MBI

The Composition and Structure of the Online Community around Corporate Twitter accounts

A study on the online communities evolvement of the top 25 product software companies in Europe

Utrecht University

Farhad Andalibi
Utrecht University
17-07-2014

ABSTRACT

With a user base of over 517 million registered users, Twitter is the second most used popular social networking site, behind Facebook. A growing body of research has examined the online communities, which are built on the Twitter platform. This research is a follow-up study to a previous study, in which the network structure and composition of the corporate Twitter accounts of the twenty-five largest product software vendors in Europe were analysed. The objective of study is to analyse the development of the Twitter follower networks and see if change patterns can be identified in these developments. After the collection of the network data of all companies, that had a Twitter account, social network analysis was utilized to analyse the network structure. Analysis on the network structure has shown that the networks exhibit stronger small-world characteristics than the networks from the previous study. In addition, we also analysed the composition of the networks. The results on the composition has shown that a) the internal audience of the networks has increased, b) only two companies are actively monitoring their competitors, and c) there is an increase in the number of followers that follow multiple companies. Based on these findings, we performed several correlation tests to identify patterns in the development of the networks and found that a) the number of unfollowers is related to the number of tweets posted by the companies, b) the number of new followers is related to the amount of retweets and hashtags used, and c) the increase of the network size has strengthened the small-world characteristics of the networks.

PREFACE

ACKNOWLEDGEMENTS

First, I would like to express my deepest respect and sincere gratitude to my supervisors Prof. Remko Helms and Jan Martijn van der Werf, and Amir Saedi. I would like to thank my professor and supervisor, Remko, for providing me the opportunity to conduct this research under his guidance, for his continuous support, patience, motivation, enthusiasm, and immense knowledge. Without his guidance and persistent help, especially for the data collection phase, this thesis would not have been possible. I would like to thank Amir for his continuous feedback on the thesis and motivational support and enthusiasm, whenever I gave him a visit. Amir was not only a great supervisor, but also a great friend. Due to obligations, Amir was unable to continue as my second supervisor in the final stage of the project and from that moment, Jan Martijn operated as my second supervisor. I would like to thank Jan Martijn for his willingness to become my second supervisor at such a short notice and his feedback on my research.

Collecting the data for this thesis brought us some difficulties and in the end in was a lengthy process. However, in the end all necessary data was complete! Many thanks again to my supervisor, Remko, for using his PC's, both at home and work, to collect the data using NodeXL. Same goes for my former fellow student, Karl Werder, who helped us in this process as well. Finally, I would also like to thank Mehdi el Fadil, an entrepreneur, who utilized his product, Tribalytics, to collect the data of the largest Twitter accounts.

Moreover, I would like to thank several people for their motivational support. During my Master's program, I met a lot of great people and made a lot of new friends. But I would like to thank, in particular, these great friends who were always there by offering help or support: Nicole, Yiouli, Angeliki, Elena, Ivan and Shaheen.

Finally, I would like to thank my girlfriend, Samra. Without doubt, she gave me the greatest motivational support to finalize this thesis. In addition, she also provided help by reviewing my thesis and giving suggestions. Furthermore, I would like to thank my brother and my parents. My parents are responsible for the fact that I am studying here, in this country. Without them, I would not have been here and I appreciate everything what they have done for me. Without their love and support, I would not have been able to finalize my study.

NOTICE OF ORIGINALITY

I declare this thesis is my own work and that information derived from published or unpublished work of others has been acknowledged in the text and has been explicitly referred to in the list of references. All citations in the text are between quotation marks (" ") .

Farhad Andalibi

TABLE OF CONTENTS

1. INTRODUCTION

Since the development of the World Wide Web (WWW), a large number of social networking sites (SNSes) have been introduced. The first major SNS, SixDegrees.com, was launched in 1997 (Ellison, 2007). It was named after the six degrees of separation theory which states that any human knows any other by six acquaintances or relatives (Guare, 1992). SixDegrees.com didn't last more than four years (Donath, & Boyd, 2004), but shortly after they went offline SNSes boomed. Nowadays, SNSes have become an important element of our lives and many of us cannot live without it. They enable us to connect by creating personal profiles, inviting our friends and colleagues to have access to our profiles, and sending e-mails and instant messages between each other (Kaplan, Haenlein, 2010). It's the way we communicate with our family and friends but it is also a means to empower our professional life (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011).

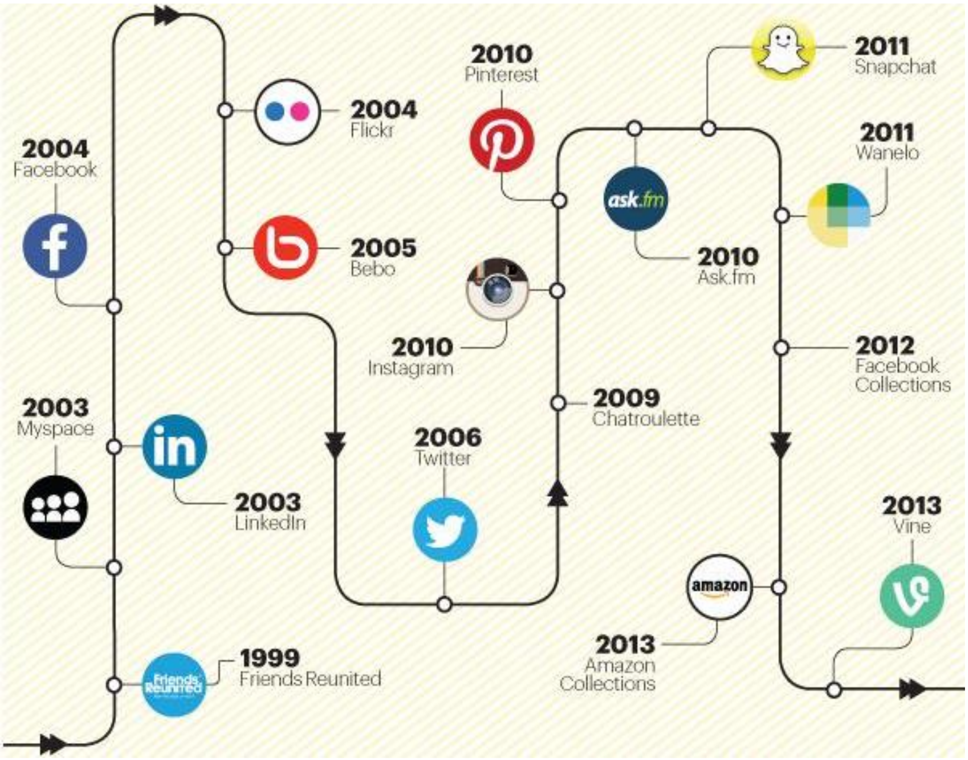

Figure 1: timeline of social network sites (Stocker, 2013)

The rise of the SNSes attracted the attention of companies who saw possible opportunities that could add value to their businesses. One of the SNSes that came under the attention of those companies is Twitter.

Launched in March 2006, Twitter is a micro-blogging service where users can post short messages up to 140 characters in length, also known as tweets, to a network of associates, also known as followers (Jansen, Zhang, Sobel, & Chowdury, 2009). The significant growth of Twitter, having more than 517 million registered users in the world (Lunden, 2012) made companies aware of the added values it could have for their organization.

Research on the use of Twitter by companies has shown that Twitter can have a great positive impact on the popularity of a company and even increase its revenue (Mulvaney, 2012). However, research also shows that Twitter is not only used for promotional activities but also to support the customer support and product development line of companies (Culnan et al., 2010; McWilliam, 2012). If used in the right manner, Twitter can add significant value to companies. Individual user actions could be aggregated to a specific pattern from which valuable information can be derived for the companies. Using the user actions, which are related to how Twitter is used by the users, companies could see a pattern that can be used for their marketing strategy.

In order to be successful and effective in that part, companies need to know to manage their online community on Twitter and engage with their followers (Ang, 2011). However, this cannot be done without having a good knowledge about the community itself and the target audience.

