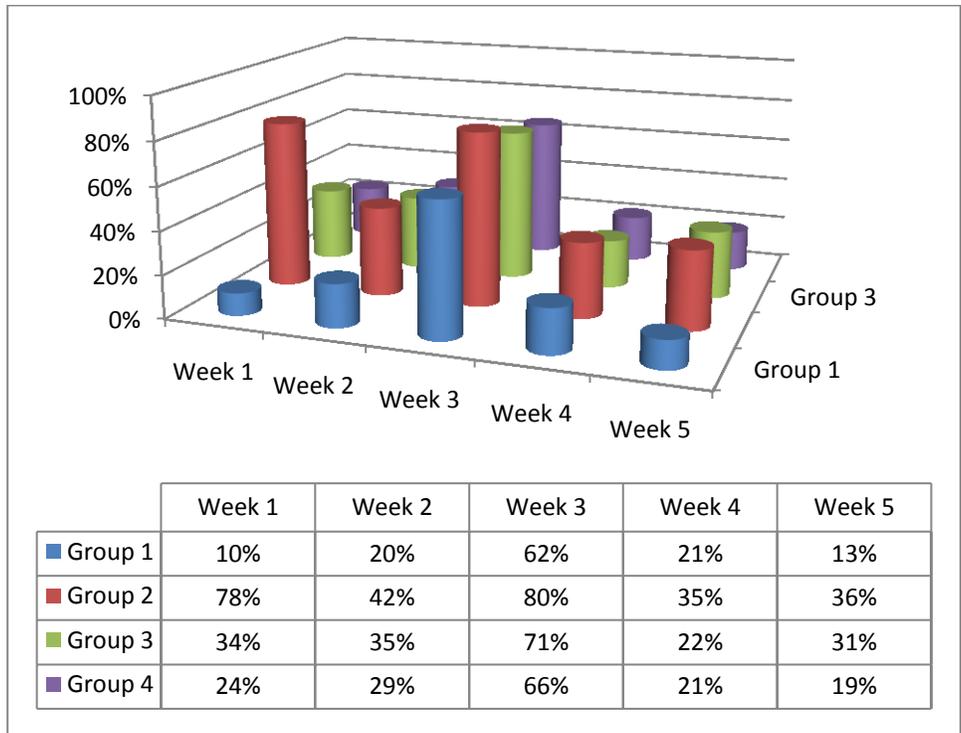


Figure 73 Percentages of Mentions

4.4.3 Relationships

Relationships are an integral part of user interaction on Twitter. As demonstrated by Figure 74, fifty followers and friends was the highest bin for both groups. In other words, for the people at the conference who tweeted, most of the participants had approximately fifty followers and followed fifty friends. It also demonstrates that the distribution of followers and friends is normal and thus relatively identical. When graphing the number of followers and friends for the people who tweeted at the conference, the participants had a normal distribution.

However, averages were different. For the participants in this study the average was 142 friends and 281 followers. Additionally, 200 of the users in this study had more friends than follower. On average, they had 44 more. One hundred and seventy-six participants had more followers than friends.

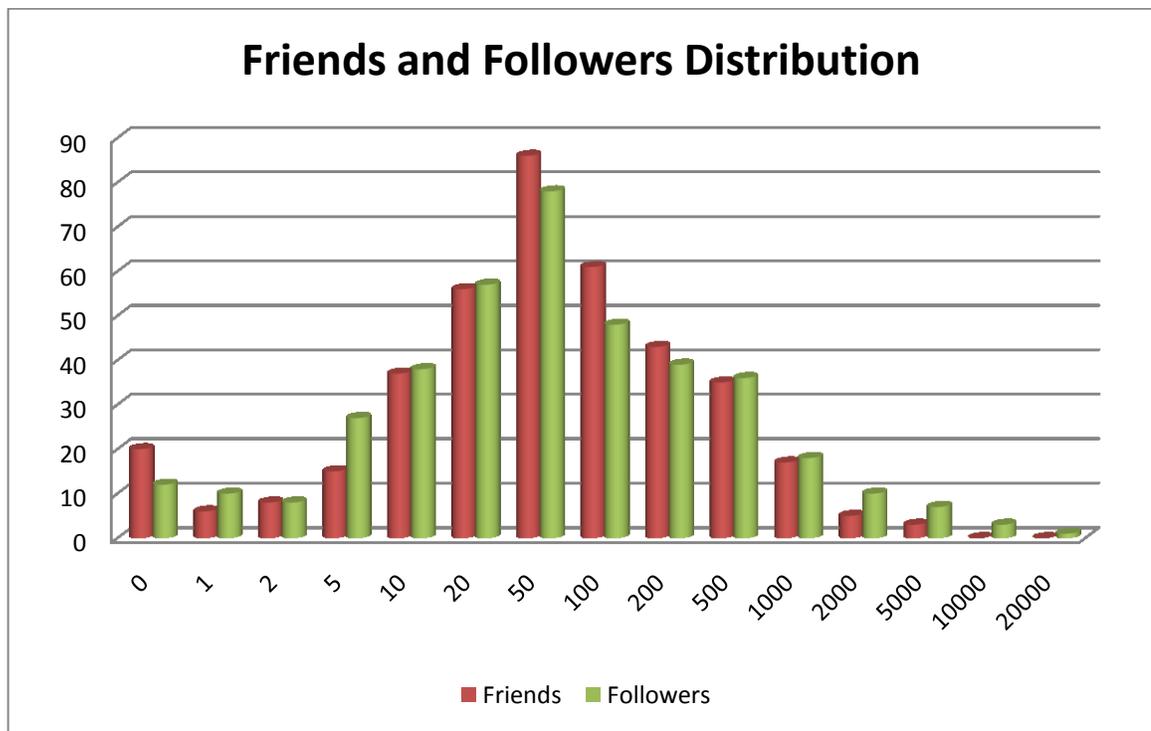


Figure 74 Friends and Followers Distribution

4.4.4 Influence & Centrality

To understand influence and centrality, we used a social network analysis. The definitions that are relevant to the finds are below:

- Node – each object in a graph.
- Edges – Also referred to as links joins two distinct nodes.

- Degree – total number nodes that are adjacent to the node. These nodes are connected by edges.
- Diameter - The diameter of a graph is the maximal distance between any two points on the graph. If the graph is not connected, its diameter is infinite. The maximal distance between all pairs of nodes.
- Density- Measures how many edges are in a set of edges compared to the maximum possible number of edges between vertices in a set of nodes. Measures how close the network is to complete. A complete graph has all possible edges and density equal to 1
- Radius- Measures how many edges are in a set of edges compared to the minimum possible number of edges between vertices in set of nodes.
- Modularity - This structure, often called a community structure, describes how the network is compartmentalized into sub-networks. These sub-networks (or communities) have been shown to have significant real-world meaning. A high modularity score indicates sophisticated internal structure.
- Average degree - The Average Degree is based on the Degree. It shows how many neighbors (adjacent nodes) a node in the network has in average.
- Average weighted degree- Similar to Average Degree, is averaged over all vertices of the graph.
- Strong Connected Components - A directed graph is strongly connected if there is a path from each vertex in the graph to every other vertex. In particular, this means paths in each direction; a path from a to b and also a path from b to a.
- Weakly Connected Components - A directed graph is weakly connected if replacing all of its directed edges with undirected edges produces a connected (undirected) graph.
- Network centralization– Also known as global centrality, measures the degree to which an entire network is focused around a few central nodes.
- Network Centralization Index – Its? minimum value, zero, occurs when all the actors have exactly the same betweenness index. The average difference between the relative centrality of the most central point. “Sum of the differences between the maximal centrality of one point and the centrality of every other point” (Degenee & Forse 1999, p 134). He explains that values of the standardized index lie between 0 and 1, and will be 0 if all nodes achieve the same centrality score and 1 if one actor dominates over all others.
- Network Centralization (Outdegree) – Indicated the degree to which actors in the network are chosen by, a central actor(s), this is, how cohesive the network is around a center in regard to messages being sent out from it
- Network Centralization (Indegree) – demonstrates the degree to which the actors in the network choose a central actor(s), that is, how cohesive the network is around a center in regards to the messages coming into it.

The scores for the network of participants that we studied are in Table 8. It shows a flattening of the social network over time that has more connections, strongly connected components, and shorter paths. This means that the conference had a strong effect on the participants in this study and their network. In particular, the reduction in modularity shows that the community may be getting stronger?.

Table 8 Statistics of Social Graph Evolution Over Time

Date	9/14	9/21	9/28	10/5	10/12	10/19
Nodes	267	288	295	316	318	318
Edges	1160	1352	1437	1875	1911	1921
Average Degree	8.69	9.39	9.74	11.87	12.02	12.08
Average In Degree	4.35	4.69	4.87	5.93	6.01	6.04
Average Out Degree	4.35	4.69	4.87	5.93	6.01	6.04
Average Weighted Degree	17.28	18.54	19.36	23.57	23.86	23.97
Diameter	5	6	5	7	7	7
Radius	0	0	0	0	0	0
Average Path length	2.27	2.27	2.23	2.28	2.29	2.28
Number of shortest paths	42941	53282	56697	68115	69777	69777
Density	0.02	0.01	0.01	0.01	0.01	0.01
Modularity	0.33	0.34	0.34	0.30	0.30	0.31
Number of Communities	19	15	19	23	25	21
Weakly Connected Components	1	1	1	1	1	1
Strongly Connected Components	111	107	106	103	100	100
Network Centralization Index	39.64%	42.73%	43.28%	44.66%	45.56%	45.17%
Network Centralization (Outdegree)	69.68%	74.23%	75.13%	72.95%	73.10%	73.09%
Network Centralization (Indegree)	47.79%	50.81%	52.94%	55.12%	55.38%	55.37%

We measured the betweenness centrality for the social network as a whole, but also with just friends and then followers. Betweenness centrality measures how often a node appears on the shortest paths between nodes in the network.

Table 9 demonstrates relevant issues in regards to the betweenness centrality of the key users in the social network. It demonstrates that those who were at the top usually stayed there. This could be because they were most visible to the network and thus stayed that way. Second, the GHC organization had the highest score. However, this is expected. Additionally, it is important for this user to have a high score, because without them the network would be loosely coupled. Lastly, there were a handful of people who made large gains throughout the experiment and emerged at the front of the pack (e.g. JessiDG).

Table 9 Top 25 User Based on Betweenness Centrality Over Time Full Graph

9/14	9/21	9/28
------	------	------

ghc	28041	ghc	35182	ghc	37397
anitaborg_org	15431	anitaborg_org	18057	anitaborg_org	18651
csimard	1159	csimard	1228	csimard	1080
RcktGrl	910	yardi	1062	conniesmallwood	897
conniesmallwood	809	conniesmallwood	879	yardi	784
ajaimk	505	RcktGrl	625	RcktGrl	568
yardi	481	sbardzell	553	LatinasInC	461
gailcarmichael	460	gailcarmichael	494	LadyLaurian	451
LadyLaurian	457	LadyLaurian	465	gailcarmichael	418
obiwankimberly	347	jpdimond	412	catgrev	405
storming	318	jbhertel	409	obiwankimberly	386
creepyed	295	obiwankimberly	379	corey_latislaw	384
nabilar	276	creepyed	351	creepyed	361
corey_latislaw	275	storming	342	jpdimond	326
HilaryP	275	ajaimk	307	nabilar	306
jennthom	267	nabilar	300	jennthom	301
kpanciera	264	HilaryP	296	HilaryP	297
gillianmsmith	253	xeeliz	296	mynatt	293
jennifer_el	253	jennthom	293	jbhertel	290
jpdimond	219	kpanciera	284	Mediamum	286
sarahsharp	218	mynatt	283	sbardzell	285
terriko	210	jennifer_el	275	storming	283
jenred	207	gillianmsmith	275	jennifer_el	281
OrganizeFISH	192	mfh	271	gillianmsmith	281
jennifermarsman	181	corey_latislaw	265	ajaimk	279
10/5		10/12		10/19	
ghc	44306	ghc	45777	ghc	45388
anitaborg_org	19734	anitaborg_org	19785	anitaborg_org	20083
JessiDG	1736	JessiDG	1799	JessiDG	1829
csimard	1023	csimard	1021	csimard	1012
conniesmallwood	952	conniesmallwood	959	conniesmallwood	958
corey_latislaw	938	corey_latislaw	934	corey_latislaw	912
annejsimmons	894	annejsimmons	900	annejsimmons	900
nancymar	623	nancymar	627	nancymar	627
wordymusic	611	wordymusic	611	gailcarmichael	612
rebeccaparsons	602	gailcarmichael	609	rebeccaparsons	606
gailcarmichael	600	rebeccaparsons	606	ajaimk	600
obiwankimberly	569	ajaimk	600	wordymusic	564
RcktGrl	545	obiwankimberly	569	obiwankimberly	556
yardi	533	RcktGrl	551	RcktGrl	548
kpanciera	489	yardi	535	yardi	519
catgrev	423	kpanciera	493	kpanciera	494

jenred	397	catgrev	467	lhawthorn	467
storming	392	lhawthorn	466	catgrev	467
jpdimond	372	jenred	405	jenred	460
Mediamum	347	storming	401	storming	397
OrganizeFISH	347	jpdimond	372	travelingsw	360
HilaryP	340	travelingsw	360	jpdimond	357
travelingsw	336	OrganizeFISH	357	OrganizeFISH	354
mfh	327	HilaryP	353	HilaryP	350
nabilar	327	Mediamum	349	jennifermarsman	340

Researchers created the following parallel coordinate visualization by taking the betweenness centrality numbers for each user and divided by that week's maximum betweenness centrality to derive a percentage. Two users were left off intentionally because their scores were an order of magnitude larger than the rest of the group.

	Week 1	Week 2	Week 3	Week 4	Week 5
ghc	100%	100%	100%	100%	100%
anitaborg_org	55%	51%	50%	45%	44%

Figure 75 The two user left off the parallel coordinate visualization

Figure 75 depicts the participant who held the top 25 largest percentage of betweenness centrality at the end of the five week period. Two users, JessiDG and annejsimmons, did not have any significant percentage of betweenness centrality at the beginning of the conference and then afterward rose upwards to have a large portion of the centrality. This means that these users formed connection between different groups of users in the social graph that make them the? central broker of information. Csimard, conniesmallwood, and RcktGrl all have declining centralities in the network. This implies that the network is less dependent on these users to act as a broker between groups.

Figure 76 demonstrates other larger, general trends. First, a large portion of the users stay relatively the same. Second, a few years decrease in their betweenness score. Lastly, few users drastically increase their betweenness score. This means that they may be better at networking than others who are on Twitter.