1.1 PROBLEM STATEMENT

With a user base of over 517 million registered users, Twitter is the second most used popular SNS, behind Facebook. A growing body of research has accompanied the rise of Twitter as researchers assess the usage and behavior of users on Twitter (Java, Song, Finin, & Tseng, 2007; Kwak, Lee, Park, & Moon, 2010; Cha, Haddadi, Benevenuto, & Gummadi, 2010).However, while the amount of research on the usage of Twitter has been increasing, previous studies primarily focus on the study of the whole Twittersphere (Java et al., 2007; Kwak et al., 2010), the Twitter follower network of a specific user in the Twittersphere (Gruzd, Wellman, & Takhteyev, 2011) or the online community in a particular event or situation (Cheong& Cheong, 2011). What is missing in the literature, is the study on the composition of the online communities and the development of the Twitter follower networks over time.

The lack of research on the composition of online communities on Twitter motivated Remko Helms and Karl Werder (2013) to study the composition and structure of the online community around corporate Twitter accounts. Together with Remko Helms, one of the authors of this study, Karl Werder carried out a social network and statistical analysis to analyse the network structure and composition of the Twitter follower network of the top twenty-five major software vendors in Europe (Helms, & Werder, 2013). The results of their study showed that the networks have strong small-world characteristics. However, networks are not a static phenomenon by nature as they constantly change in terms of composition and structure. This, and the lack of research on the development of Twitter follower networks, shows the need to do a follow-up study and analyse how the network structure and composition has evolved in the last year.

1.2 RESEARCH OBJECTIVE

As explained in the previous section, there is a need to analyse how the network structure and composition evolves over time. Therefore, the main objective of this study is to identify the change patterns in the development of the structure and composition of the Twitter follower networks. In the next sections, we explain the needed steps to reach our research objective in detail.

1.3 RESEARCH QUESTIONS

This sub-section will give an overview of the research question and the sub questions that have been formulated to support the research question. In order to formulate the main research question, we summarize our research objective into three smaller objectives. First, we want to gain a better understanding about Twitter follower network structure. Second, we want to see the composition of the Twitter follower network; i.e. who are the followers? And finally, we want to know whether we can identify change pattern in the way these communities evolve over time. This leads us to the following research question:

> *"What is the structure and composition of online Twitter communities and how does it evolve?"*

This research question is supported by the following sub research questions:

SQ1: What are online communities?

In order to have a suitable preparation for our research, we first need to expand our knowledge about the research domain and social network analysis (SNA). At the end of this sub-question, we expect to have enough knowledge regarding online communities and SNA.

SQ2: What particular structure do the networks of the top 25 product software companies in Europe have? (e.g. small network)

In the second sub-question of this study, we look at the network structure of the Twitter follower networks. Results of the previous study showed that the networks have small-world characteristics and a high number of reciprocal relations. Does this still hold on or has it changed?

SQ3: Who are the followers of these product software companies?

In the third sub-question of this study, we expect to get a good overview of the composition of the Twitter follower networks. Our main goal for this sub-question is to identify the target audience and their characteristics. In addition to that, we would like to know whether companies have identified their target audience themselves, by analysing their tweeting strategy.

SQ4: To what extend does the composition and structure evolve over time?

Finally, in our final sub-question of this study, we hope to reach our research objective. This question concerns the identification of the change patterns in the development of the Twitter follower network structure and composition. In order to identify these patterns, we will use the content (tweets) of the corporate Twitter accounts and identify the impact it had on the changes in the network structure and composition.

1.4 RELEVANCE

This section describes how relevant the problems investigated in this research are and the research objective. Two perspectives are employed, a scientific perspective which describes the benefits to the academic world, and a societal perspective which describes the benefits of this research to the society.

A. SCIENTIFIC CONTRIBUTION

From an academic perspective, this research contributes in several ways. A look into related studies from the literature, shows us several previously conducted researches that used a SNA and/or statistical analysis research approach. Some notable examples are a SNA study on how students interact with each other on the Twitter platform (Ullrich, Borau, & Stepanyan, 2010), a statistical analysis to measure the user influence on several Twitter topics (Cha et al., 2010) and a SNA study on tweets during the Australian 2010-2011 floods (Cheong, & Cheong, 2011). However, as explained earlier, these researches primarily focus on the study of the user behavior on the whole Twittersphere or the online communities on specific topics or time periods. Little is known about the composition of the Twitter follower networks and how these networks evolve over a period of time. For that reason, we will be trying to cover that in this study to provide researchers a better insight into the composition and evolution of Twitter follower networks.

B. SOCIETAL CONTRIBUTION

As explained earlier, it is important for companies to identify their target audience. In order to do that, they'll have to gain insight into the composition of their Twitter follower network. Using our method, companies will be able to identify the characteristics of their target audience. This enables them to adapt their Twitter strategy and strengthen their Twitter strategy and/or presence.

1.5 GLOSSARY

- Micro-blogging service
  A micro-blogging service is an Internet-based application which belongs to the social network family and belongs allows users to exchange short sentences, individual images, or links (Kaplan, & Haelnlein, 2011).
- Tweets
  Tweets are defined as short messages which are posted on the micro-blogging service (Jansen et al., 2009).
- Online Twitter community / Twitter follower network
  Online Twitter community or Twitter follower network is the community which is formed through the follower network of a Twitter account (Helms, & Werder, 2013).
- Small-world network
  A small-world network is defined as a complex with a high degree of local clustering and where the length of the shortest path between any pair of actors (followers) tends to be small (Humphries, & Gurney, 2008).
- SQL
  Structured query language (SQL) is a scripting language which is used to interact with relation databases (Date, & Darwen, 1987).
- CSV
  A comma-separated values (CSV) file is a simple and widely supported text file, that represents tabular data by using commas to separate the values in each line of data (Shafranovich, 2005).

2. THEORETICAL BACKGROUND

In order to give an overview of what is already known in the literature regarding online communities and social network analysis, the theoretical background is provided.

As the aim of this study is to research the structure and composition of twitter communities and their development in a year's time, it is important to provide a framework regarding the knowledge on online communities and social networks. As a result, the theoretical background is divided into three parts. The first part will give an overview of what is known in the literature regarding online communities. The second part will emphasize on literature regarding Twitter, and the final and third part, will focus on literature regarding the SNA research method.

2.1 ONLINE COMMUNITIES

Since the invention of the WWW by Tim Bernes-Lee in 1989 (Bernes-Lee, 1992), all kind of different online communities have been arising as well. An online community could potentially be viewed as an extension of a traditional community. A traditional community can be defined as a group of people with common interest or practices who would meet and interact (Driskell, & Lyon, 2002), whereas an online community is based on the same principle, except that the communication is done in a virtual environment (Preece, 2000; Ellis, Oldridge, & Vasconcelos, 2004).

In the nineties, three types of online communities that were used frequently and interchangeably were:

- **E-mail lists**
  An e-mail list is based on the concept of e-mail, introduced in 1978 by Shiva Ayyadurai (Ayyadurai, 2013). The rise in the penetration rate of the Internet and adoption of e-mail, resulted in the development of announcement and discussion lists, the two forms of e-mail lists (Weibel, 1995). The difference between the two is that in an announcement list you usually have one author and multiple recipients known as subscribers, whereas in a discussion list all subscribers have the possibility to reply to messages. These messages would then be sent to all other subscribers which eventually would end up in a group dialog (Sloan, 2006).
- **Chat rooms**
  The story of online chat rooms dates back to 1972, when Doug Brown and David R. Woolley developed Talkomatic, a virtual environment that enabled up to five people to exchange messages simultaneously (Woolley, 1972). Chat rooms were very popular and widely used in the nineties and early 2000s (Taubman, 2002). The key feature of chat rooms and at the same time difference with e-mail lists and discussion boards, was the ability to exchange messages real-time (Meola, & Stormont, 2000).
- **Discussion boards**
  Discussion boards, also referred to as “forums” or “bulletin boards”, can be seen as an extension and web-based version of discussion lists. The main difference between the two is that conversations in discussion lists can get very chaotic, whereas conversations on discussion are stored in a structured way on a web-server. As the popularity of discussion boards began to increase, programmers with an entrepreneurial mindset started to develop

their own forum software. Nowadays, many kind of paid and open source forum software can be found on the Internet (Breslin, Harth, Bojars, & Deckers, 2005). Despite the large amount of different forum software in different programming languages, they all usually rely on one common structure which is shown in Figure 2.