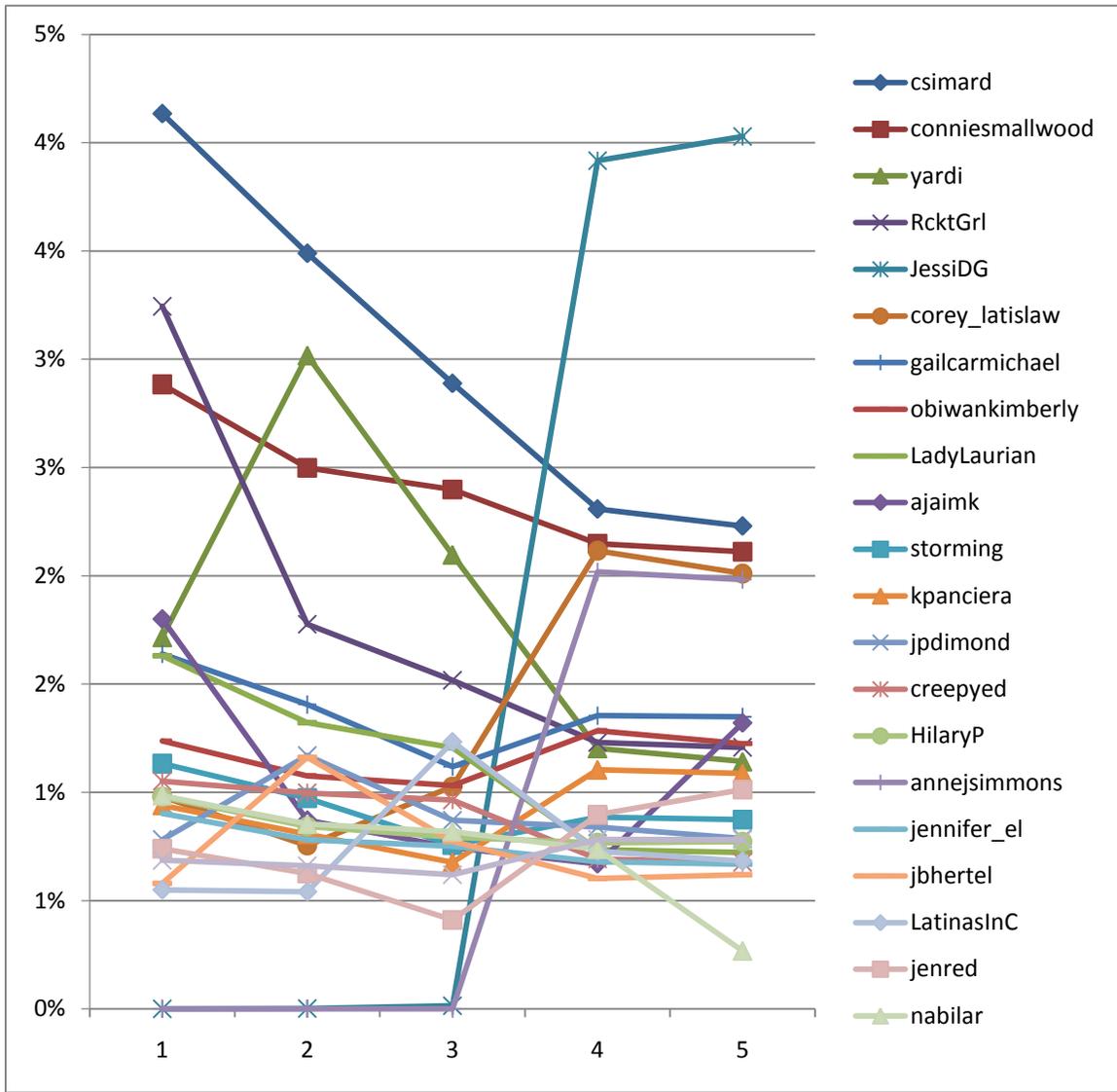


Figure 76 Parallel Coordinates Visualization of Top 25 Users Over Time

Figure 77 shows the betweenness centrality numbers by group for the study period. Group 4 – Immersed had an increased level of centrality compared to the other groups. Additionally, Group 4 had a bump in their betweenness centrality score in the weeks prior to the conference. This group makes up the central users in the social graph for the attendees on Twitter.

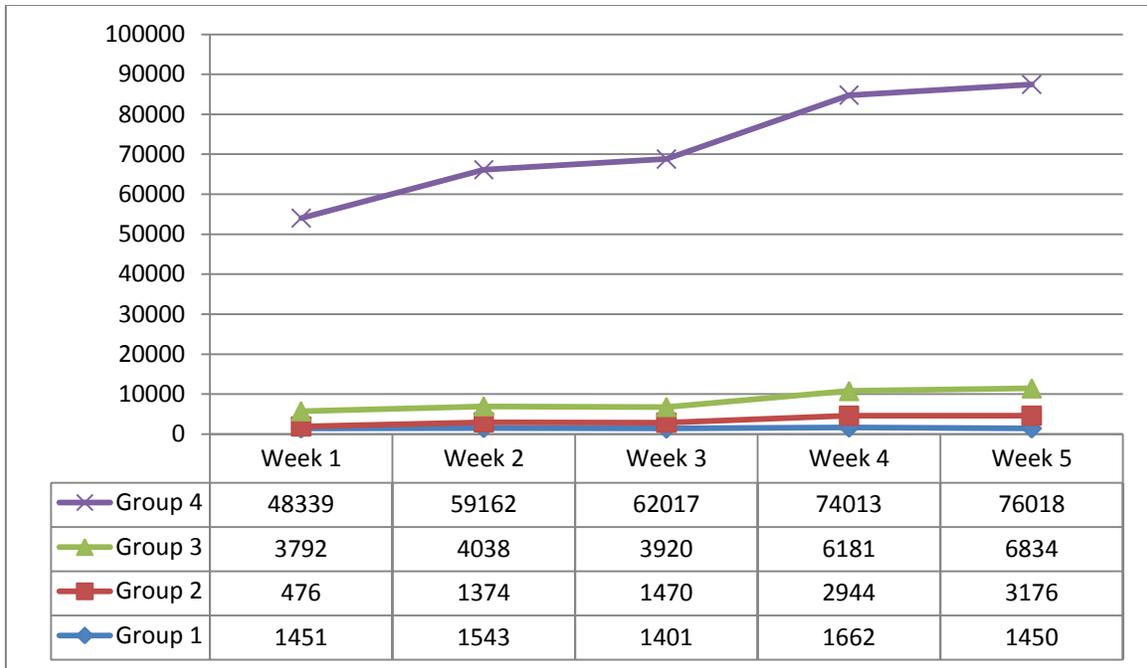


Figure 77 Parallel Visualization by Group on the Sum of their betweenness centrality

When focused only on the impact of friends and time, Table 10 demonstrates a few key trends. The first is that these same people were present in Table 9. The similarity indicates that these are still key players and not flukes. Participants could have also gained betweenness centrality scores by following everyone who was at the conference. However, this table demonstrates that people, in fact, followed these people back. This means that the network is strong.

Table 10 Top 25 User Based on Betweenness Centrality Over Time Friends Graph

9/14		9/21		9/28	
ghc	27291	ghc	34122	ghc	36357
anitaborg_org	15023	anitaborg_org	17425	anitaborg_org	18219
csimard	1141	csimard	1200	csimard	1062
RcktGrl	903	yardi	990	conniesmallwood	879
conniesmallwood	791	conniesmallwood	858	yardi	692
ajaimk	493	RcktGrl	618	RcktGrl	563
yardi	471	sbardzell	538	LatinasInC	454
gailcarmichael	452	gailcarmichael	482	LadyLaurian	445
LadyLaurian	447	LadyLaurian	453	gailcarmichael	410
obiwankimberly	340	jpgimond	403	catgrev	395
storming	312	jbhertel	399	corey_latislaw	389
creepyed	290	obiwankimberly	371	obiwankimberly	379
corey_latislaw	272	creepyed	341	creepyed	359
HilaryP	271	storming	335	storming	324

nabilar	270	ajaimk	298	jpdiamond	323
jennthom	261	nabilar	293	nabilar	300
kpanciera	255	HilaryP	292	HilaryP	294
gillianmsmith	247	xeeliz	287	jennthom	291
jennifer_el	247	jennthom	284	mynatt	287
jpdiamond	216	kpanciera	279	jbhertel	283
sarahsharp	213	mynatt	277	Mediamum	282
terriko	205	gillianmsmith	268	sbardzell	278
jenred	202	jennifer_el	268	jennifer_el	275
OrganizeFISH	189	corey_latislaw	262	gillianmsmith	275
jennifermarsman	177	sarahsharp	238	ajaimk	273
10/5		10/12		10/19	
ghc	42145	ghc	43593	ghc	43211
anitaborg_org	19011	anitaborg_org	19057	anitaborg_org	19353
JessiDG	1701	JessiDG	1764	JessiDG	1794
csimard	1017	csimard	1015	csimard	1007
corey_latislaw	941	conniesmallwood	942	conniesmallwood	941
conniesmallwood	935	corey_latislaw	937	corey_latislaw	915
nancymar	612	nancymar	617	nancymar	616
gailcarmichael	593	gailcarmichael	602	gailcarmichael	605
wordymusic	578	ajaimk	590	ajaimk	590
obiwankimberly	560	wordymusic	579	obiwankimberly	547
RcktGrl	534	obiwankimberly	560	RcktGrl	538
yardi	524	RcktGrl	540	wordymusic	531
kpanciera	480	yardi	526	yardi	510
catgrev	415	kpanciera	484	kpanciera	484
jenred	389	lhawthorn	459	lhawthorn	459
storming	386	catgrev	457	catgrev	458
jpdiamond	368	jenred	397	jenred	451
OrganizeFISH	340	storming	395	storming	391
Mediamum	340	jpdiamond	368	jpdiamond	353
HilaryP	334	travelingsw	352	travelingsw	353
travelingsw	329	OrganizeFISH	350	OrganizeFISH	348
LadyLaurian	321	HilaryP	347	HilaryP	344
nabilar	320	Mediamum	343	jennifermarsman	334
mfh	320	LadyLaurian	324	Mediamum	328
LatinasInC	317	nabilar	323	LadyLaurian	323

A corollary to Table 10, Table 11 lists the betweenness centrality of followers over time. The people listed are the ones that others find to be the most popular.

What is interesting is the comparison between the people listed on Table 11 and the people listed on Table 10. They are relatively the same people. This is a further indicator that users tend to keep the number of followers and friends the same.

Table 11 Top 25 User Based on Betweenness Centrality Over Time Followers Graph

9/14		9/21		9/28	
ghc	27802	ghc	34922	ghc	37409
anitaborg_org	15388	anitaborg_org	18020	anitaborg_org	18661
csimard	1143	csimard	1211	csimard	1064
RcktGrl	906	yardi	1061	conniesmallwood	897
conniesmallwood	808	conniesmallwood	878	yardi	785
ajaimk	505	RcktGrl	623	RcktGrl	567
yardi	481	sbardzell	553	LatinasInC	461
gailcarmichael	460	gailcarmichael	495	LadyLaurian	451
LadyLaurian	457	LadyLaurian	465	gailcarmichael	419
obiwankimberly	345	jpdiamond	411	catgrev	405
storming	317	jbhertel	409	corey_latislaw	384
creepyed	295	obiwankimberly	376	obiwankimberly	383
nabilar	276	creepyed	351	Creepyed	361
corey_latislaw	275	storming	342	Jpdiamond	325
HilaryP	274	ajaimk	307	Nabilar	306
jennthom	267	nabilar	300	Jennthom	301
kpanciera	264	xeeliz	296	HilaryP	297
gillianmsmith	253	HilaryP	295	Mynatt	293
jennifer_el	253	jennthom	293	Jbhertel	290
jpdiamond	219	kpanciera	284	Mediamum	286
sarahsharp	218	mynatt	283	Sbardzell	285
terriko	210	jennifer_el	275	Storming	283
jenred	207	gillianmsmith	275	jennifer_el	281
OrganizeFISH	193	mfh	271	Gillianmsmith	281
jennifermarsman	181	corey_latislaw	265	Ajaimk	279
10/5		10/12		10/19	
ghc	43189	ghc	44658	ghc	44220
anitaborg_org	19740	anitaborg_org	19796	anitaborg_org	20144
JessiDG	1701	JessiDG	1763	JessiDG	1789
csimard	1006	csimard	1005	csimard	999
corey_latislaw	958	corey_latislaw	956	conniesmallwood	954
conniesmallwood	947	conniesmallwood	955	corey_latislaw	937
nancymar	621	nancymar	625	nancymar	624
wordymusic	612	gailcarmichael	617	gailcarmichael	623
gailcarmichael	608	wordymusic	614	ajaimk	598

obiwankimberly	564	ajaimk	598	wordymusic	567
RcktGrl	542	obiwankimberly	563	obiwankimberly	551
yardi	534	RcktGrl	548	RcktGrl	546
kpanciera	487	yardi	535	yardi	521
travelingsw	466	travelingsw	492	travelingsw	495
catgrev	420	kpanciera	491	kpanciera	491
jenred	396	lhawthorn	465	lhawthorn	466
storming	390	catgrev	465	catgrev	465
jpdiamond	370	jenred	404	jenred	459
Mediamum	345	storming	399	storming	396
OrganizeFISH	345	jpdiamond	370	jpdiamond	356
HilaryP	338	OrganizeFISH	356	OrganizeFISH	355
LadyLaurian	330	Mediamum	348	HilaryP	344
mfh	325	HilaryP	346	jennifermarsman	335
nabilar	325	LadyLaurian	333	LadyLaurian	334
LatinasInC	317	nabilar	328	Mediamum	333

Figure 78, Figure 79, Figure 80, Figure 81, Figure 82, and Figure 83 depict the network diagram of the relations between attendees in our study held during the conference. Each figure represents the time slices that were analyzed in this study. The size of the labels for each node represents the betweenness centrality score of the user. Each line represents a relationship link between two users. The application colors the lines and nodes based on the modularity the structure. This is to say, how the network compartmentalizes into sub-networks, therefore, users with like colors are in the same sub-networks. These sub-networks (or communities) have significant real-world meaning. The application colors the nodes by which modality group the user is in and colors the lines by blending the colors of the connecting nodes.

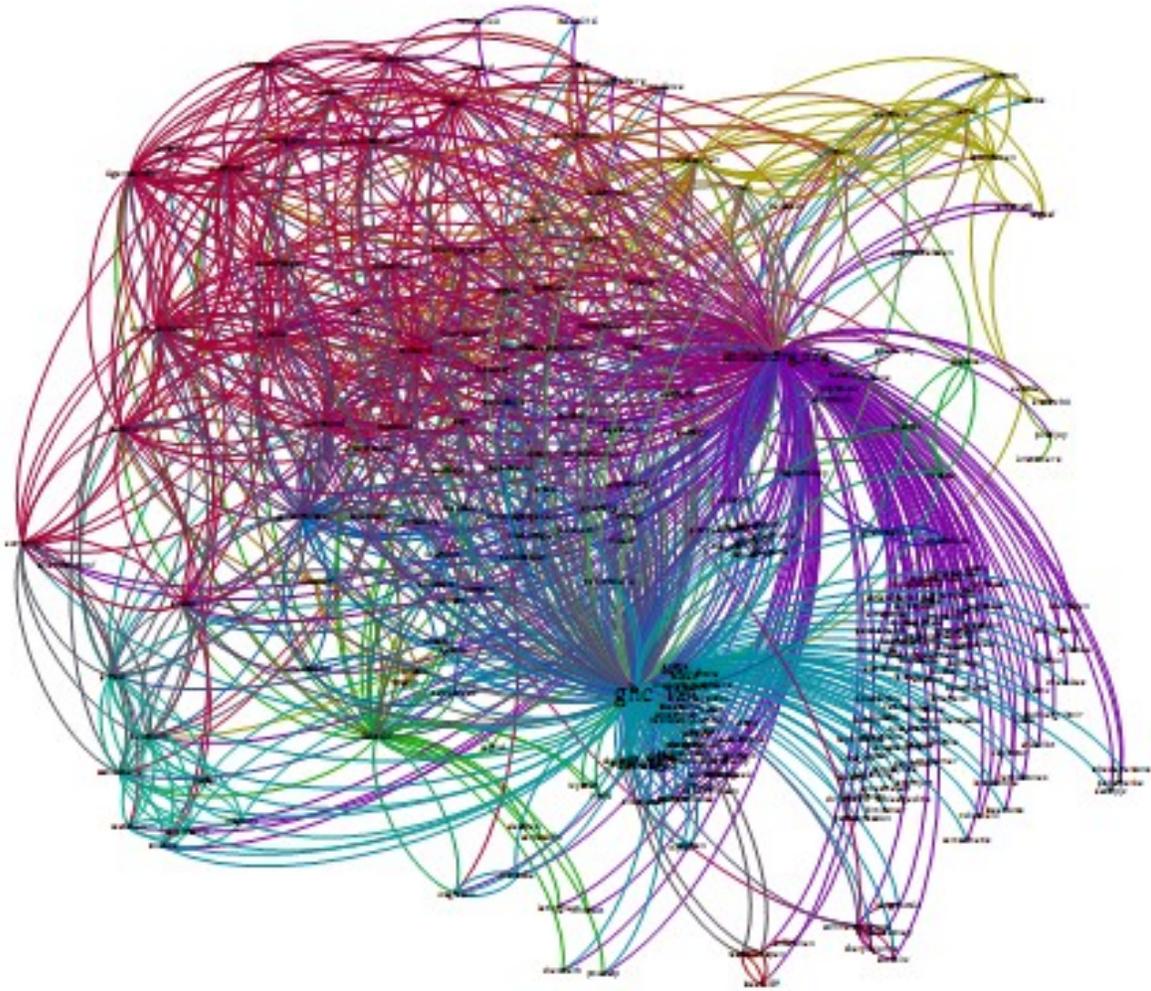


Figure 78 Social Graph time slice at the date of 9/14

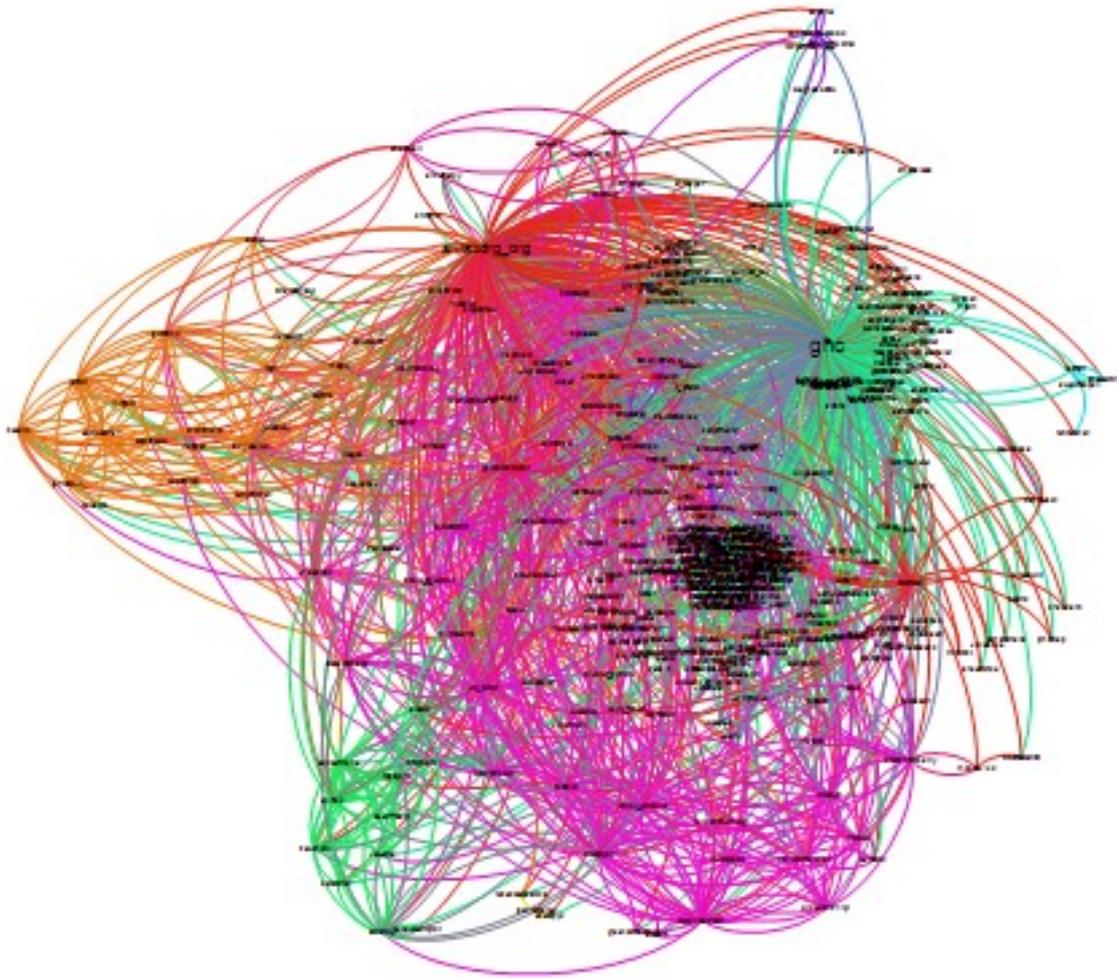


Figure 79 Social Graph time slice at the date 9/21

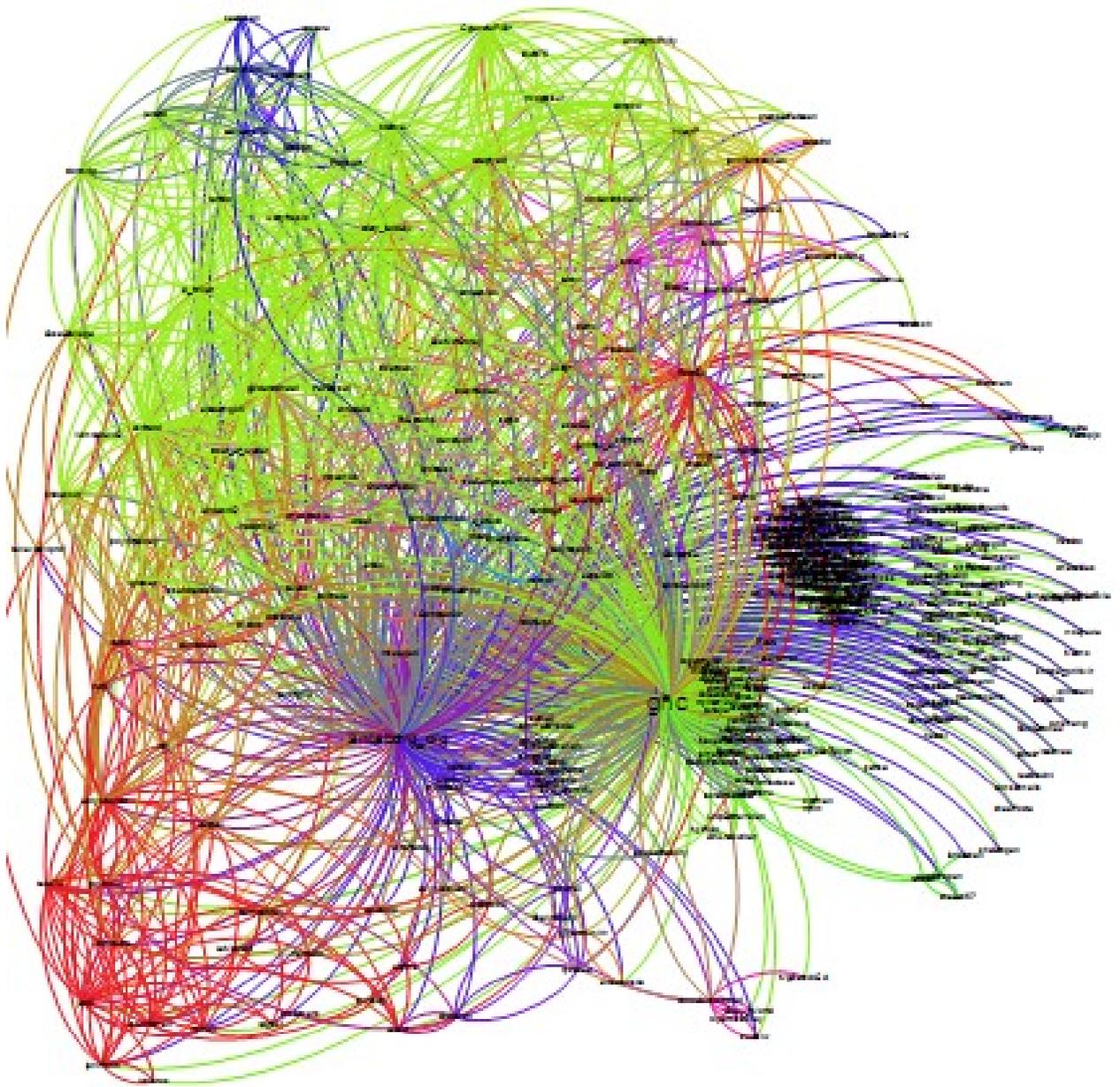


Figure 80 Social Graph time slice at the date 9/28

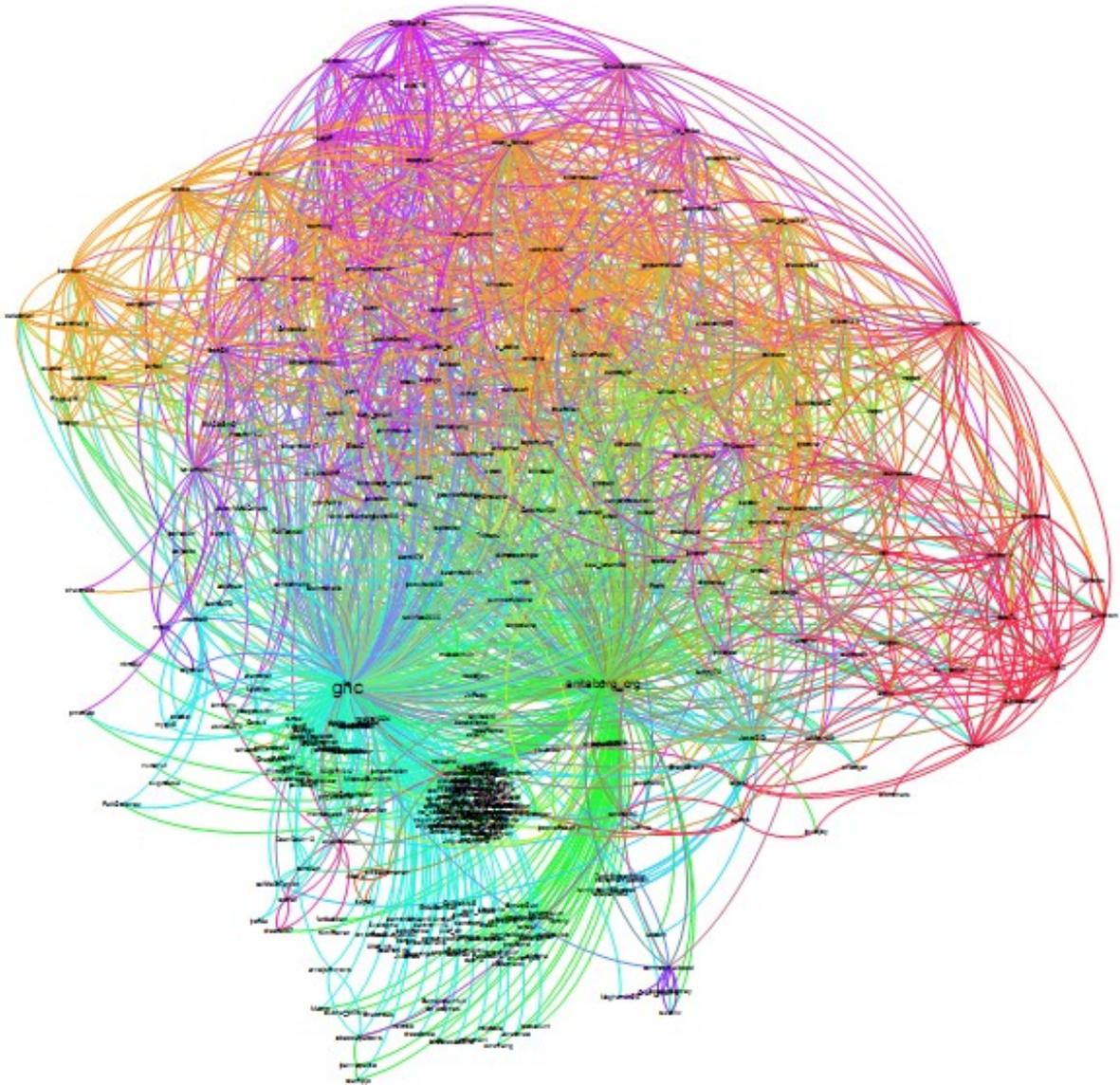


Figure 81 Social Graph time slice at the date 10/5

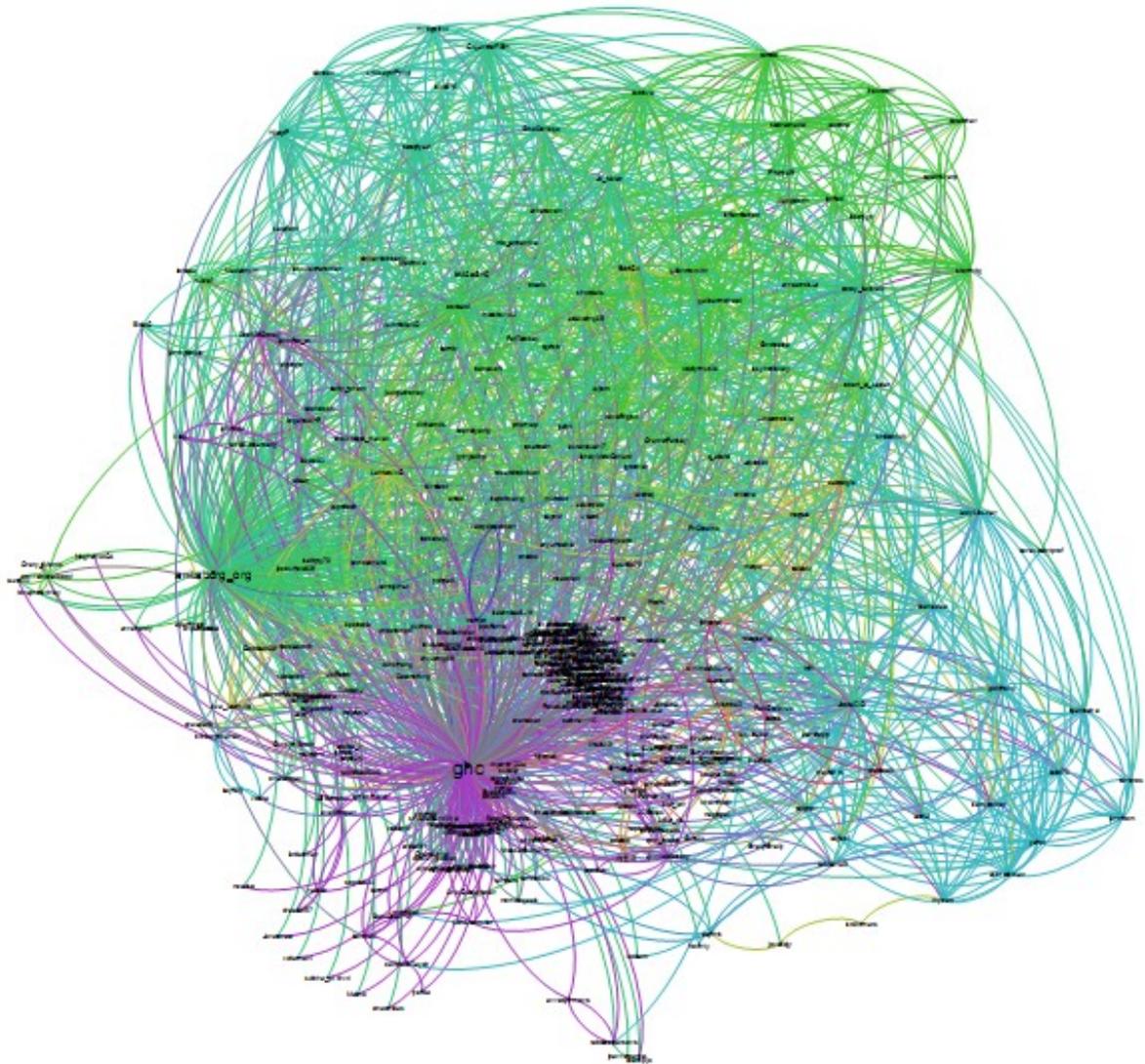


Figure 82 Social Graph time slice at the date 10/12

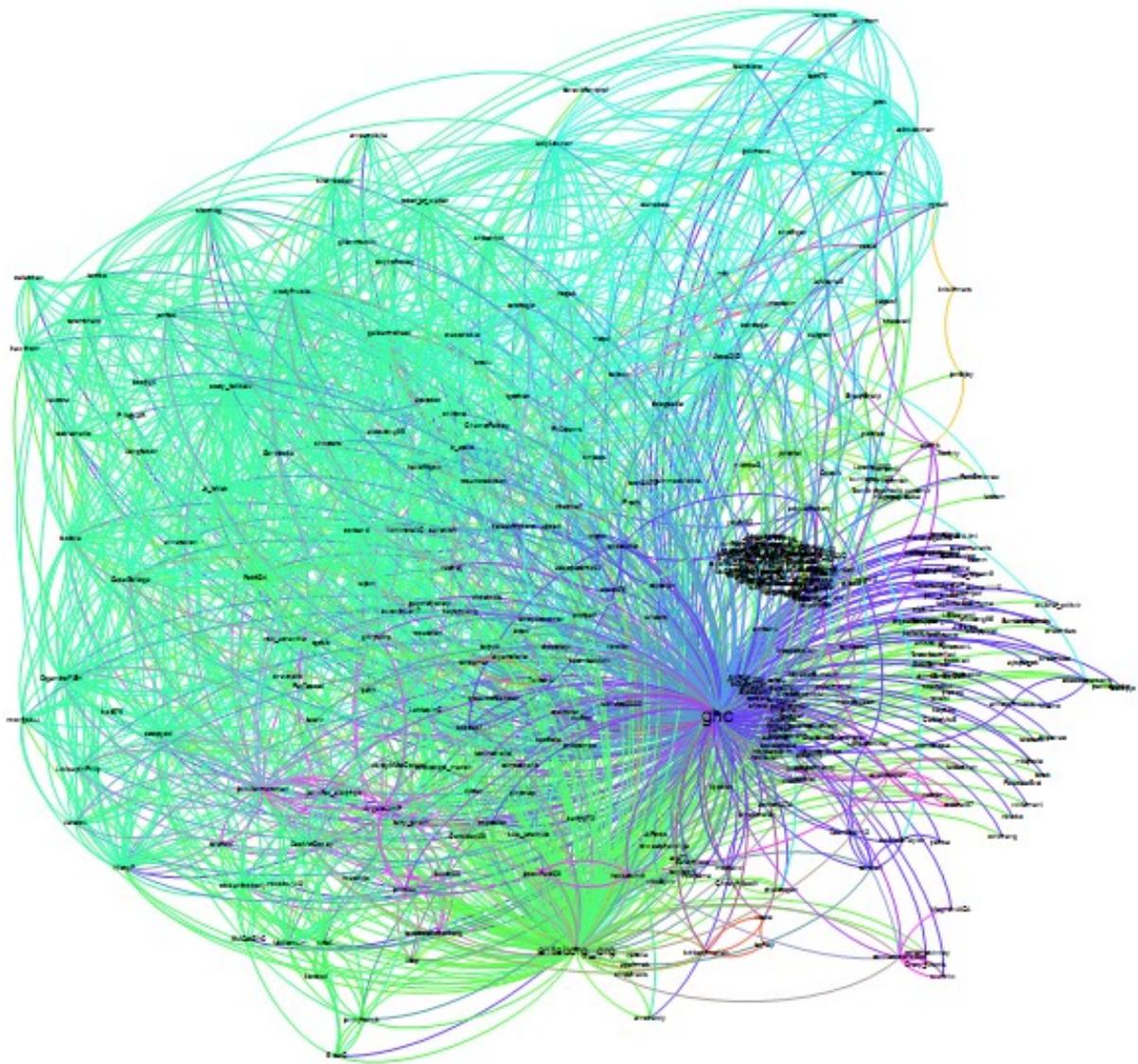


Figure 83 Social Graph time slice at the date 10/19

4.4.5 Retweeting

A retweet is a way for a user to share a Tweet that is in another user’s timeline. When retweeted, Twitter will display it in all the timelines of the followers of that user. There is an official retweet function inside of Twitter and an unofficial one. The numbers displayed below are of only the official retweets.

4.4.5.1 Group 1- Experimenter

Group 1 - Experimenter used 28 retweets or used retweets in 13% of tweets.

Week 1	Week 2	Week 3	Week 4	Week 5
--------	--------	--------	--------	--------

# Retweets	4	2	11	9	2	28	<i>total</i>
# Tweets	29	25	97	39	31	221	<i>total</i>
% of Total	14%	8%	11%	23%	6%	13%	<i>average</i>

4.4.5.2 Group 2- Part-timer

Group 2 - Part-timer used 29 retweets or used retweets in 9% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Retweets	6	20	33	22	11	92	<i>total</i>
# Tweets	107	159	521	158	132	1077	<i>total</i>
% of Total	6%	13%	6%	14%	8%	9%	<i>average</i>

4.4.5.3 Group 3- Active

Group 3 - Active used 506 retweets or used retweets in 14% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Retweets	73	83	216	69	65	506	<i>total</i>
# Tweets	509	618	1408	449	548	3532	<i>total</i>
% of Total	14%	13%	15%	15%	12%	14%	<i>average</i>

4.4.5.4 Group 4- Immersed

Group 4 – Immersed used 2123 retweets or used retweets in 13% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Retweets	254	506	685	320	358	2123	<i>total</i>
# Tweets	2296	3018	5506	2821	2679	16320	<i>total</i>
% of Total	11%	17%	12%	11%	13%	13%	<i>average</i>

4.4.5.5 Unofficial Retweets

Research (boyd, Golder et al. 2010) have found other syntax used to mark retweets: 'RT: @', 'retweeting @', 'retweet @', '(via @)', 'RT(via @)', 'thx @', 'HT @', 'r @', and ' @'.

Out of the tweets discovered by looking for the official Grace Hopper Conference Hashtags (Table 1 and Table 2), I found trends in the retweets pulled from the tweets from the participants in the study, as demonstrated in Table 12 and Table 13. The first trend is that both groups of participants mostly used the official retweet mechanism (i.e., "RT"). Another trend was that apart from the official retweet mechanism, the "via @" and "RT(via @)" mechanisms were predominant. Users may have used these styles because users prefer adding their comments to a tweet before reposting it.

Table 12 Unofficial Retweets from List