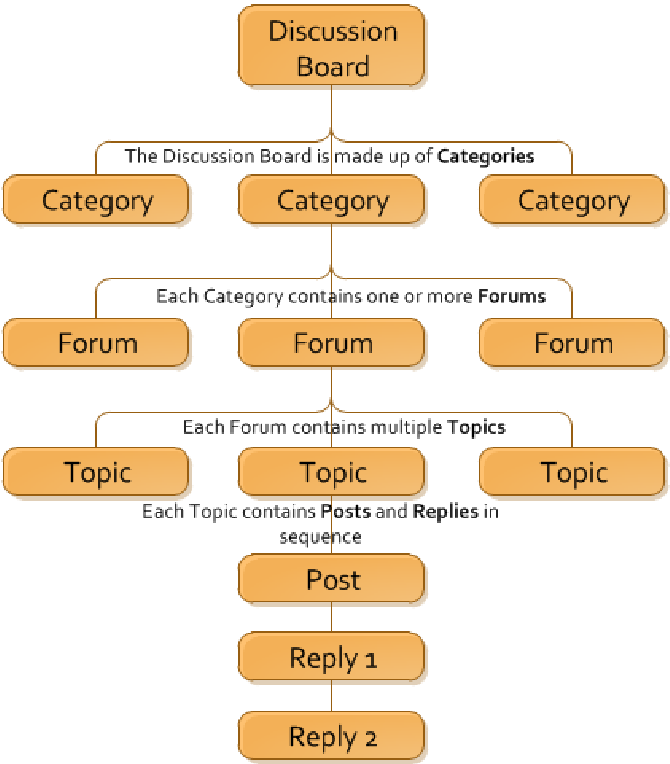

Figure 2: Structure of a discussion board

The continuous development of the Internet resulted in another new type of online community in the late nineties: social networking sites. In the early stages of the SNSes, the key difference between SNSes and previous online communities was the fact that SNSes were mainly organized around individuals, and not interests. However, the continuous development of SNSes changed this pattern and current SNSes are now not only organized around individuals but also interests. Without doubt, SNSes are currently the most prominent type of online community on the Internet. Before continuing to emphasize on SNSes, let's first take a look at its definition:

> *"We define social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site." (Ellison, 2007)*

The term social network site and social networking sites are often used interchangeably as Boyd and Ellison (2007**)** state:

> *"While we use the term "social network site" to describe this phenomenon, the term "social networking site" also appears in public discourse, and the two terms are often used interchangeably. We chose not to employ the term "networking" for two reasons: emphasis and scope. "Networking" emphasizes relationship initiation, often between strangers. While networking is possible on these sites, it is not the primary practice on many of them, nor is it what differentiates them from other forms of computer-mediated communication (CMC)."*

As explained by Boyd and Ellison (2007), the difference between the two terms is the emphasis of the term "networking" on relationship initiation on SNSes. Due to the fact that this study emphasizes on the study of follower relations in a Twitter follower network, we have decided to only use the term social networking sites (SNSes) in this study.

The emergence of the SNSes started with the introduction of SixDegrees.com (Kasvana, Nusair, & Teodosic, 2010). Named after the theory of the six degrees of separation, it quickly became a popular SNS. At its height, the website had over 3.5 million registered members (Ahmad, 2011). However, the enormous rise of new SNSes in the beginning of the 21th century, eventually led to the closure of SixDegrees.com (Donath, & Boyd, 2004). In the following period, a high degree of innovation was seen in the development of the new SNSes. While previous SNSes were focused on providing a service which enabled people to connect and share information with each other, a lot of new SNSes emerged with different terms in scope and functionality. Facebook, the most frequently used SNS at this moment[1], was initially only available for students of universities in the United States (Govani, & Pashley, 2005). In 2003, a year before Facebook was introduced, LinkedIn was founded. LinkedIn, became one of the first SNSes which was developed around professional individuals, providing them a service to maintain and strengthen their professional network (Ellison, 2007). The continuously development of new types of SNSes continued and eventually led to the development of Twitter.

As explained before, SNSes have currently become an important element in our lives. Since the last decade, more and more people have been using it and nowadays, we cannot live without it anymore. The enormous rise of SNSes, has also attracted the attention of companies who have discovered the possibilities SNSes could have for their business. The following part of the theoretical background, elaborates on the research on Twitter. In one of the sub-sections, we will also look at the literature regarding the usage of Twitter for business purposes.

2.2 TWITTER

Launched in 2006, Twitter is the most notable micro-blogging service on the internet. Prior to the rise of Twitter, the term micro-blogging service was not used that frequently. The term originates from the concept of tumblelogs (Siles, 2012). Tumblelogs, was introduced by Jason Kottke in 2005 in a proposal where he described a new way of blogging (Kottke, 2005).

[1] http://www.alexa.com/topsites

Over the last years, Twitter has received a lot of attention from researchers in various disciplines. Research on Twitter focuses on different aspects. Before an overview of these different aspects is provided, we will first discuss the term community on Twitter.

Online Twitter community

In this study, we consider the online Twitter community or Twitter follower network, as the community around a corporate Twitter account. However, it is important to realize that this is not the only perspective a community on Twitter can have. Another option is to consider the online community as the whole Twittersphere and study the relations among the users on the whole Twitter network (Kwak et al., 2010). In addition to these two different perspectives, there are also different interpretations of the relations on Twitter. While most researchers, such as ours, focus on the following relations among the users, there are researches where only reciprocal relations are considered as relations (Gruzd et al., 2011). This goes even one step further in a particular research where a relation is only considered when the users exchange messages with each other (Huberman, Romero, & Wu, 2008).

TWITTER ADOPTION

Early research on Twitter was mainly focused on the usage of Twitter by its users (Java et al., 2007; Hughes, & Palen, 2009). Later on, researchers also focused on the motivation to use Twitter (Hargittai, & Litt, 2011). Researchers found that users use Twitter in different kind of ways. In the early years of Twitter, users primarily used Twitter to share their opinion regarding their daily life or purchasing decisions (Kim, & Srivastava, 2007; Zhou, Zhang, & Zimmerman, 2011). However, with the enormous growth of Twitter, the role of Twitter was also evident in the health care where doctors used the platform to interact with their patients (Hawn, 2009). In addition, the growth of Twitter, also attracted researchers to focus on how Twitter can contribute and add value to businesses and companies (Culnan, McHugh, & Zubillaga, 2010). Companies, also saw opportunities to use Twitter as an interaction method with their (potential) customers (Kafai, Fields, & Burke, 2010). Even though, other SNSes, such as, Facebook were used as well, Twitter had a huge advantage over other SNSes, due to the ease and simplicity of attracting new followers on Twitter. This enabled companies to use innovative ways to increase their brand awareness, drive engagement with their followers, or even attract new employees (Fosso, Wamba, & Carter, 2013). However, over the years, companies have not only gained positive effects from Twitter. Researchers has shown that in some cases, offensive Tweets by companies or employees, has had an enormous negative impact on the businesses of these companies (Ojeda-Zapata, 2008; Hutchings, 2012; Kierkegaard, 2010). For this reason, companies started to create Twitter strategies for their internet marketers (Burton, & Soboleva, 2011; Wilson, Guinan, Parise, & Weinberg, 2011).