Week 1	Week 2	Week 3	Week 4	Week 5	Total
--------	--------	--------	--------	--------	-------

thx @	1	1	0	6	0	8
via @	26	31	80	39	24	200
r @	0	0	1	0	0	1
RT: @	2	1	1	1	5	10
RT(via @)	20	21	43	15	20	119
RT	877	1217	2222	1029	1004	6349
Total	926	1271	2347	1090	1053	6687

Table 13 Unofficial Retweet from GHC HashTags

	Week 1	Week 2	Week 3	Week 4	Week 5	Total
thx @	0	0	0	3	0	3
via @	2	1	59	2	0	64
r @	0	0	1	0	0	1
RT: @	0	0	1	0	0	1
RT(via @)	3	4	54	5	0	66
RT	93	175	1684	73	36	2061
Total	98	180	1800	83	36	2196

4.4.5.6 Summary

In all, there were a total of 2,749 tweets that are retweets made by the participants of this study. In this study, retweets were used in 13% of all tweets.

As defined by Liere (2010), retweeters serve as a kind of information broker in a community. By this definition, and by examining Figure 84 and Figure 85, we can see that Group 4 acted as the biggest information brokers. However, retweeting was higher for all types of users, especially the week of the conference.

In terms of percentages, as demonstrated in Figure 85, Group 4 retweeted the most the week prior to the conference. This could be because people in this group started to establish connections to others who would be attending the conference.

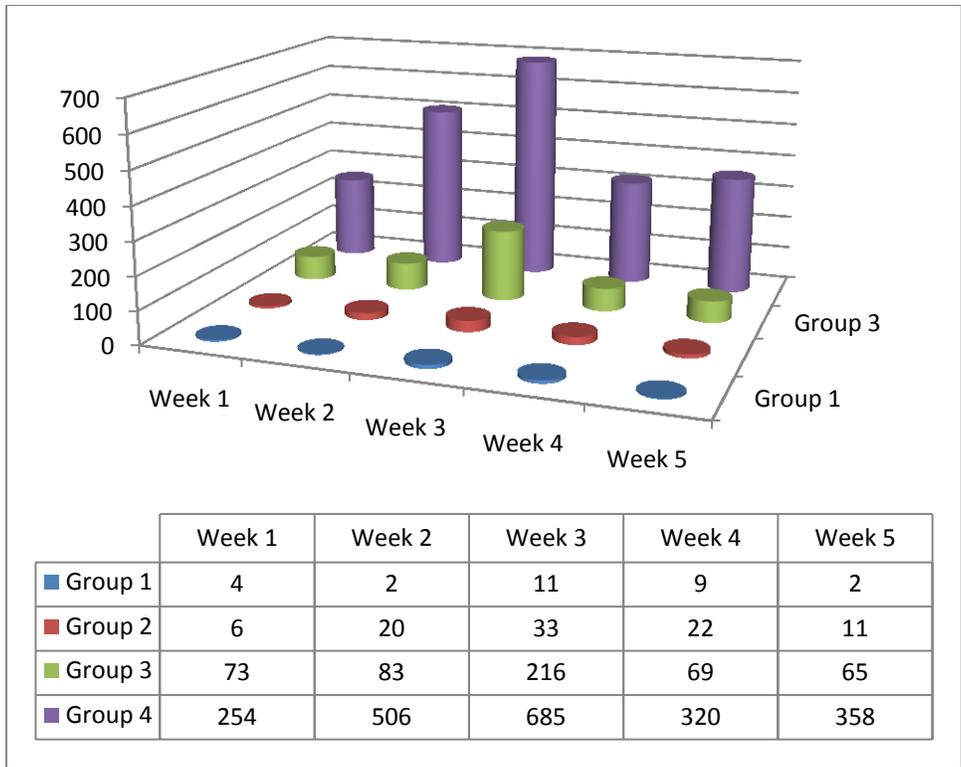


Figure 84 Number of Retweets

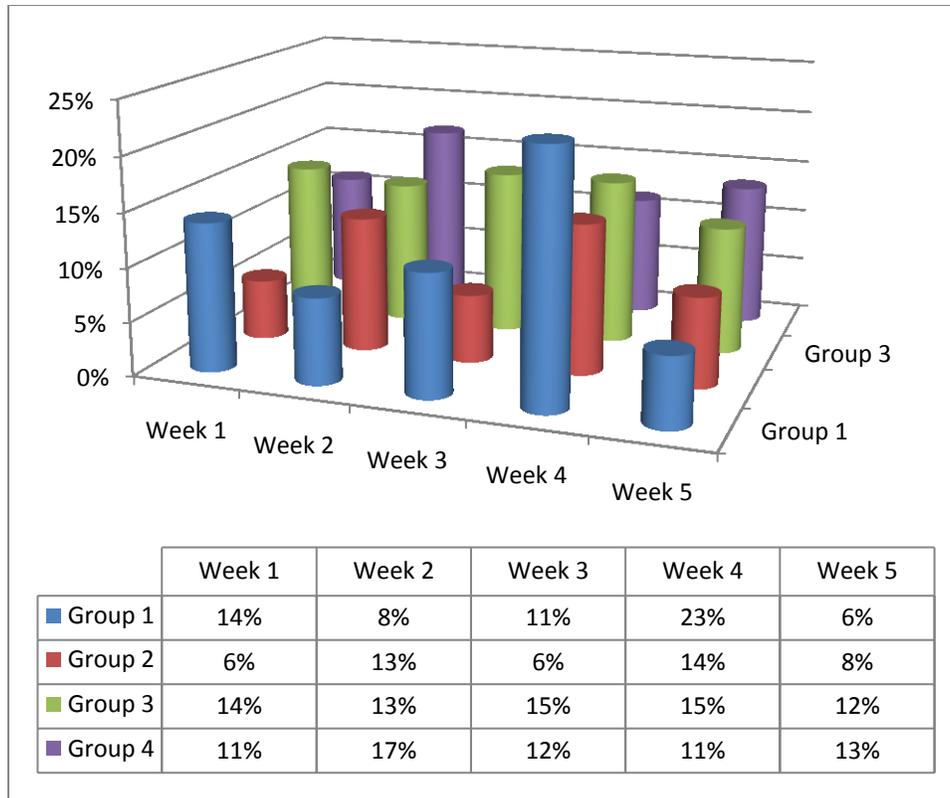


Figure 85 Percentages of Retweets

4.4.6 Replies

Twitter has the official capacity to reply to a tweet. Twitter identifies a tweet as being in reply when a tweet has the `in_reply_to_status_id` field filled in. All official Twitter clients have this feature. If a user only mentions the user, but their client does not include `in_reply_to_status_id` field, it not considered a reply. In this study, if a Tweet has the `in_reply_to_status_id` field filled in it is counted as a reply message, otherwise it is not. These are tweets in which users replied to a previous tweet.

4.4.6.1 Group 1- Experimenter

Group 1 - Experimenter replied to others 9 times or replied in 4% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5	
# Replies	2	2	5	0	0	9 <i>total</i>
# Tweets	29	25	97	39	31	221 <i>total</i>
% of Total	7%	8%	5%	0%	0%	4% <i>average</i>

4.4.6.2 Group 2- Part-timer

Group 2 - Part-timer replied to others 126 times or replied in 12% of tweets.

Week 1	Week 2	Week 3	Week 4	Week 5
--------	--------	--------	--------	--------

# Replies	16	22	62	17	9	126	<i>total</i>
# Tweets	107	159	521	158	132	1077	<i>total</i>
% of Total	15%	14%	12%	11%	7%	12%	<i>average</i>

4.4.6.3 Group 3- Active

Group 2 - Part-timer replied to others 670 times or replied in 20% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Replies	78	130	238	93	131	670	<i>total</i>
# Tweets	509	618	1408	449	548	3532	<i>total</i>
% of Total	15%	21%	17%	21%	24%	20%	<i>average</i>

4.4.6.4 Group 4- Immersed

Group 4 - Immersed replied to others 5727 times or replied in 36% of tweets.

	Week 1	Week 2	Week 3	Week 4	Week 5		
# Replies	808	1037	1697	1086	1099	5727	<i>total</i>
# Tweets	2296	3018	5506	2821	2679	16320	<i>total</i>
% of Total	35%	34%	31%	38%	41%	36%	<i>average</i>

4.4.6.5 Summary

In all, there were a total of 6,532 replies made by the Twitter users of this study, thus making replies 31% of user activity, as shown in Figure 86 and Figure 87.

Figure 86, in particular, demonstrates a few important trends. The first is that Group 4 replied to most of the tweets. This means that they held most of the conversations. The second is that all groups increased their replies in the week of the conference. Additionally, Groups 1 and 2 had more activity before the conference, whereas, Groups 3 and 4 had more activity after the conference.

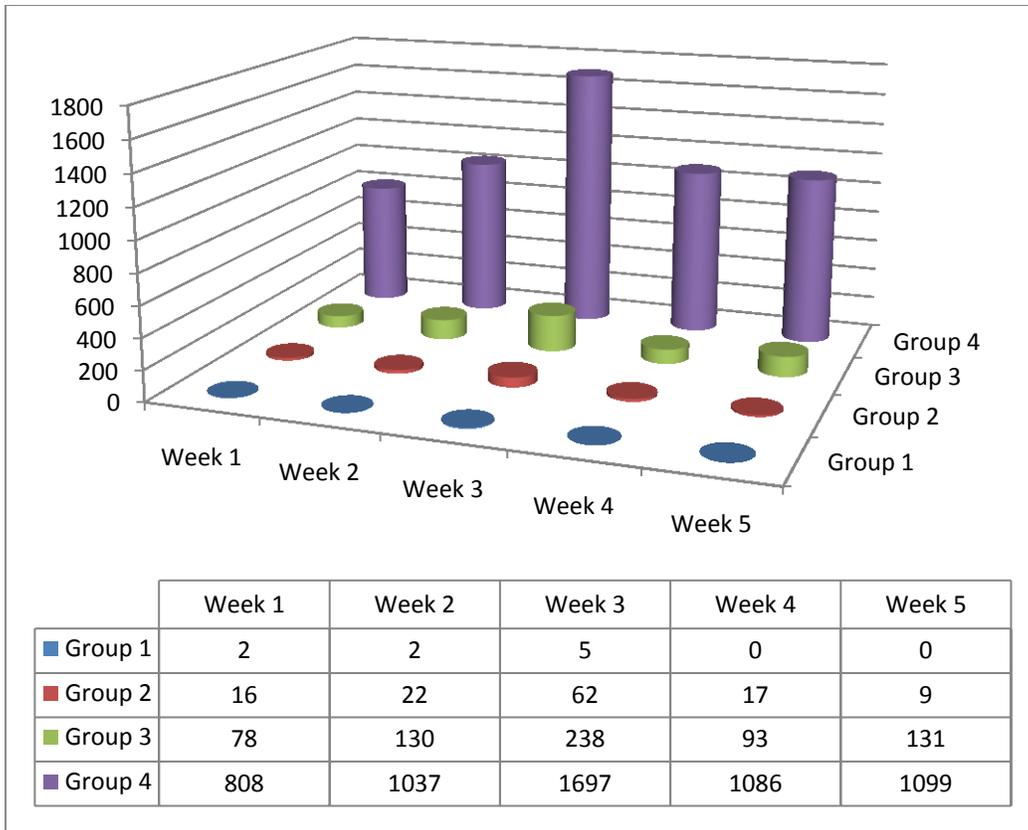


Figure 86 Number of Replies

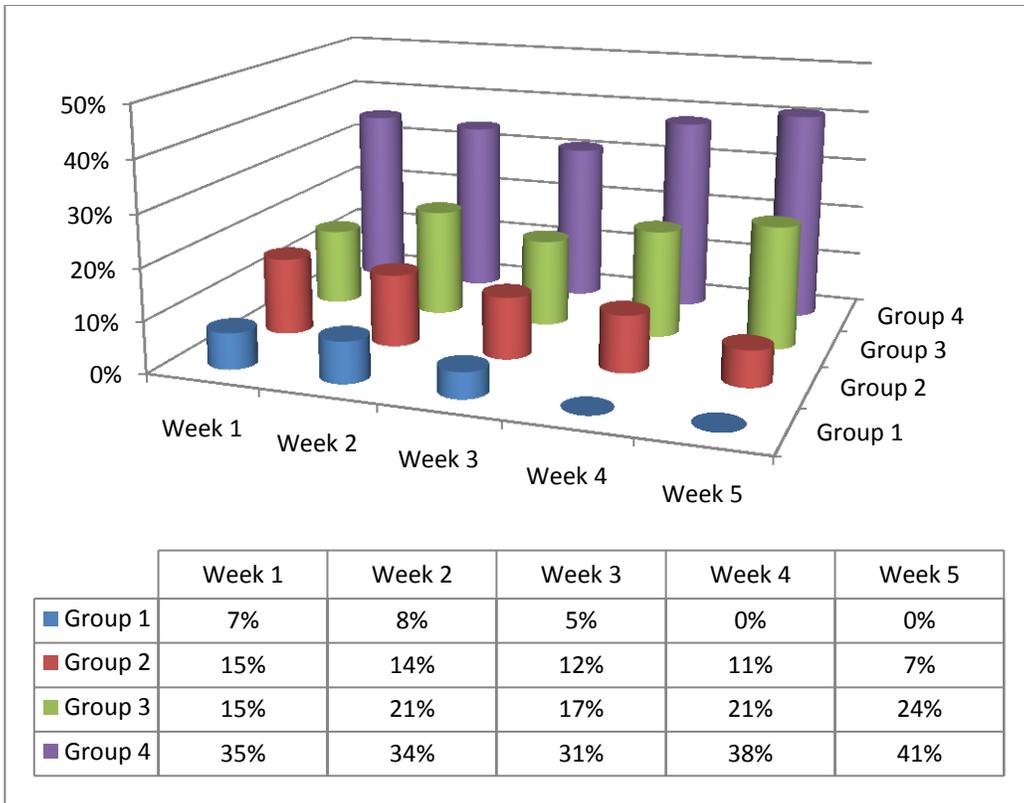


Figure 87 Percentages of Replies

5 Discussion

This thesis and all other work like it reflects an area of research. In particular, it focuses on the need to try and understand the growing trend of online social networks and their use. People are using OSNs for marketing, online meeting places, customer relationship management, and personal and professional networking tools. People use these tools as part of their everyday personal and professional lives.

The measurements taken in this study give a picture of users' activities on Twitter. Not only do these measurements allow us to detect the user's activity, but also how other users are interacting with that individual. We took these measurements because we were interested in studying:

- The people using Twitter at conferences.
- The type of conversations they are having over Twitter.
- The community that forms around these users and uses of Twitter.
- The overall activity at the conference, in particular when compared to the two weeks before and the two weeks following it.

We studied the Grace Hopper Conference and its users. Grace Hopper is a conference held annually to help encourage and celebrate women in computing. In 2010, the conference took place in October at the Hyatt Regency Hotel in Atlanta, Georgia.

There are particular characteristics about this conference that are important to mention and thus situate the findings of this research. First, the conference is predominantly attended by females. Gendered behaviors may skew the finding in this study. Second, this conference was set up to encourage the use of Twitter. For example, the organization running the conference provided sample hashtags for sessions and there was a moderator (@ghc) who actively retweeted and encouraged interconnections between people attending the conference. Conferences without such encouragement may not have similar results. Lastly, this conference is particularly large. Smaller conferences may not have the same kind of critical momentum to garner the same sense of community and context that this conference establishes.

Beyond those limitations, I gathered and presented the findings using a two-pronged method in this thesis. First, I collected all of the tweets in the two weeks prior to the conference, the week of the conference, and the two weeks after the conference. The data collected consists of 399 users and 21,150 tweets. Secondly, 78 tweeters at the conference participated in an online survey that asked demographic questions and questions related to tweeting at the conference.

5.1 Tweeple

In this study, we presented four different types of users based on the number of total tweets. As a way to compare and contrast the measurements collected, we created the four groups to see if tweeting activity corresponded to other Twitter behaviors. Some of the results were direct results of this categorization while others were not. The Tweeple in my study were broken down into the groups: Experimenters, Part-timers, Active, and Immersed users. Two other groups we identified: Inactive users and Lurkers. Inactive users did not tweet or did not create any new friendships during the collection period. Lurkers did not tweet, but did create friendships.

Group 4 (Immersed users) had the most conversations and tweet activity during the study. Group 4 were the users who created emergent hashtags and also participated in the emergent conversations in the network.

Group 2 (Part Timers) reduced their Twitter activity the week after the conference to below starting levels, while the other groups had higher level of Twitter participation after that conference. We could call this a hangover of over tweeting during the conference. The hangover could have been caused by the amount of tweeting this group did at the conference, the conference itself, the amount of work needed to be done after the conference, or a combination of all of the above. This could be because they are the only group who are either still experimenting or not actively engaged with Twitter on a daily basis. There is no anchor to keep them in the system after such over exposure.

Relationships are an integral part of user interaction on Twitter. As demonstrated by the histogram in Figure 74, fifty followers and friends was the highest bin for both groups. In other words, of the people at the conference who tweeted, most of the participants had approximately fifty followers and followed fifty friends. When graphing the number of followers and friends for the people who tweeted at the conference, the participants had a normal distribution.

For the participants in this study the average was 142 friends and 281 followers. Additionally, 200 of the users in this study had more friends than followers. On average, they had 44 more. One hundred and seventy-six participants had more followers than friends. Even though the largest bin in the histogram, in Figure 74, had around 50 friends and follower, the averages were skewed much higher. Therefore, many participants in this study had significantly more friends and followers than the histogram indicates. The number of friends in a person's network could relate to how engaged a user is in Twitter and, therefore, activity levels.

5.2 Activity

For review, people used Twitter for personal and professional reasons. 60.3% participants reported that they used Twitter for both personal and professional purposes. 25.6% use Twitter for personal purposes, and 14.1% use Twitter for professional purposes. Conferences also serve objectives both professionally and personally for the attendee. Conferences pass knowledge to

their attendees and also allow for the expansions of personal networks by engaging with new people and friends.

All groups had increased activity during the conference. When people are collocated, like at conferences, information sharing and conversation, which happens in the physical network, is reflected into the electronic network. This means physical interaction affects electronic interaction, as defined by the centrality of the network. More importantly it is the event that causes the increased activity in the electronic space. This is important because we always want to think about supporting both the physical and electronic interactions of users.

All groups used mentions across the five weeks of the study— the most of any type of tweet. However there was one instance where, in Week 3, the use of mentions for Group 2 (Part-Timers) decreased and hashtags greatly increased. We can speculate that, because Group 2 are still fairly infrequent Twitters, during the conference they wanted to engage in the larger conversations. The mentions only facilitate being able to list a few people, but hashtags allow users to share in a larger group's context, such as using '#ghc10.' This shift is important for conference attendees, because it means that, while one-on-one connections are important and are the primary way to communicate with people in Twitter, people have found way to connect with communicates on a much larger scale.

Word clouds demonstrated that the content of tweets leading up to the conference converged. Two weeks before the conference, there were only a few clues that indicated people were going to the conference. However, by the second week, words related to the conference became prominent. By the third week of the conference, the word 'ghc10' became all encompassing. This is to be expected because all the users we were studying were focused on and most likely at the conference. The Tweet contents are variable due to the personal and professional lives of the users. However, the content of the tweets related to the conference was ephemeral by the fifth week. Nevertheless, other data shows that this is not true, such as the conversations between the attendees and the continuing relationships.

Lurkers identified themselves by joining the list of attendees at the conference and then having no measurable activity. One of the main reasons users reported using Twitter is as a news source. Lurkers could have used Twitter as their news source and created an account to accomplish this. Lurkers could also have asked to join the list for visibility, and as a way to keep track of the list itself, which reinforces the use of Twitter as a news source.

Mentions went up in both amount and percentage regarding the amount of times they were used during the week of the conference. Mentions are a way to start or continue conversations in Twitter. This was the only measurement that showed both numbers increase. The percentage points doubled during the week of the conference. It would make sense that during a conference users would be having and trying to start conversations in order to foster new relationships.

Retweeting is a way for users to pass information to their followers. It is a vote that this is a good tweet and should be paid attention to. Even though Twitter has an official retweet button, some users still c retweet the older way. In review, I found a total of 2,749 official retweets, which equates to around 12% of tweets being official retweets. I found 6,687 retweets including the unofficial ones, which makes 31% of tweets being retweets. If there is an official mechanism to retweet, why do so many tweets continue to have older retweet styles? The main reason that I speculate is that because the official retweet style does not allow commenting, individuals choose the old style which does allow annotation and commenting on the original tweet. More importantly, the amount of retweets indicates that there is a community that is passing information on to their followers, a virtual word of mouth.

5.3 Community

Prior work has demonstrated that Twitter is useful for communities. For example, prior work found that it is useful for conferences (Reinhardt, Ebner et al. 2009), that habit, satisfaction, and perceived usefulness of Twitter affected the growth of Twitter (Barnes & Böhringer 2009), and that people use Twitter because it creates gratification when communicating with a community (Johnson & Yang 2009).

This is because Twitter has many features that encourage communities to interact. For example, the work of Honeycutt and Herring (2009) demonstrated that people being able to have conversations through microblogging bonded the community. Additionally, the role of relationships (i.e., the use of followers and friends in Twitter) was directly linked to people interacting with one another (Blood 2004; Lento, Welser et al. 2006), and the ability to retweet a person's tweet allowed people to engage in information transferring (Liere 2010).

This prior work demonstrates that people have many tools to engage in the community of their social network. It also shows that these tools influence how people engage and interact, which is also reflected by the data in this study.

One of the main goals for attending a conference is networking. In order to network new relationships, communication must occur. Users employ Twitter as a medium for having conversations, therefore making it an important subject to study for this research. It is clear from the results that if communication around a subject is desired, a conference is the way to achieve the intended effect. The participants of the study started most of the conversations during the week of the conference. The week before and the week after had more conversations than the first or last week of the study. Therefore, most of the conversation between participants happened in the time during and around the conference. Even though there are long term effects of building community through conferences, most community building is done during the time period of a conference. In terms of activity alone, the number of pieces of conversational involvement increased with each group. Group 1 had the least and Group 4 had the most, as was expected.

This means that conferences have a large effect on electronic communication. While this is to be expected, it is something that was unthought-of ten years ago. The question becomes, how does this online interaction affect the offline interaction? And, how does offline interaction affect online interaction? The peak of interaction online and offline the week of the conference indicates that collocation critically impacts online interaction. Additionally, online interaction (particularly between people who did not know each other previously) can then reciprocally impact offline interaction. People may meet for the first time because of retweeting; people may recognize each other from retweets.

As activity on Twitter increased, the length of the conversations shortened. In the study, time was measured between the first tweet and the last tweet in a conversation thread. A conversation was defined as having a least one reply. This could be because the more active users are regularly on Twitter, and they can then engage in conversations much quicker. For example, Group 2 (Part-timers) in the fifth week took on an average of 137 hours to have a conversation.

The length of conversations is relevant when thinking about tool support for conferences. This is because conferences encourage networking and communication. If online conversations have a delay of 137 hours the purpose of networking is muted. Tool designers should think of new ways to encourage networking and decrease conversation time, especially for the types of users in Group 1 and Group 2 after the conferences.