CONTENT AND AUTHORS OF MESSAGES

Most research on the content and authors on Twitter is performed to measure the influence of the users on Twitter. The most important factor to measure a user influence on Twitter is the retweet rate of the tweets posted by a particular user (Cha et al., 2010; Suh, Hong, Pirolli, & Chi, 2010). Retweets are defined as tweets which are shared (retweeted) by another Twitter user which is different than the original poster of the tweet (Jansen et al., 2010). Analysis on the retweet ratio has shown that on average retweeted content (tweets) reach approximately 1000 users, no matter the importance or popularity of the original poster (Kwak et al., 2010). Using the retweets mechanism and the already existing PageRank algorithm[2], researchers, such as Lee et al. (2010), introduced new algorithms to measure the influence of users on Twitter. In a follow-up study on the research of Lee et al. (2010), the algorithm was extended, and the new algorithm was found to be more accurate for identifying the influence of Twitter users (Weng, Lim, & Jiang, 2010). In addition, to the research on the influence of a user, we found that in some cases, Twitter is also used to track people in particular events and analyse the crowdedness of a particular area (Lee, Wakamiya, &Sumiya, 2011; Hosseini, Unankard, Zhou, & Sadiq, 2014). In the next sub-section, we will also point out how the content on Twitter is used to give recommendations or make predictions.

PREDICTIONS

Big data has become very important and crucial in the last few years. Many companies are nowadays focusing on how to analyse data and turn it into useful information which could add value to their businesses. In this research, we will be comparing two data sets which could provide us information regarding the factors which contribute to the success of a Twitter follower network. The popularity of big data has also been visible on Twitter. Recently, a number of services have been launched to make recommendations regarding news, TV shows or products (Abel, Gao, Houben, & Tao, 2013; Ariyasu, Fujisawa, & Sunasaki). The recommendations are based on the content of a particular user on Twitter. When emphasizing on the predictions area, we can also see cases where the content of Twitter has been used to predict elections (Tumasjan, Sprenger, Sandner, & Welpe, 2010). In addition to that, we have also seen several researchers studying how the content on Twitter can make predictions regarding the stock market (Bollen, Mao, & Zeng, 2011; Zhang, Fuehres, & Gloor, 2011). This even led to the creation of multiple start-ups which developed tools to predict the stock market by analysing the tweets on Twitter (Hassan, Abbasi, & Zeng, 2013).

SOCIAL NETWORKS

As our study is based on a social network analysis on Twitter, we also provide an overview of several previous studies on Twitter where social network analysis was applied. One of the first researches on Twitter, Java et al. (2007), demonstrated how Twitter was used. In order to analyse the usage of Twitter, Java et. al (2007) collected all tweets between the first of April and 30th of May in 2007. Their analysis showed that the overall Twitter network had a high degree correlation and reciprocity which indicated the relational closeness among the users. In addition to that, they also categorized the type of users on Twitter by analysing their edges on Twitter and understanding the reasons of why they use Twitter. The following categories were identified:

- Information Source
  Information sources were identified as users who have a large number of followers. In some cases these could also be automated bots or spam accounts.

[2]The PageRank algorithm is an algorithm developed by Google in measure of websites on their search engine and rank them accordingly (Rogers, 2002).

- Friends
  Most of the edges were found to be mutual friends following each other such as family and co-workers.
- Information seeker
  Finally, the final category of people which was identified were the information seekers. They were seen as users who don't update their Twitter status frequently but instead follow a lot other users.

While the research of Java et al. (2007) showed a high reciprocity among the users on Twitter, this was not the case in one of the biggest and most famous research studies on Twitter (Kwak, Lee, Park, & Moon, 2010). Their findings showed a very low reciprocity rate, especially when compared to other SNSes such as Facebook and Flickr. A possible explanation for this difference could be that the research of Java et al. (2007) was conducted just a year after Twitter was introduced and thus the possibility of users knowing and following each other was most probably higher at the start of Twitter.

While there are many examples of studies available where social network analysis is applied to study the network structure on Twitter, there is no research available on the development of Twitter follower networks over a specific period of time. The closest research we found on this phenomenon was the study of Cao, Wan, Lu, Quan, and Chen (2010) who studied the development of the network structure on Twitter during a natural disaster.

2.3 SOCIAL NETWORK ANALYSIS

DEFINTION

The main objective of this study is to explore the online Twitter community and look at its structure and relations between the followers. In order to do that, we use social network analysis (SNA) in our study. SNA can be defined as a study on the "social relationships in terms of network consisting of nodes and ties" (Jin, Girvan, & Newman, 2001). Nodes, also referred to as vertices are described as the individuals within the networks (Wasserman, 1994). In our study on the online Twitter community, these are the followers. Ties, also referred to as links, connections or edges define the relationships between the individuals (Wasserman, 1994).

As explained earlier, SNA can be applied to various disciplines and in different situations. One researcher who contributed a lot in reviewing the development of SNA is Linton C. Freeman (2011). In one of his works, he characterized SNA as an approach which consists of the following four properties (Freeman, 2011):

1. Recognition of the importance of relations among the actors (nodes)
2. Based on the collection and analysis of network data which consists of actors and ties (edges) between those actors
3. Usage of graphs as a way to visualize the ties between the actors and reveal possible patterns
4. Usage of mathematical and computational models to analyse and describe those ties and patterns

DIRECTED VERSUS UNDIRECTED GRAPHS

As explained in the previous sub-section, graphs can be used in SNA to visualize the ties between the actors in the network. What is important to know is that there are two types of graphs: directed and undirected graphs (Hanneman, & Riddle, 2005). The difference between these two types is the way the ties in the network are interpreted. In a directed network or graph, the ties also reveal the direction of the connection while in an undirected network or graph, the directions are unknown (Hanneman, & Riddle, 2005). An illustration of both graphs is shown in Figure 3.

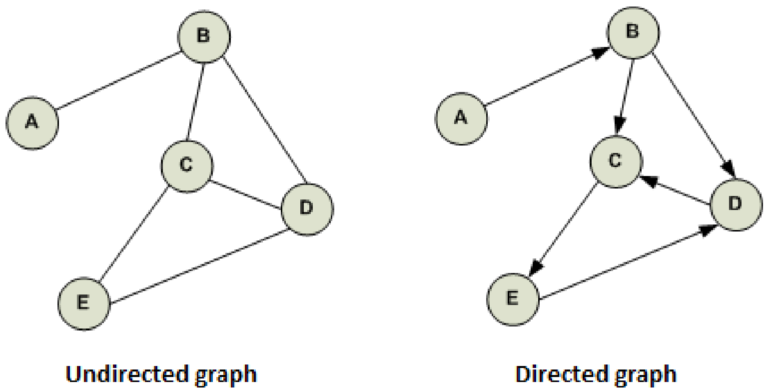

Figure 3: Undirected and directed graph

SNA TOOLS

In the final section of the literature concerning social network analysis, we decided to look into several tools which could be used for the analysis and visualization of the Twitter follower networks.

NODEXL

NodeXL is an open-source template extension created for Microsoft's Excel. NodeXL enables the end-user to import raw network data from different SNSes (Smith et al., 2009). Furthermore, it includes several features to analyse and explore the network after the data has been successfully imported. From recent studies, we have seen that NodeXL is slowly gaining more popularity among SNA approaches on Twitter (Kim, Lee, Choi, Bae, Ko, & Kim, 2013; Yep, & Shulman, 2014). Other than NodeXL, not many open-source tools facilitate the import of networks from SNS.

GEPHI

Gephi is a powerful modular software application which enables the end-user to explore and manipulate networks (Latha, & Sathiyakumari, 2012). As it is the case of modular software, Gephi enables users to extend the application with plugins which are created by third parties. This makes it possible to import data from different data sources such as CSV and MySQL (Cherven, 2013). From the literature, we have found that Gephi is very popular for visual analysis of network graphs (Cherven, 2013; Heymann, & Le Grand, 2013).

UCINET

UCINET is a freeware network analysis tool which can be used for free for 90 days (Borgatti, Everett, & Freeman, 2002). The first version of UCINET which included numerical analysis was developed in 1987. It is one of the oldest and most notable network analysis software tools currently available. The application was widely used in the SNA book of Hanneman and Riddle (2005) and it can be considered as a powerful and comprehensive tool in the network analysis field. Contrary to Gephi, UCINET does not accept multiple types of file formats as the tool can only read matrix data from a text file saved in Excel format.