Social networks can indicate what is happening with a group of users in the physical space. Conferences are designed to intentionally bring people together and for them to form connections. It is important to study what happens on Twitter in relation to the network to see if the main purpose of the conference is at least being accomplished in the virtual world. From the results, we can see there is a definite effect of the graph during the week of the conference. Participants created more edges in that week than in all the other weeks combined. Also, the number of communities increased, displaying the fact that the participants are less reliant on fewer individuals, which can be also compared with the number of strongly connected components and modularity scores.

The roles of the @ghc and @anitaborg_org are vital to the network studied³⁹. This is demonstrated in numerous data point as the results, such as having the highest betweenness centrality scores by an order of magnitude, the largest number of followers and friends of conference attendees, and the visual depictions of the graphs themselves. The two Twitter accounts work at creating connections between the users by only connecting people via the accounts themselves, but also by making connections between people via @ghc and @anitaborg_org actions on Twitter such as retweeting, replying, and creating lists. These accounts also allow the lurkers to follow these accounts and use them as news sources. The data

³⁹ The same individual is responsible for both of these accounts

shows that if I manually remove these two users from the social graph the number of distinct communities in the network would jump from the tens into the hundreds because of a large fragmentation of the network. It is important for organizers who want to create a strong community around them on Twitter to have an active user there.

5.4 Hashtags

Conference organizers have attempted to create complex and simple hashtag schemes. The Grace Hopper conference created a list of hashtags and publicized them heavily, but the only heavily used hashtag from the participants was '#ghc10.' Table 4 shows that even though large percentages in the list are conference hashtags, some emergent hashtags beat out the heavily publicized ones. In my opinion, it is not worth the time to create complex schema for different sessions, but instead create a hashtag for the conference and then sub-hashtags for large events at the conference, such as '#keynote.' Events are the driving force being Twitter usage.

6 Conclusion

In this thesis, I have presented an analysis of microblogging at a conference. Microblogging is a growing phenomenon where people post short updates to a list of followers. Through microblogging, people provide information about their lives, their professions, and have conversations with other microbloggers.

In this thesis I have focused: (1) on the specific microblogging platform, Twitter; and, (2) on a specific conference, the Grace Hopper Conference. Tweeting and Twitter have become the fastest growing microblogging platform, and it has fundamentally shifted the way people relay and communicate (Kwak, Lee et al. 2010).

We collected this data as an exploratory study of a new phenomenon. This means that the study did not have any strict guiding hypothesis. The main hypothesis instead was an assumption that collocation did affect online and offline interaction. This means that the methods used to evaluate this loose hypothesis were exploratory.

We used two sources of data and analyzed them many different ways to understand Twitter use at conferences. We have divided this analysis and the related work into topic areas where findings were interrelated. These sections are facts about tweeples, or people who tweet, content and description of tweets, or an evaluation of what is included inside a microblog, communities and interactions, or the effect of a large group of people interacting and communicating, and, tools.

In summary, all data supports that collocation had a large impact on offline and online interaction. We expected this confirmation, especially when viewed through the lens of Communication Theory (Ruggiero 2000). Communication Theory states that users' needs will influence the media that they choose. Additionally, it states that active participants in the information exchange process media will compete for attention.

The question of whether or not microblogging is important, when viewed this way, becomes irrelevant. Instead, the question becomes: how are tweets adding to the enjoyment of online and offline interaction as an information source? This question, when reviewing the data presented in this thesis, is critical. The data demonstrates that people find unique ways to participate and engage with the community that they belong to.

Conferences, particularly ones like the Grace Hopper Conference, that span both personal and professional boundaries, are nothing but streams of information sharing. Yet, as the data suggests, users want new and valuable ways to share and engage with what is happening at conferences.

6.1 Future Work

First, the content analysis of the tweets in this study was limited to word clouds. While word clouds provided a thin overview of the content of the data, they are insufficient. We suggest a

thorough reading of a random subset of the tweets. This reading could lead to a collection of unique codes through the use of Grounded Theory. Also, independent researchers could then code the remaining body of tweets in the hopes of inter-rater reliability.

Second, we did this study with a metaphorical bird's eye view of the data and the activities of users. We would like to shadow a subset of users at a conference to study when they tweet and study what prompts them to tweet. We could also ask participants what they gathered from reading their Twitter feed. This kind of focused study would help answer specific questions like, "Why did someone decide to follow that person?" and "Why does this user find that quote salient?" in addition, this is really the only way to measure the impact of collocation on Twitter.

Third, researchers should study additional conferences using the same method. For example, researchers should study larger and smaller conferences to see what affect conference size has on behaviors. Additionally, researchers could study conferences with varying focuses. The conference that we studied was mostly female and technical. This means that future work could explore conferences that are non-technical or predominantly male. This would validate that the results from this study would also be consistent across all conferences.

Finally, researchers should consider tools as a form of design probe to try and further understand this phenomenon. For example, researchers and designers should consider how to automatically create and display lists of attendees. Or, as another example, how the use of ambient displays of tweets might affect the adoption and use of microblogging tools. As has been shown by this exploratory study into the effects of conferences on tweeple, there is more for researchers to study in regard to this topic.

7 References

- (2010). "History of the Conference." Retrieved 2/3/2011, 2011, from <http://gracehopper.org/2011/about/history-of-the-conference/>.
- (2010, 4/15/2010). "Twitter Donates Entire Tweet Archive to Library of Congress." Retrieved 2/6/2011, 2011, from <http://www.loc.gov/today/pr/2010/10-081.html>.
- Bakshy, E., J. M. Hofman, W. A. Mason and D. J. Watts (2011). Everyone's an influencer: quantifying influence on twitter. Proceedings of the fourth ACM international conference on Web search and data mining. Hong Kong, China, ACM: 65-74.
- Barnes, S. J. and M. Böhringer (2009). Continuance Usage Intention In Microblogging Services: The Case of Twitter. Proceedings of the 17th European Conference on Information Systems (ECIS). Verona, Italy.
- Bennett, K., P. Layzell, D. Budgen, P. Brereton, L. Macaulay and M. Munro (2000). Service-based software: the future for flexible software. Proceedings of the Seventh Asia-Pacific Software Engineering Conference, IEEE Computer Society: 214.
- Blood, R. (2004). How blogging software reshapes the online community. Commun. ACM.
- boyd, d. (2010). "spectacle at Web2.0 Expo... from my perspective." Retrieved 01 January 2010, from http://www.zephoria.org/thoughts/archives/2009/11/24/spectacle_at_we.html.
- boyd, d., S. Golder and G. Lotan (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. System Sciences (HICSS), 2010 43rd Hawaii International Conference on.
- Brin, S. and L. Page (1998). The anatomy of a large-scale hypertextual Web search engine. Proceedings of the seventh international conference on World Wide Web 7. Brisbane, Australia, Elsevier Science Publishers B. V.: 107-117.
- Brooks, A. C., E.F. (2010). Tune in, Tweet on, Twit out: Information Snacking on Twitter. CHI 2010 Workshop on Microblogging
- Buzeck, M. and J. Müller (2010). TwitterSigns: microblogging on the walls. Proceedings of the international conference on Multimedia. Firenze, Italy, ACM: 819-822.
- Cha, M., H. Haddadi, F. Benevenuto and K. P. Gummadi (2010). "Measuring user influence in Twitter: The million follower fallacy." Proceedings of the Fourth International AAAI Conference on Weblogs and Social.
- Cheong, M. and V. Lee (2009). Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. Proceeding of the 2nd ACM workshop on Social web search and mining. Hong Kong, China, ACM: 1-8.
- Cormode, G. and B. Krishnamurthy (2008). "Key Differences between Web1.0 and Web2.0." First Monday Volume 13 Number 6.
- Cummings, J. N., B. Butler and R. Kraut (2002). "The quality of online social relationships." Commun. ACM **45**(7): 103-108.
- Facebook. (2011). "Statistics." Retrieved 02/20/2011, 2011, from <http://www.facebook.com/press/info.php?statistics>.
- Galuba, W., K. Aberer, D. Chakraborty, Z. Despotovic and W. Kellerer (2010). Outtweeting the twitterers - predicting information cascades in microblogs. Proceedings of the 3rd conference on Online social networks. Boston, MA, USENIX Association: 3-3.
- Golovchinsky, G. and M. Efron (2010). Making sense of Twitter Search.
- Greengard, S. (2009). "The first internet president." Commun. ACM **52**(2): 16-18.

- Hansen, D., B. Shneiderman and M. Smith (2010). Analyzing Social Media Networks with NodeXL: Insights from a Connected World, Morgan Kaufmann.
- Hazlewood, W. R., W. Ryan and K. Makice (2008). Twitterspace: a co-developed display using Twitter to enhance community awareness. Proceedings of the Tenth Anniversary Conference on Participatory Design 2008. Bloomington, Indiana, Indiana University: 230-233.
- Honeycutt, C. and S. C. Herring (2009). Beyond Microblogging: Conversation and Collaboration via Twitter.
- Huberman, B., D. M. Romero and F. Wu (2009). "Social networks that matter: Twitter under the microscope." First Monday Volume 14, Number 1.
- Java, A., X. Song, T. Finin and B. Tseng (2007). Why we twitter: understanding microblogging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. San Jose, California, ACM: 56-65.
- Johnson, P. and S.-I. Yang (2009). Uses and Gratifications of Twitter: An Examination of User Motives and Satisfaction of Twitter User. Association for Education in Journalism and Mass Communications. Sheraton Boston, Moston, MA.
- Kempe, D., J. Kleinberg and É. Tardos (2003). Maximizing the spread of influence through a social network. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, D.C., ACM.
- Khonsari, K. K., Z. A. Nayeri, A. Fathalian and L. Fathalian (2010). Social Network Analysis of Iran's Green Movement Opposition Groups Using Twitter. Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining, IEEE Computer Society: 414-415.
- Knoke, D. and S. Yang (2008). Social Network Analysis (Quantitative Applications in the Social Sciences), Sage Publications, Inc.
- Krazit, T. (2010). "Watch What You Tweet: Google's Archiving." CBS News Retrieved 2/19/2011, 2011, from <http://www.cbsnews.com/stories/2010/04/14/tech/cnettechnews/main6395383.shtml>.
- Krishnamurthy, B. (2009). A measure of online social networks. Proceedings of the First international conference on COMMunication Systems And NETWORKS. Bangalore, India, IEEE Press: 190-199.
- Kumar, R., J. Novak and A. Tomkins (2006). Structure and evolution of online social networks. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. Philadelphia, PA, USA, ACM: 611-617.
- Kwak, H., C. Lee, H. Park and S. Moon (2010). What is Twitter, a social network or a news media? Proceedings of the 19th international conference on World wide web. Raleigh, North Carolina, USA, ACM: 591-600.
- Lee, R. and K. Sumiya (2010). Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks. San Jose, California, ACM: 1-10.
- Lento, T., H. T. Welser and L. Gu (2006). The Ties that Blog: Examining the Relationship Between Social Ties and Continued Participation in the Wallop Weblogging System
- Leskovec, J., J. Kleinberg and C. Faloutsos (2005). Graphs over time: densification laws, shrinking diameters and possible explanations. Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. Chicago, Illinois, USA, ACM: 177-187.

- Leskovec, J., J. Kleinberg and C. Faloutsos (2007). "Graph evolution: Densification and shrinking diameters." ACM Trans. Knowl. Discov. Data **1**(1): 2.
- Liere, D. v. (2010). How far does a tweet travel?: Information brokers in the twitterverse. Proceedings of the International Workshop on Modeling Social Media. Toronto, Ontario, Canada, ACM: 1-4.
- Makice, K. (2009). Phatics and the design of community. Proceedings of the 27th international conference extended abstracts on Human factors in computing systems. Boston, MA, USA, ACM: 3133-3136.
- McNely, B. (2009). Backchannel persistence and collaborative meaning-making. Proceedings of the 27th ACM international conference on Design of communication. Bloomington, Indiana, USA, ACM: 297-304.
- McPherson, M., L. Lovin and J. Cook (2001). "Birds of a Feather: Homophily in Social Networks." Annual Review of Sociology **27**(1): 415-444.
- Michelson, M. and S. A. Macskassy (2010). Discovering users' topics of interest on twitter: a first look. Proceedings of the fourth workshop on Analytics for noisy unstructured text data. Toronto, ON, Canada, ACM: 73-80.
- Müller, J., J. Exeler, M. Buzeck and A. Krüger (2009). ReflectiveSigns: Digital Signs That Adapt to Audience Attention. Pervasive Computing. H. Tokuda, M. Beigl, A. Friday, A. Brush and Y. Tobe, Springer Berlin / Heidelberg. **5538**: 17-24.
- Naaman, M., J. Boase and C.-H. Lai (2010). Is it really about me?: message content in social awareness streams. Proceedings of the 2010 ACM conference on Computer supported cooperative work. Savannah, Georgia, USA, ACM: 189-192.
- Oulasvirta, A., E. Lehtonen, E. Kurvinen and M. Raento (2010). "Making the ordinary visible in microblogs." Personal Ubiquitous Comput. **14**(3): 237-249.
- Perez, S. (2009). "10 Ways to Archive Your Tweets." ReadWriteWeb Retrieved 2/19/2011, 2011, from http://www.readwriteweb.com/archives/10_ways_to_archive_your_tweets.php.
- Qiu, T., J. Feng, Z. Ge, J. Wang, J. Xu and J. Yates (2010). Listen to me if you can: tracking user experience of mobile network on social media. Proceedings of the 10th annual conference on Internet measurement. Melbourne, Australia, ACM: 288-293.
- Reinhardt, W., M. Ebner, G. Beham and C. Costa (2009). How People are using Twitter during Conferences. Proceeding of 5. EduMedia Conferece. Salzburg: 145-156.
- Rejaie, R., M. Torkjazi, M. Valafar and W. Willinger (2010). "Sizing up online social networks." Netwrk. Mag. of Global Internetwkg. **24**(5): 32-37.
- Rodriguez, M. G., J. Leskovec and A. Krause (2010). Inferring networks of diffusion and influence. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. Washington, DC, USA, ACM: 1019-1028.
- Rott, N. (2010, 12/20/2010). "Planning A Flash Mob? Better Keep It Quiet." Digital Life Retrieved 3/20/2011, 2011, from <http://www.npr.org/2010/12/20/132205587/planning-a-flash-mob-better-keep-it-quiet>.
- Ruggiero, T. (2000). "Uses and Gratifications Theory in the 21st Century." Mass Communication & Society **3**(1): 3-37.
- Sharma, P. (2008, 10/28.2008). "Core Characteristics of Web 2.0 Services." TechPluto Retrieved 2/20/2011, 2011, from <http://www.techpluto.com/web-20-services/>.
- Smith, M., V. Barash, L. Getoor and H. W. Lauw (2008). Leveraging social context for searching social media. Proceeding of the 2008 ACM workshop on Search in social media. Napa Valley, California, USA, ACM: 91-94.

- Smith, M., D. L. Hansen and E. Gleave (2009). Analyzing Enterprise Social Media Networks. Computational Science and Engineering, 2009. CSE '09. International Conference on.
- Smith, M. A., B. Shneiderman, N. Milic-Frayling, E. M. Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer and E. Gleave (2009). Analyzing (social media) networks with NodeXL. Proceedings of the fourth international conference on Communities and technologies. University Park, PA, USA, ACM: 255-264.
- Sutton, J., L. Palen and I. Shklovski (2008). Backchannels on the front lines: Emergent use of social media in the 2007 Southern California fire. Information Systems for Crisis Response and Management Conference (ISCRAM), Washington DC.
- Swift, M. (2011, 02/14/2011 12:16:00 AM PST). "LinkedIn, Facebook now let you map your social connections." Retrieved 03/02/2011, 2011.
- Weil, K. (2011, Monday, February 22, 2010). "Measuring Tweets." Retrieved 2/6/2011, 2011, from <http://blog.twitter.com/2010/02/measuring-tweets.html>.
- Weng, J., E.-P. Lim, J. Jiang and Q. He (2010). TwitterRank: finding topic-sensitive influential twitterers. Proceedings of the third ACM international conference on Web search and data mining. New York, New York, USA, ACM: 261-270.
- Westman, S. and L. Freund (2010). Information interaction in 140 characters or less: genres on twitter. Proceeding of the third symposium on Information interaction in context. New Brunswick, New Jersey, USA, ACM: 323-328.
- Yoshioka, T., G. Herman, J. Yates and W. Orlikowski (2001). "Genre taxonomy: A knowledge repository of communicative actions." ACM Trans. Inf. Syst. **19**(4): 431-456.

8 Appendix A: IRB



VirginiaTech

Office of Research Compliance
Institutional Review Board
2000 Kraft Drive, Suite 2000 (0497)
Blacksburg, Virginia 24060
540/231-4606 Fax 540/231-0959
e-mail irb@vt.edu
Website: www.irb.vt.edu

MEMORANDUM

DATE: September 23, 2010

TO: Manuel A. Perez-Quinones, Edgardo Vega, Andrea L. Kavanaugh, Steve Harrison

FROM: Virginia Tech Institutional Review Board (FWA00000572, expires June 13, 2011)

PROTOCOL TITLE: Twitter Use at Conferences

IRB NUMBER: 10-335

Effective September 23, 2010, the Virginia Tech IRB Chair, Dr. David M. Moore, approved the amendment request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report promptly to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at <http://www.irb.vt.edu/pages/responsibilities.htm> (please review before the commencement of your research).

PROTOCOL INFORMATION:

Approved as: **Expedited, under 45 CFR 46.110 category(ies) 6, 7**

Protocol Approval Date: **9/10/2010**

Protocol Expiration Date: **9/9/2011**

Continuing Review Due Date*: **8/26/2011**

*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals / work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

Invent the Future

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
An equal opportunity, affirmative action institution

9 Appendix B: Survey

Section 1: Informed Consent

Investigators: Dr Manuel A Perez-Quinonez and Edgardo Vega

I. Purpose of this Research

The purpose of this research is to investigate users of Twitter, before, during, after a specific event we hope to be able to discover a pattern in postings. Examining these patterns will facilitate a better understanding user practices and guide the design of current and future systems in this space.

II. Procedures

After agreeing to this informed consent form, you will asked a series of questions. There are two sections after this one. First is demographic information and the second is on your Twitter usage.

The survey can be interrupted for job related tasks. Without interruption we anticipate this interview to take between 20 and 30 minutes. At the end of the survey you will be thanked for your time.

III. Risks

The only known risk in this study is legal. We do not wish to see any information that is confidential or illegal to share. However, all information and audio will be stored in a password-protected computer. Files that are printed will only be named with your participant number that is written at the top right of this form.

IV. Benefits

The long term benefits of this study is that the information you provide will facilitate a better understanding user practices and guide the design of current and future systems in this space. This will help in providing better system that can be used in providing more value to going to conferences.

However, no promise or guarantee of benefits has been made to encourage you to participate in this research. If at a later date, if you are interested in the results of this study, you can contact the researches listed below for a summary.

V. Extent of Anonymity and Confidentiality

At the beginning of the study you will be assigned a random participant number to assure your confidentiality. You can see this number at the top of the informed consent. There are a few places you identification will be stored. One is on our copy of this form. The second will be in a digital file that associates your name with your participant number. These file will all digitized and secured under a password, and the informed consents will be stored in a locked file cabinet under the protection of the investigators. At no time will the researchers release identifying information to anyone other than individuals working on the project without your written consent.

It is possible that the Institutional Review Board (IRB) may view this study's collected data for auditing purposes. The IRB is responsible for the oversight of the protection of human subjects involved in research.

Unless the participant reveals information regarding abuse to him/herself and/or others, this confidentiality will be maintained under all circumstances.

VI. Compensation

If you choose to opt into the lottery, by entering any piece of contact information that we could use to contact you if you win (e.g phone, email, twitter name, etc). You could win 50 dollars. Each user will be place in a number list and a random number generator will choose the winner. This will be done transparently in front of a group of people (minimum of 3), chosen by the PI and videotaped. This drawing will happen two Mondays after the conference ends to allow plenty of time for the survey to be filled out by all those who wish to take it. You may contact the researches listed below for the video.

VII. Freedom to Withdraw

You are free to withdraw from a study at any time without penalty. You are free not to answer any questions or respond to experimental situation that you choose without penalty.

There may be circumstances under which the investigator may determine that you should not continue as a subject. You will be thanked for your time.

VIII. Subject's Responsibilities

I voluntarily agree to participate in this study. I have the following responsibilities: to let the experimenter know if I am feeling overly frustrated and need to take a break; to let the experimenter know that I need to leave the study.

IX. Contact

Should I have any pertinent questions about this research or its conduct, and research subjects' right, and whom to contact in the event of a research-related injury to the subject, I may contact:

Edgardo Vega, Dr Manuel A Perez-Quinonez (Faculty Advisor Computer Science Department

2202 Kraft Drive

Blacksburg, VA. 24060

{evega, perez}@cs.vt.edu,

Phone: 540-231-7409

Dr. Ryder

Department Head, Computer Science Department

2202 Kraft Drive

Blacksburg, VA 24060

Ryder@cs.vt.edu

Phone: 540.231.8452

David M. Moore

Chair, Virginia Tech Institutional Review

Board for the Protection of Human Subjects

Office of Research Compliance

2000 Kraft Drive, Suite 2000 (0497)

Blacksburg, VA. 24060

540.231.4991

MooreD@vt.edu

X. Subject's Permission

I have read the Consent Form and conditions of this project. I have had all of my questions answered. I hereby acknowledge the above and give my voluntary consent:

Yes/No

Section 2: Demographics

What is your primary role at the conference?

Speaker

Attendee

Organizer

Attending Virtually

Other

What is the highest degree or level of school you have completed?

some high school will there really be middle school kids or high school drop outs at this conference?

high school graduate

some college

trade/technical/vocational training

college graduate

some postgraduate work

post graduate degree

Where do you currently reside?

USA

Africa

Canada

South America/Mexico

Western Europe

Eastern Europe

Asia

Other

What is your age?

18-25

26-35

36-45

46-55

56-65

66+

What is your gender?

Male

Female

No answer

What is your race?

American Indian or Alaska Native

Asian

Black or African American

Native Hawaiian or Other Pacific Islander

White

Hispanic, Latino or Spanish origin (may be of any race)

Is your race is Hispanic, Latino or Spanish origin

No, not of Hispanic, Latino, or Spanish origin

Yes, Mexican, Mexican Am., Chicano

Yes, Puerto Rican

Yes, Cuban

Yes another Hispanic, Latino, or Spanish origin for example, Argentinean, Columbian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.

Section 3: Twitter

I. Current Twitter Use

How did you start using Twitter? (e.g., someone told me about it, I wanted to try something new)

About how many people do you follow on twitter?

How do you decide whom to follow on Twitter?

About how many people follow you on twitter?

Do you have more than one twitter account?

Yes

No

If you have more than one Twitter account, explain why? Please explain the differences in use.

Do you primarily use your twitter account for personal purposes or professional purposes?

Personal Purposes

Professional Purposes

Both

Why do you use Twitter (Select All the Apply)?

Personal Communications

Professional Communications

Networking (overall)

News Source (current events, industry news, etc.)

Other (please specify)

What is the MAIN reason that you use Twitter?

Personal Communications

Business Communications

Networking (overall)

News Source

Industry News Source

Other

What do you primarily use twitter for?

Share Resources

Communicate with others

Participate in Parallel Discussion

Find news and information

Take Notes

Establish Online Presence

Pose organization Questions [what does this mean?]

How long have you had a twitter account?

Less than a month;

One month to six months;

six months to a year;
more than one year.

How frequently do you tweet on a typical day?

0-10

11-25

26-60

61-99

100-150

151-200

201+

II. Conference Twitter Use

Did you have a twitter account before coming to the conference?

Yes

No

Did you have a twitter account before registering for the conference?

Yes

No

Did you create a twitter because you were attending this conference?

Yes

No

Have you actively used twitter to communicate during the conference?

Yes

No

If you have more than one account do you use them differently at a [THIS?] conference? How?

While at the conference do you use Twitter primarily to communicate with individuals at the conference or outside the conference?

At the conference

Outside the conference

Both

Please estimate how many times you tweet during {THIS} conference per day]? In as separate question you can ask “has your twitter use at this conference differed from your twitter use at previous conferences?” If so, how?

0-10

11-25

26-60

61-99

100-150

151-200

201+

Have you discussed topics with other other users of Twitter (Tweeple)by sending direct messages?

Yes (Why?)

No (Why not?)

Have you discussed topics with other other users of Twitter (Tweeple) by replying to tweets?

Yes (Why?)

No (Why not?)

How have you been primarily using twitter during the conference so far?

Sharing Resources

Communicate with others

Participate in Parallel Discussion

Jot down Notes

Establish Online Presence

Pose organization Questions

What is your primary reason for using Twitter during the conference? (this is not different from the preceding question)

Sharing Resources

Communicate with others

Participate in Parallel Discussion

Jot down Notes

Establish Online Presence

Pose organization Questions

What is your primary method for interacting with Twitter?

Desktop Client

Twitter.com

Mobile Device Application

What is your primary method for interacting with Twitter at [this] conference?

Desktop Client

Twitter.com

Mobile Device Application

Please select the operating system that your smart phone uses

Blackberry OS

iPhone OS

Android OS

Other Smartphone _____

Don't have a smartphone

If using your mobile device to interact with Twitter which of the following do you use primarily?

Text Messaging

An application

m.twitter.com

Why have you been using Twitter at this conference (Select All the Apply)?

Personal Communications

Business Communications

Networking (overall)

News Source

Industry News Source

Other

What is the main reason you have been using Twitter at this conference?

Personal Communications

Business Communications

Networking (overall)

News Source

Industry News Source

Other

Have you used hashtags in Twitter?

Yes

No

Are you using them currently?

Yes

No

Are you following the conference hashtag (e.g. #GHC10)?

Yes

No

If you follow the conference hashtag (e.g. #GHC10) how do you filter what is and is not important?

Do you use hashtags outside of this conference? If so, how?

What language do you primarily tweet in?

Have you tweeted at other conferences? If yes, What language do you primarily tweet in at conferences (like this one)?

Do you ever read the (global) public timeline?

Yes

No

If yes, why ?

with all the different communication media, such as Twitter, connecting people, do you think it is still important to meet face to face at a conference like this?

Yes

No

If yes, why?

10 Appendix C: Code

This code used Twitterizer version 2.3 and a SQL backend. The code interacts with the database by using LINQ to SQL.

```
using System;
using System.Collections.Generic;
using System.Diagnostics;
using System.Linq;
using System.Net;
using System.Runtime.Serialization.Json;
using System.Text;
using System.Timers;
using System.Web;
using System.Net.Mail;

namespace TwitterCollection
{
    class Program
    {
        public const string CONSUMER_KEY = "";
        public const string CONSUMER_SECRET = "";
        public const string ACCESS_TOKEN = "";
        public const string ACCESS_TOKEN_SECRET = "";

        static Timer t;
        const int CHECK_INTERVAL = 1; //in minutes
        static string body = string.Empty;

        static void Main(string[] args)
        {
            Twitterizer.OAuthTokens oAuthTokens = new Twitterizer.OAuthTokens();
            oAuthTokens = GetTokens();

            try
            {
                t = new Timer();
                t.Interval = CalculateNextExactMin();
                t.Elapsed += new ElapsedEventHandler(t_Elapsed);
                t.Start();
            }
            catch (Exception ex)
            {
                Console.WriteLine(ex.ToString());
                SendMail(0, "Could not start in main: " + ex.ToString());
            }
        }
    }
}
```

```

        while (Console.Read() != 'q')
        {
            ; // do nothing...
        }

    }

    static void SendMail(int cycle, string body)
    {
        try
        {
            MailMessage message = new MailMessage("edgardo.vega@gmail.com", "edgardo.vega@gmai
l.com", "Cycle" + cycle, body);
            SmtplibClient smtp = new SmtplibClient("smtp.gmail.com");
            smtp.EnableSsl = true;
            System.Net.NetworkCredential SMTPUserInfo = new System.Net.NetworkCredential("edga
rdo.vega@gmail.com", "");
            smtp.UseDefaultCredentials = false;
            smtp.Credentials = SMTPUserInfo;

            smtp.Send(message);
        }
        catch (Exception ex)
        {
            Console.WriteLine(ex.Message);
        }
    }

    static void t_Elapsed(object sender, ElapsedEventArgs e)
    {
        try
        {
            t.Stop();

            //Get all users in list
            //Get all users from ghc10#
            //Contently get all ghc10 tweets
            //go through all users and get followers and following userid
            Twitterizer.OAuthTokens oAuthTokens = new Twitterizer.OAuthTokens();
            oAuthTokens = GetTokens();

            int nextCycle = GetNextCycle();
            Console.WriteLine("Cycle" + nextCycle.ToString());
            body += "Cycle" + nextCycle.ToString() + Environment.NewLine;

            SendMail(nextCycle, "Cycle Started");

            Console.WriteLine("Get List Users Start");
            body += "Get List Users Start" + Environment.NewLine;

            GetListUsers(oAuthTokens, nextCycle);

            Console.WriteLine("Get List Users End");
            body += "Get List Users End" + Environment.NewLine;
            Console.WriteLine("Process List Users Start");
            body += "Process List Users Start" + Environment.NewLine;

            ProcessListUser(oAuthTokens, nextCycle);
        }
    }
}

```

```

        Console.WriteLine("Process List Users End");
        body += "Process List Users End" + Environment.NewLine;

        TwitterCollectionDataContext db = new TwitterCollectionDataContext();
        Cycle cycle = (from c in db.Cycles
                      where c.id == nextCycle
                      select c).Single();

        cycle.ended = DateTime.Now;
        Console.WriteLine("End Cycle" + nextCycle.ToString() + " " + DateTime.Now.ToShortT
imeString());
        db.SubmitChanges();

        //Search
        Console.WriteLine("Search Start");
        body += "Search Start" + Environment.NewLine;
        Search(oAuthTokens, nextCycle);

        Console.WriteLine("Search Stop");
        body += "Search Start" + Environment.NewLine;

        t.Interval = CalculateNextExactHour();
        t.Start();
        SendMail(nextCycle, "Cycle Ended" + Environment.NewLine + "RunTime: " + (cycle.ended - cycle.started).ToString() + Environment.NewLine + body);
        body = string.Empty;
    }
    catch (Exception ex)
    {
        Console.WriteLine("Error: " + ex.ToString());

        t.Interval = CalculateNextExactMin();
        SendMail(-1, "Error: " + ex.ToString() + Environment.NewLine + body);
        body = string.Empty;
        t.Start();
    }
}

static void Search(Twitterizer.OAuthTokens oAuthTokens, int cycle)
{
    TwitterCollectionDataContext db = new TwitterCollectionDataContext();

    Twitterizer.SearchOptions s0 = new Twitterizer.SearchOptions();
    s0.ResultType = Twitterizer.SearchOptionsResultType.Recent;
    s0.NumberPerPage = 100;
    //find max ID of status from db
    try
    {
        var maxID = (from s in db.SearchResults
                    orderby s.id descending
                    select s.id).First();

        s0.SinceId = Convert.ToInt64(maxID);
    }
    catch (Exception)
    {
    }
}