R

The final application which we are discussing is R, a free software package and programming language for statistical data analysis (Team, 2005). Contrary to the previously discussed software applications, R is very simplistic in terms of the user interface as the applications runs through its own command-line tool. Compared with the other tools, this makes R a bit harder to learn. However, previous SNA studies have shown that R can be a very powerful tool when handling large data sets (Bader, Chaudhuri, Rothberg, & Chant, 2004; Thijs, Lemmens, &Fieuws, 2008). Similar to Gephi, R can also be extended with plugins. We found several SNA plugins which not only provide structural and visual analysis but also include features to import network data from different data sources (Butts, 2007; Butts, 2008; Csardi, & Nepusz, 2006).

3 RESEARCH METHOD

3.1 RESEARCH STRATEGY

This research provides, in an explorative and quantitative way, an overview of the Twitter follower network structure and composition of the of the top twenty-five major software vendors in Europe according to the Truffle100 list. The ranking is based on the software revenue of the companies from the year of 2010 (Appendix B). Table 1 gives an overview of these sub questions and the related research methods.

	Question	Method
SQ1	What are online communities?	Literature research
SQ2	What particular structure do the networks of the top 25 product software companies in Europe have? (e.g. small network)	Structural analysis
SQ3	Who are the followers of these product software companies?	Statistical analysis and text mining
SQ4	To what extend does the network overlap and structure change over time?	Statistical analysis

Table 1: Research questions and the related research methods

In order to gain a deeper insight in the research strategy, a Process-Deliverable Diagram (PDD) is created. A PDD is a diagram which is based on a meta-modeling technique invented by Sjaak Brinkkemper (1996). The diagram gives an overview of all activities (processes) and the corresponding concepts (deliverables) for a particular method.

An overview of the constructed PDD for the research method used in this study is shown in Figure 4. On the left side of the diagram, the activities of the method are shown. These activities are further described in Table 2, the activity table (Brinkkemper, 1996). On the right side of the diagram, the deliverables for each activity are shown. Further information about these deliverables can be found in Table 3, the concept table (Brinkkemper, 1996), where a clear definition has been provided for all deliverables.

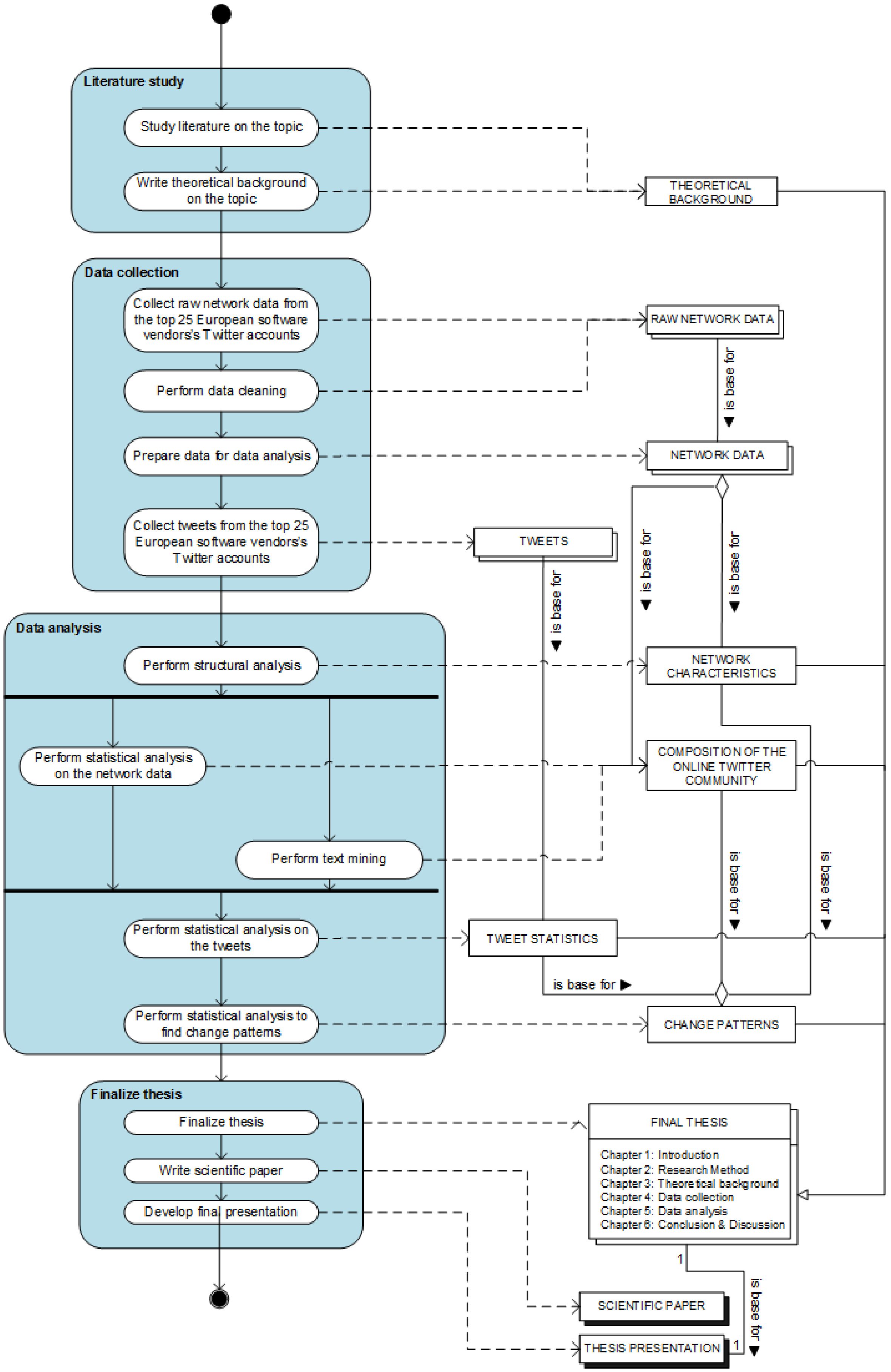

Figure 4: Process-deliverable diagram (PDD) of the research strategy

Activity	Sub-Activity	Description
Literature study	Study literature on the topic	In the first sub-activity of the literature study, the goal is to find, select and study papers. This activity will result in the initial version of the THEORETICAL BACKGROUND.
	Write theoretical background on the topic	The second and final sub-activity of the literature study, will lead to the finalization of the THEORETICAL BACKGROUND framework.
Data collection	Collect raw network data from the top 25 European software vendors' Twitter accounts	In the first sub-activity of the data collection, the Twitter follower network data will be collected and stores as RAW NETWORK DATA.
	Perform data cleaning	This sub-activity concerns the removal of duplicate and inaccurate records from the RAW NETWORK DATA.
	Prepare data for data analysis	In this sub-activity, the RAW NETWORK DATA will be converted into multiple formats. The end-result of the conversion is stored as NETWORK DATA and is ready for the data analysis phase.
	Collect tweets from the top 25 European software vendors' Twitter accounts	For each of the corporate Twitter accounts, the tweets will be collected and stored as TWEETS.
Data analysis	Perform structural analysis	By applying SNA metrics on the NETWORK DATA, we can now identify the NETWORK CHARACTERISTICS of the Twitter follower networks.
	Perform statistical analysis on the network data	Statistical analysis on the NETWORK DATA will provide us information regarding the network size and overlap which are part of the COMPOSITION OF THE ONLINE TWITTER COMMUNITY.
	Perform text mining	Text mining on the Twitter account names and user names will provide us information regarding the internal audience which is also part of the COMPOSITION OF THE ONLINE TWITTER COMMUNITY.
	Perform statistical analysis on the tweets	Statistical analysis on the TWEETS will result in the TWEET STATISTICS which provides information regarding the usage of the Twitter accounts by the companies.
	Perform statistical analysis to find change patterns	Statistical analysis on the relation between the concepts of the previous sub-activities on the data analysis phase will eventually result in the identification of the CHANGE PATTERNS.
Finalize thesis	Finalize thesis	The main deliverable of this study will be the FINAL THESIS which includes the results from all previous activities.
	Write scientific paper	In this activity, a SCIENTIFIC PAPER will be written for this study.
	Develop final presentation	Finally, the study will be finalized with an oral presentation which is based on a Powerpoint presentation stored as THESIS PRESENTATION.

Table 2: Activity table

Concept	Description
THEORETICAL BACKGROUND	The theoretical background in this study refers to the theoretical background chapter of the final thesis deliverable. The theoretical background provides the reader with an overview of the research topics in this study.
RAW NETWORK DATA	The raw network data refers to the output of the NodeXL application. It contains the data regarding the nodes (follower accounts) and edges (relations between the followers).
NETWORK DATA	The network data concerns the storage of the cleaned and structured raw network data.
TWEETS	Tweets are defined as short messages which are posted on the micro-blogging service (Jansen et al., 2009).
NETWORK CHARACTERISTICS	The network characteristics include metrics which provide information regarding the network structure.
COMPOSITION OF THE ONLINE TWITTER COMMUNITY	The composition of the online Twitter community provides information regarding the followers of the Twitter accounts.
TWEET STATISTICS	The Tweet statistics provide information regarding the usage of the Twitter accounts by the companies.
CHANGE PATTERNS	The change patterns are patterns where patterns related to the changes in the network structure and composition are identified.
FINAL THESIS	The final thesis is the most important deliverable of this study which contains a written information of all the research stages and the deliverables.
SCIENTIFIC PAPER	The scientific paper is a condensed version of this thesis which will be used for submission purposes.
THESIS PAPER PRESENTATION	The final paper presentation concerns the Powerpoint presentation which involves the results of this study.

Table 3: Concepts table

3.2 LITERATURE REVIEW

Prior to the actual data collection and data analysis phase, a literature review will be conducted. The literature review is an important part of a research project as it conceptualizes the research areas, and sets the basis and directions for the actual research phase (Webster & Watson, 2002). In order to develop the theoretical background for this study, the three-stage literature review process, proposed by Levy and Ellis (2006) will be used. Levy and Ellis (2006) suggest a three-step literature review process to guide researchers in the Information Systems research domain. The actual steps are displayed in Figure 5, where 1) Input stands for the paper selection phase, 2) Processing for the analysis phase and 3) Output for the actual literature review.

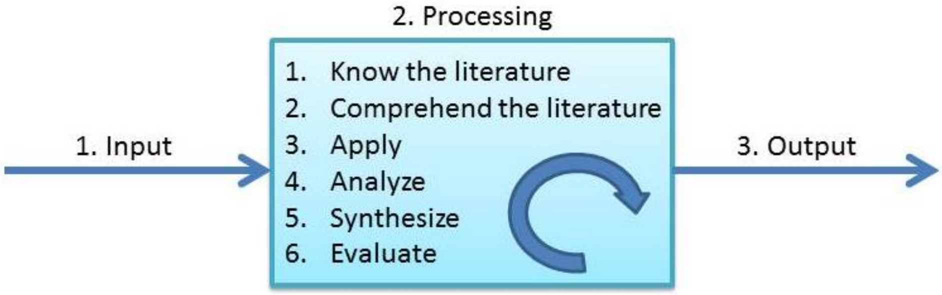

Figure 5: Three-stage literature review process (Levy, & Ellis, 2006)

3.3 DATA COLLECTION

The data collection phase will be conducted in the following stages:

1. **Twitter follower networks**
   In this stage, NodeXL 1.0.1.251[3] will be used to collect the Twitter follower network data from the corporate Twitter accounts.
2. **Tweets**
   In the second stage, a self-developed script will be used to collect the tweets of the corporate Twitter accounts.
3. **Data cleansing**
   In the third stage, the collected data will cleaned to make sure duplicate or inaccurate records are removed from the data set.
4. **Data preparation**
   In the final stage of the data collection phase, a self-developed script will be used to convert the collected data into multiple formats. The formats used in this study include: (1) SQL, used for transferring the collected data into a MySQL database, (2) GDF, in order to analyse the data in Gephi 0.8.2[4], and (3) CSV, in order to analyse the data in R 3.1.0[5].

[3] http://nodexl.codeplex.com/releases/view/117300
[4] https://www.gephi.org/users/download/
[5] http://cran.r-project.org/bin/windows/base/

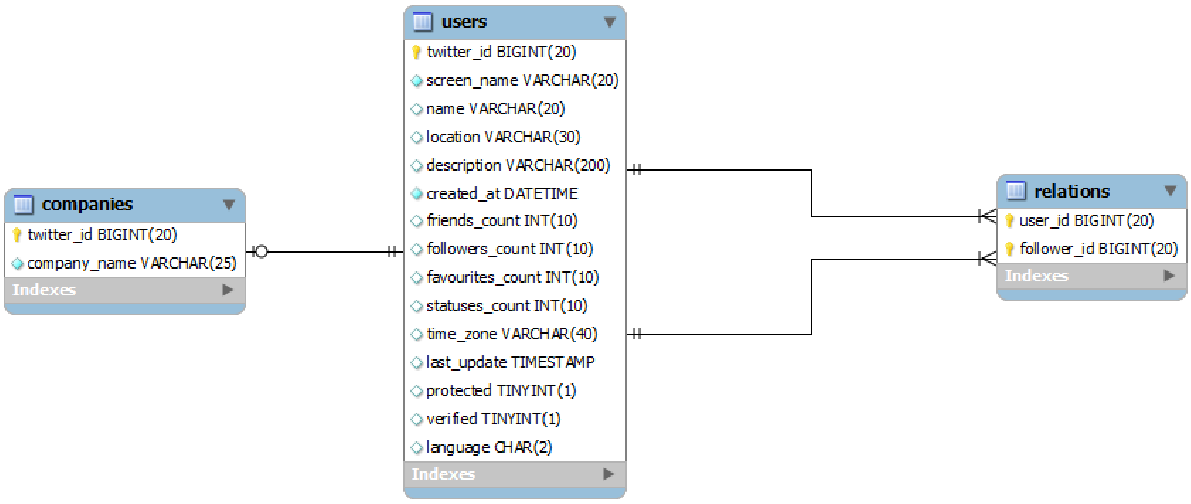

Figure 6: MySQL database diagram used for storing the follower network

3.4 STRUCTURAL ANALYSIS

As explained earlier, the software tools used in the data analysis phase are Gephi and R. The motivation to use Gephi in this study is its ease of use and our familiarity with Gephi. The motivation to use R is mainly its robustness and ability to handle large and complex data. Based on our previous experiences with Gephi, we found that Gephi is very unstable when analysing a large network data set.

In order to analyse the network, we will be using several metrics. As we are looking to compare the current network structure with the network structure from another time, we will be using the same metrics as used in the previous study (Helms, & Werder, 2013). The metrics, which are used in this study, are discussed below.

RECIPROCITY

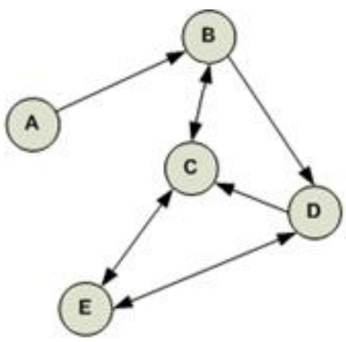

Figure 7: Network with reciprocal relations

Reciprocity refers to the ratio of reciprocal relations in a network (Hanneman, & Riddle, 2005). A concrete example to describe a reciprocal relation would be a situation where user A follows user B, and user B follows back user A. The reciprocity can be defined as the ratio of reciprocal relations x between the nodes n in a network, and total number of potential reciprocal relations in the network, which is calculated using the following formula, $\frac{n \cdot (n-1)}{2}$. The reciprocity in Figure 7 would for example be 0.30. However, in SNA, researches use two different approaches to calculate the reciprocity in a network. As both approaches are employed in this study, we will give an overview of both approaches. The first approach, referred to as the arc method, focuses on the ties in a network and calculates the number of involved ties which have a reciprocal relation divided by the total number of ties (Hanneman, & Riddle, 2005). The second approach, referred to as the dyad method, focuses on the actors in the network. In this method the calculation is based on the amount of reciprocal ties divided by the total of potential ties among the actors (Hanneman, & Riddle, 2005). Using the dyad and arc method, the reciprocity values in Figure 7 would be 0.50 ($\frac{3}{6}$) and 0.67 ($\frac{6}{9}$), respectfully.