```

```

        string searchQuery = GetSearchQuery();

        Twitterizer.TwitterResponse<Twitterizer.TwitterSearchResultCollection> searchResponse
= new Twitterizer.TwitterResponse<Twitterizer.TwitterSearchResultCollection>();

        Twitterizer.TwitterSearchResultCollection searchResults = new Twitterizer.TwitterSearchResultCollection();

        var tweetsPerPage = 100;

        var ceiling = Math.Ceiling(1500m / tweetsPerPage);

        // Twitter API allows up to 1500 total results
        for (var i = 1; i <= ceiling; i++)
        {
            sO.PageNumber = i;
            RetryUtility.TwitterRetryAction(() => searchResponse = Twitterizer.TwitterSearch.Search(oAuthTokens, searchQuery, sO));

            if (searchResponse.Result != Twitterizer.RequestResult.Success)
            {
                RetryUtility.TwitterRetryAction(() => searchResponse = Twitterizer.TwitterSearch.Search(oAuthTokens, searchQuery, sO));
            }

            foreach (var searchResult in searchResponse.ResponseObject)
            {
                searchResults.Add(searchResult);
            }

            if (searchResponse.ResponseObject.Count() != 100)
            {
                break;
            }
        }

        // 100 tweets per page, 15 page ceiling, 1500 total result statuses
        Console.WriteLine("Search Results Found: " + searchResults.Count);
        body += "Search Results Found: " + searchResults.Count + Environment.NewLine;

        foreach (var searchResult in searchResults)
        {
            SearchResult sR = new SearchResult();

            sR.CreatedDate = searchResult.CreatedDate;

            sR.FromUserScreenName = searchResult.FromUserScreenName;
            sR.id = searchResult.Id;
            sR.Language = searchResult.Language;
            sR.Location = searchResult.Location;
            sR.Source = searchResult.Source;
            sR.Text = searchResult.Text;

            sR.ToUserScreenName = searchResult.ToUserScreenName;

            db.SearchResults.InsertOnSubmit(sR);
        }
    }
}

```

```

        db.SubmitChanges();
    }
    catch (Exception ex)
    {
        Console.WriteLine("Search Error: " + ex.Message);
        body += "Search Error: " + ex.Message + Environment.NewLine;
    }
}

static string GetSearchQuery()
{
    string query = string.Empty;
    string hashtag = "#";

    List<string> keywords = new List<string>();
    keywords.Add(hashtag + "ghc10");
    keywords.Add(hashtag + "acad");
    keywords.Add(hashtag + "award");
    keywords.Add(hashtag + "career");
    keywords.Add(hashtag + "info");
    keywords.Add(hashtag + "pract");
    keywords.Add(hashtag + "ugrad");
    keywords.Add(hashtag + "grad");
    keywords.Add(hashtag + "early");
    keywords.Add(hashtag + "indst");
    keywords.Add(hashtag + "invtech");
    keywords.Add(hashtag + "k12");
    keywords.Add(hashtag + "ldrshp");
    keywords.Add(hashtag + "sci");
    keywords.Add(hashtag + "steer");
    keywords.Add(hashtag + "stdnts");
    keywords.Add(hashtag + "exec");

    keywords.Add(hashtag + "hci");
    keywords.Add(hashtag + "osrc");
    keywords.Add(hashtag + "tech");
    keywords.Add(hashtag + "theme");
    keywords.Add(hashtag + "build");
    keywords.Add(hashtag + "sci");

    keywords.Add(hashtag + "phd1");
    keywords.Add(hashtag + "phd2");
    keywords.Add(hashtag + "phd3");
    keywords.Add(hashtag + "phd4");
    keywords.Add(hashtag + "newinv1");
    keywords.Add(hashtag + "newinv2");
    keywords.Add(hashtag + "newinv3");

    //bofs
    string bofs = string.Empty;
    for (int i = 1; i < 15; i++)
    {
        bofs += hashtag + "bof" + i + " OR ";
    }

    foreach (string key in keywords)
    {
        query += key + " OR ";
    }
}

```

```

        query += bofs;

        query = query.Remove(query.Length - 4, 4);
        return query;
    }
    static int CalculateNextExactMin()
    {
        DateTime now = DateTime.Now;
        DateTime nowPlus = now.AddMinutes(CHECK_INTERVAL);
        nowPlus = nowPlus.AddSeconds(-1 * nowPlus.Second);

        TimeSpan interval = nowPlus - now;
        Console.WriteLine("Next StartTime: " + nowPlus);
        body += "Next StartTime: " + nowPlus + Environment.NewLine;
        return (int)interval.TotalMilliseconds;
    }

    static int CalculateNextExactHour()
    {
        DateTime now = DateTime.Now;
        DateTime nowPlus = now.AddMinutes(60);
        nowPlus = nowPlus.AddSeconds(-1 * nowPlus.Second);
        nowPlus = nowPlus.AddMinutes(-1 * nowPlus.Minute);

        TimeSpan interval = nowPlus - now;

        Console.WriteLine("Next StartTime: " + nowPlus);
        body += "Next StartTime: " + nowPlus + Environment.NewLine;
        return (int)interval.TotalMilliseconds;
    }

    static int CalculateTimerInterval(int minute)
    {
        if (minute <= 0)
            minute = 60;
        DateTime now = DateTime.Now;

        DateTime future = now.AddMinutes((minute -
(now.Minute % minute))).AddSeconds(now.Second * -1).AddMilliseconds(now.Millisecond * -1);

        TimeSpan interval = future - now;

        return (int)interval.TotalMilliseconds;
    }

    static void UsersLookup(Twitterizer.OAuthTokens oAuthTokens, int cycle)
    {
        TwitterCollectionDataContext db = new TwitterCollectionDataContext();

        var alreadyfound = (from u in db.ExtendedInformations
                            where u.cycleID == cycle
                            select u.User.username).ToList();

        //should only get user who were removed from the list for some reason.
        var allusers = db.Users.Select(x => x.username).ToList();

        var users = allusers.Except(alreadyfound);

        Console.WriteLine("User Count: " + users.Count().ToString());
        body += "User Count: " + users.Count().ToString() + Environment.NewLine;
    }

```

```

int pageSize = 100;
int pages = Convert.ToInt32(Math.Ceiling(users.Count() / Convert.ToDouble(pageSize)));

List<string> pagedUsers;

for (int pageNum = 0; pageNum < pages; pageNum++)
{
    pagedUsers = users.Skip((pageNum * pageSize)).Take(pageSize).ToList();

    System.Collections.ObjectModel.Collection<string> screenNames = new System.Collect
ions.ObjectModel.Collection<string>(pagedUsers);
    Twitterizer.TwitterResponse<Twitterizer.TwitterUserCollection> lookup = new Twitte
rizer.TwitterResponse<Twitterizer.TwitterUserCollection>();

    RetryUtility.TwitterRetryAction(() => lookup = Twitterizer.TwitterUser.Lookup(oAut
hTokens, new Twitterizer.LookupUsersOptions { ScreenNames = screenNames }));

    foreach (var user in lookup.ResponseBody)
    {
        Console.WriteLine(user.ScreenName);
        body += user.ScreenName + Environment.NewLine;
        InsertUserInformation(user, cycle, false);
    }
}

}

static void ProcessListUser(Twitterizer.OAuthTokens oAuthTokens, int cycle)
{
    Console.WriteLine("User LookUp Started");
    body += "User LookUp Started" + Environment.NewLine;

    UsersLookup(oAuthTokens, cycle);
    Console.WriteLine("User LookUp End");
    body += "User LookUp End" + Environment.NewLine;

    TwitterCollectionDataContext db = new TwitterCollectionDataContext();

    //must exclude protected users to avoid errors
    var users = from e in db.ExtendedInformations
                where e.IsProtected == false
                && e.cycleID == cycle
                select e;

    //All the users and their friends the latest status of added or not
    var allFriendsHistory = from a in
                            (from f in db.Friends
                             group f by new { f.userID, f.FriendsID } into g
                             select new
                             {
                                 g.Key,
                                 cycleID = g.Max(f => f.cycleID)
                             })
                            join b in db.Friends
                            on new { a.cycleID, a.Key.userID, a.Key.FriendsID }
                            equals
                            new { b.cycleID, b.userID, b.FriendsID }
                            select new

```



```

        /*Followers*/
        Twitterizer.TwitterResponse<Twitterizer.UserIdCollection> Followers = new Twitterizer.TwitterResponse<Twitterizer.UserIdCollection>();
        RetryUtility.TwitterRetryAction(() => Followers = Twitterizer.TwitterFriendship.FollowersIds(oAuthTokens, new Twitterizer.UsersIdsOptions { UserID = u.userID }));

        var userFollowersAlreadyAdded = allFollowersHistory.Where(x => x.userID == u.userID && x.added == true).Select(s => s.FollowID);

        var toAddFollowers = Followers.ResponseObject.Except(userFollowersAlreadyAdded);
        foreach (var temp in toAddFollowers)
        {
            db.Followers.InsertOnSubmit(new Follower { userID = u.userID, FollowID = temp, cycleID = cycle, added = true });
        }

        //there users would need to be set as deleted
        var toDeleteFollowers = userFollowersAlreadyAdded.ToList().Except(Followers.ResponseObject);

        foreach (var temp in toDeleteFollowers)
        {
            db.Followers.InsertOnSubmit(new Follower { userID = u.userID, FollowID = temp, cycleID = cycle, added = false });
        }

        db.SubmitChanges();
    }
    Console.WriteLine();
    body += Environment.NewLine;
}

static void InsertUserInformation(Twitterizer.TwitterUser user, int cycle, bool attendee)
{
    TwitterCollectionDataContext db = new TwitterCollectionDataContext();
    var tempUsers = (from u in db.Users
                    where u.id == user.Id
                    select u);

    User tempUser = new User();
    ExtendedInformation ei = new ExtendedInformation();
    if (tempUsers.Count() != 1)
    {
        tempUser.attendee = attendee;
        tempUser.id = user.Id;
        tempUser.username = user.ScreenName;
        tempUser.discoveredCycleID = cycle;

        ei.IsProtected = user.IsProtected;
        ei.cycleID = cycle;
        ei.Location = user.Location;
        ei.NumberOfFriends = user.NumberOfFriends;
        ei.NumberOfFollowers = user.NumberOfFollowers;
        ei.Language = user.Language;
        ei.TimeZone = user.TimeZone;
        ei.NumberOfStatuses = user.NumberOfStatuses;
        ei.CreatedDate = user.CreatedDate;
        ei.ListedCount = user.ListedCount;

        tempUser.ExtendedInformations.Add(ei);

        db.Users.InsertOnSubmit(tempUser);
    }
}

```

```

    }
    else
    {
        ei.IsProtected = user.IsProtected;
        ei.cycleID = cycle;
        ei.Location = user.Location;
        ei.NumberOfFriends = user.NumberOfFriends;
        ei.NumberOfFollowers = user.NumberOfFollowers;
        ei.Language = user.Language;
        ei.TimeZone = user.TimeZone;
        ei.NumberOfStatuses = user.NumberOfStatuses;
        ei.CreatedDate = user.CreatedDate;
        ei.ListedCount = user.ListedCount;
        tempUsers.First().ExtendedInformations.Add(ei);
    }

    try
    {
        db.SubmitChanges();
    }
    catch (Exception ex)
    {
        Console.WriteLine("Error inserting user information: " + tempUsers.First().username
e + " " + cycle + " " + ex.Message);
        body += "Error inserting user information: " + tempUsers.First().username + " " +
cycle + " " + ex.Message + Environment.NewLine;
    }
}

static void GetListUsers(Twitterizer.OAuthTokens oAuthTokens, int cycle)
{
    Twitterizer.TwitterUserCollection allAttendees = new Twitterizer.TwitterUserCollection
());

    allAttendees = AllMembersinList(oAuthTokens, "ghc", "ghc10-attendees");

    var names = allAttendees.Select(x => x.ScreenName).Distinct();

    var jkad = from j in allAttendees
                where names.Contains(j.ScreenName)
                select j;

    Console.WriteLine("List Users: " + jkad.Count());
    body += "List Users: " + jkad.Count() + Environment.NewLine;
    foreach (Twitterizer.TwitterUser user in jkad)
    {
        InsertUserInfromation(user, cycle, true);
    }
}

static int GetNextCycle()
{
    int nextCycle = 1;
    TwitterCollectionDataContext db = new TwitterCollectionDataContext();

    var maxCycle = (from c in db.Cycles
                    orderby c.id descending
                    select c);

    if (maxCycle.Count() != 0)
    {

```

```

        nextCycle = maxCycle.First().id + 1;
    }

    Cycle cycle = new Cycle();

    cycle.id = nextCycle;
    cycle.started = DateTime.Now;

    db.Cycles.InsertOnSubmit(cycle);
    db.SubmitChanges();

    return nextCycle;
}

static Twitterizer.OAuthTokens GetTokens()
{
    Twitterizer.OAuthTokens oAuthTokens = new Twitterizer.OAuthTokens();

    oAuthTokens.AccessToken = ACCESS_TOKEN;
    oAuthTokens.AccessTokenSecret = ACCESS_TOKEN_SECRET;
    oAuthTokens.ConsumerKey = CONSUMER_KEY;
    oAuthTokens.ConsumerSecret = CONSUMER_SECRET;

    return oAuthTokens;
}

static Twitterizer.TwitterUserCollection AllMembersinList(Twitterizer.OAuthTokens oAuthTok
ens, string userName, string listName)
{
    Twitterizer.TwitterUserCollection usersInTheList = new Twitterizer.TwitterUserCollecti
on();
    Twitterizer.TwitterUserCollection distinctUsersInTheList = new Twitterizer.TwitterUser
Collection();

    Twitterizer.TwitterResponse<Twitterizer.TwitterUserCollection> response;

    response = Twitterizer.TwitterList.GetMembers(oAuthTokens, userName, listName);

    Console.WriteLine(response.RateLimiting.Remaining + " " + response.RateLimiting.ResetD
ate);

    if (response.Result != Twitterizer.RequestResult.Success)
    {
        throw new Exception();
    }

    usersInTheList = response.ResponseObject;

    if (usersInTheList != null && usersInTheList.Count == 20)
    {
        response = usersInTheList.NextPage();

        if (response.Result != Twitterizer.RequestResult.Success)
        {
            throw new Exception();
        }
        var tempUsers = response.ResponseObject;

        while (tempUsers != null)
        {
            foreach (var member in tempUsers)
            {
                usersInTheList.Add(member);
            }
        }
    }
}

```

```

    }
    var distinctCount = usersInTheList.Select(x => x.ScreenName).Distinct().Count(
);

    if (tempUsers.Count == 20 && usersInTheList.Count() == distinctCount)
    {
        response = tempUsers.NextPage();

        if (response.Result != Twitterizer.RequestResult.Success)
        {
            throw new Exception();
        }
        tempUsers = response.ResponseBody;
    }
    else
    {
        tempUsers = null;
    }
    }
}

return usersInTheList;
}

static void stream_OnStatusReceived(Twitterizer.TwitterStatus status)
{
    throw new NotImplementedException();
}

static void stream_OnStatusDeleted(Twitterizer.TwitterStatus status)
{
    throw new NotImplementedException();
}

static void stream_OnFriendsReceived(List<decimal> friendList)
{
    throw new NotImplementedException();
}
}
}

```