AVERAGE DEGREE

The degree refers to the amount of ties a node (follower) has in the network and could help to reveal the most powerful individuals in a network (Hanneman et al., 2005). Furthermore, the degree of the actors can be used to reveal the degree distribution of the network. The result of the degree distribution can be used to identify a small-world structure in a network (Watts, & Strogatz, 1998). As we are focusing on the metric of average degree, this metric gives the average degree value of all actors in the network.

AVERAGE PATH LENGTH

The average path length is a metric which refers to the average of path length (number of steps) between any possible pair of actors in the network (Coward, & Jonard, 2004). The average path length can provide information regarding how fast the information can flow from one actor to another actor (Coward, & Jonard, 2004). The formula to calculate the average path length can be defined as following: $l = \frac{1}{n \cdot (n-1)} \cdot \sum_{i \neq j} d(v_i, v_{j)}$, where n refers to the network size, and d to the shortest distance between node a (v_i) and node b (v_j).

AVERAGE CLUSTERING COEFFICIENT

The clustering coefficient measures the degree to which the nodes in a network tend to cluster together (Watts, & Strogatz, 1998). As this is quite a complex metric to explain, we use a simple example to describe it. In most networks, the probability is high that if node A is connected with node B and node B is connected with node c_i, then the likelihood is high that node A is connected with node C as well. In SNA terms, a situation like this is called a 3-vertex clique (triangle). Based on that, the clustering coefficient c_i is defined as the ratio between the number of connections among the neighbors k_i of node i and the maximum of potential connections, m_i. The latter is calculated using the following formula: $m_i = \frac{k_i \cdot (k_i - 1)}{2}$. In our study, we are going to focus on the average clustering coefficient, c, which is the mean value of the c_i of all nodes.

SMALL-WORLD-NESS INDEX

The final metrics which we will be using is used to calculate the degree of the small-worldness of a network. The calculation of the small-worldness index is performed in three steps (Humphries, & Gurney, 2008). The first step is to calculate the ratio of the clustering coefficient in the network and the clustering coefficient of a Erdős–Rényi (ER) graph with the same size of nodes. ER is a random graph which uses the Bernoulli distribution with parameter p to generate the random edges (Namayandeh, Didehvar, & Shojaei, 2013). The second step is to calculate the ratio of the average path length and the average path length of the random graph which was initialized in the previous step. Finally, the ratio of the output of the first and second step will result into the small-world-ness index.

3.5 STATISTICAL ANALYSIS AND TEXT MININGON COMPOSITION

During the next stage of our data analysis phase, we will conduct statistical analysis and text mining to analyse the composition of the followers' network. In order to do so, several analysis will be performed:

1. **Internal versus external audience**
   The first analysis concerns the relation between the followers and the companies behind the corporate Twitter accounts. Our objective here is to see what proportion of the followers are actually somehow related to the companies (e.g. staff members, employees or consultants). In order to do so, we will be applying text mining on the screen names and bio (short biography) of the Twitter accounts. Users with screen names and short biographies that contain one or more keywords related to the name of the companies, are then identified as part of the internal audience.
2. **Unique followers**
   In the second analysis, we want to see whether there is an overlap between the followers of the corporate Twitter accounts. Due to the geographical closeness of the companies and the fact that they are operating in the same field (IT), we would assume that there are people who follow company A, but also company B. These users could be marketers, suppliers, competitors or potential customers. In order to determine the overlap, we will be using the Jaccard index (Bassecoulard & Zitt, 1999). The Jaccard index is a well-known static which is used to measure the similarity between two data sets. The motivation to use the Jaccard index is mainly based on the fact that the measure was used in the previous study.
3. **Social authority**
   The following analysis concerns the analysis of the influence of the Twitter accounts and its followers. In order to do so, we will be using a metric called social authority. The social authority metric provides a score on a 1 to 100 scale, which shows the influence of an individual on Twitter (Bray, & Peters, 2013). The calculation of the social authority provides us the ability to see whether the changes in the network size (left and joined users) had a positive or negative impact on the overall social authority score.
4. **Competitors**
   In the fourth analysis, we are interested to find the number of Twitter accounts related to each of the Truffle 100 companies that follows the top 25 product software companies. This analysis enables us to see whether the product software companies are following their competitors, and which of the companies has the most active presence on Twitter.
5. **Activity**
   Finally, we want to see whether companies are using a tweet strategy, based on the activity of their followers. In order to do so, we will first use a third-party web service to identify the most active moments of the followers of each of the corporate Twitter accounts. These

results will then be compared with the actual date and time of the tweets, which are posted by the corporate Twitter accounts. This enables us, to see whether the companies have taken account the activity of their followers.

3.6 STATISTICAL ANALYSIS ON COMPANY TWEETS

The next stage of the data analysis phase concerns the analysis on the tweet rate of the corporate Twitter accounts. For each of the corporate Twitter accounts, we will only analyse the tweets which are posted during a specific time span. The time span is based on the moment the network data from a corporate Twitter account was collected in the previous study and the moment the network data will be obtained in this study. The analysis on the tweets will provide us information regarding the number of tweets, retweets or replies a company has shared on Twitter. Furthermore, we will also identify the number of retweets companies have received on their own tweets, and the number of hashtags the companies have used in a tweet. This information helps us to identify possible change patterns in the development of the corporate Twitter accounts.

3.7 ANALYSING RESULTS

In the final stage of the data analysis phase, we will use the results from the earlier analysis to identify possible change patterns in the development of the corporate Twitter accounts. In order to do so, several correlation tests will be performed. Each of the tests, will be performed, based on existing theory and our own assumptions regarding factors which influences a specific change in the Twitter follower networks.

4 DATA COLLECTION

In this chapter, the data collection process will be elaborated. The first section, provides a quick overview of the size of the sample analysed in this study. The next section, gives an overview of the type of data that was necessary for this study. The third section, describes how the data from the previous study was used in this study. The fourth section, provides an overview of the services and technologies that were used to collect the data for this study. Finally, the fifth and final section, provides an explanation about the data conversion process.

4.1 CORPORATE TWITTER ACCOUNTS

This study concerns the analysis on the composition and structure of the online community around corporate Twitter accounts of the top 25 product software companies in Europe. However, it is important to note that four of the 25 companies have no Twitter account[6]. As a result, the number of actual corporate Twitter accounts are 21.

4.2 DATA TYPES

In the following sub-sections , we describe each type of the data that was collected from the corporate Twitter accounts.

TWITTER FOLLOWER NETWORKS

The first type of data that was collected, was the network data, also referred to as the Twitter follower networks in this study. The network data consisted of two types of data which were collected asynchronously. The tool, used to collect this data, first had to identify and collect the followers of the corporate Twitter accounts. After the collection of the followers, the next step was to identify the relations (edges) between these followers (degree 1). An example of how this looks like in a graph is shown in Figure 8.

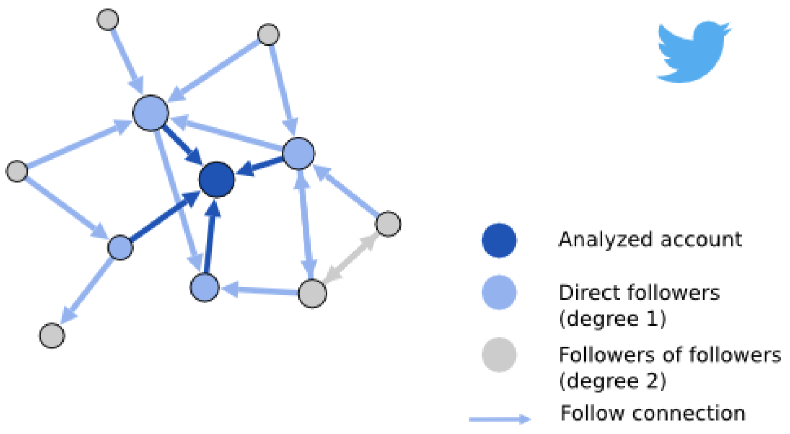

Figure 8: Visualization of a Twitter follower network ("Raw Twitter follow graph data –Tribalytics", 2014)

[6] The following companies are not present on Twitter: Wincor Nixdorf, Compugroup Holding, Murex and NIS.

TWEETS

The next step in the data collection process, was to collect the tweets of the corporate Twitter accounts. This step was not performed in the previous study. However, in this study, this data could become very useful and necessary as we believed it could possibly help to explain the changes in the network structure and composition of the Twitter follower networks.

4.3 DATA FROM PREVIOUS STUDY

As noted before, the data from the previous study only concerned the raw network data from all corporate Twitter accounts. This data was delivered to us for this study in .mdb (Microsoft Access) files. We decided to import this data into MySQL with each of the accounts having its own entity in the data model. The decision to go for this approach was taken as we believed that it would simplify the composition comparison in the next stages.

4.4 SERVICES AND TECHNOLOGIES

Due to changes in the terms and conditions on the use of Twitter's API (Sippey, 2012), we were not able to re-use the tool which was developed and used in the previous study by Helms and Werder (2013). This forced us to look into alternative ways to collect the network data.

NODEXL

For 19 out of the 21 corporate Twitter accounts, the network data was collected by using NodeXLand this process took nearly two months. As NodeXL could not handle very large Twitter accounts, we decided to look into alternative ways to collect the network data for SAP and SageUK.

TRIBALYTICS

As we were unable to collect the network data of SAP and SageUK by ourselves, we decided to contact other companies whom had experience with collecting raw network data from Twitter. Thankfully, we found a commercial party who was willing to co-operate with us for this study. Mehdi El Fadil, owner of Mango Information Systems, decided to help us by adjusting his product Tribalytics so it would just obtain raw network data. Mango Information Systems, located in Brussels, offers data analytics and social media monitoring on Twitter for companies[7]. Their product Tribalytics is a reporting tool which provides companies the possibility of getting insights about their industry's social presence on Twitter[8]. Future requests for raw data can be made through the raw data collection page on Tribalitycs' webpage[9].

[7] http://www.mango-is.com
[8] http://www.tribalytics.com
[9] http://www.tribalytics.com/tribalytics-raw

TWITTER REST API V1.1

As noted before, due to Twitter's changes on the limited usage of their API for network data collection, we were unable to use the web application from the previous study. However, what we did notice was that the limit on the collection of tweets for collecting tweets from Twitter accounts was still sufficient to obtain this particular data in a short timeframe[10]. As a result, a customized web application was built in PHP to obtain the tweets from all corporate Twitter accounts. Twitter, however, does have a limit on the maximum amount of tweets you are allowed to collect from each account you specify[11]. Because of this, we were only able to collect up to 3200 tweets for accounts which had more than 3200 tweets posted.

4.5 DATA CONVERSION

Contrary to the network data, the collected tweets were directly parsed into a MySQL database which was set-up for this study. Figure 9provides an overview of the database diagram where the converted network data and tweets were eventually stored. Further explanation will be given in the following about the data format of the stored data.

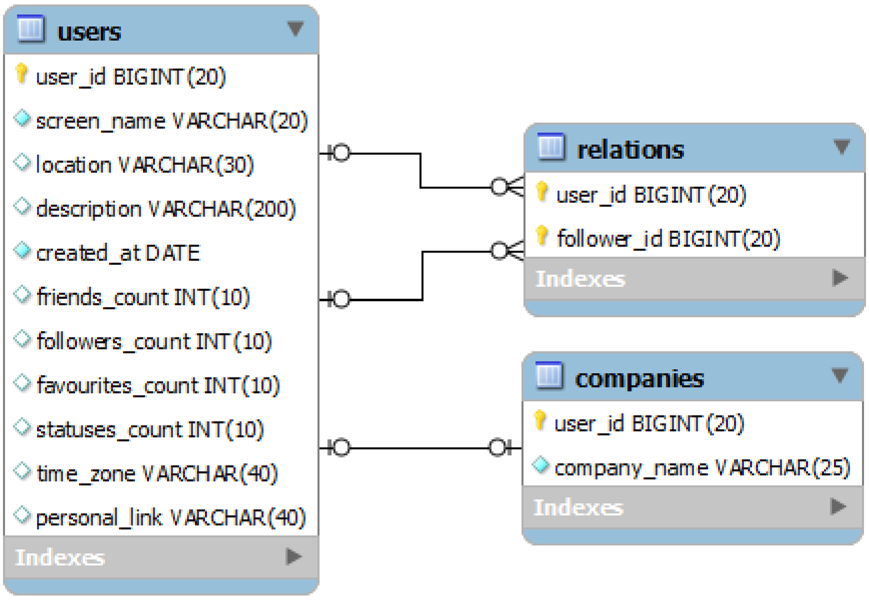

Figure 9: MySQL database model of the Twitter follower networks

SQL

The first step was to collect and store all the network data in MySQL. The decision to store all the network data into MySQL is driven by our desire to develop PHP scripts in further stages which would help analysing the data.

[10] https://dev.twitter.com/docs/rate-limiting/1.1/limits
[11] https://dev.twitter.com/docs/api/1/get/statuses/user_timeline

GDF

Based on the decision to use Gephi for the graph conversion and analysis on the network structure, we also developed a PHP script which converted all the network data into GDF files.

CSV

Finally, in later stages we noticed that Gephi was unable to process large network data sets. For this reason, we decided to use R for analysing the network structure as well. The plugins we used in R, required the input file to be of CSV format. Therefore, we converted the network data to CSV as well by using Microsoft Excel.

5 DATA ANALYSIS

In this chapter we explain how the collected data has been analysed. In the first section, we give an overview of the analysis on the tweets of the corporate Twitter accounts. In the second section, we look at the size of the network and the changes that have occurred. In the third section, the network structure is explored and analysed by applying SNA metrics. The fourth section concerns the analysis of the composition of the online Twitter community. Apart from the first section, all other sections follow the following steps:

1. Overview of the analysis of the data from the previous study
2. Overview of the analysis of the data for this study
3. Comparison of the results
4. Identification of possible patterns in the occurring changes

In order to identify the possible causes for the changes in the Twitter follower network, we carry out correlation tests by utilizing Pearson's product moment correlation (Pearson's correlation) in SPSS 20.0.0.2[12]. Pearson's correlation is a common method which can be utilized to measure how closely related two variables are. The highest possible correlation coefficient (r) is 1, which indicates a perfect correlation between the two properties. The statistical significance (P) of a Pearson's r is determined using the critical values of Pearson's r. In this study, these are typically defined as $p<0.05$, and $p<0.01$.

IMPORTANT NOTES

SAMPLE SIZE

As explained before, four of the top 25 product software companies have no Twitter account. In addition to that, we found that another four companies currently have a different Twitter account then the one they had when the previous study was conducted[13]. For that reason, we have opted to use two sample sizes in this study. For the analysis of the network- size, and structure, tweets, and first two analysis on the network composition, we used the sample size of 17. As for the other analysis on the network composition, we used a sample size of 21 as these analyses were not performed in the previous study and thus no comparison could be made.

Our decision to exclude four of the Twitter accounts which were changed since the previous study, was based on the fact that their inclusion will not provide us an accurate insight in the development of their Twitter network. This decision was justified by a pre-analysis on the network size which showed that 57% of the followers of the old Twitter account were not present in the new Twitter follower network. This decrease in followers was significantly higher than the mean decrease of 18%.

[12] http://www14.software.ibm.com/download/data/web/en_US/trialprograms/W110742E06714B29.html
[13] The following companies have moved to a new Twitter account: HP Autonomy, Invensys, Mysis and Swisslog.

INCORRECT VALUES IN NETWORK STRUCTURE DATA FROM PREVIOUS STUDY

In order to be able to compare the results of this study with the previous study, we first had to validate the old results. During the validation process, we wanted to see whether our calculation of the network structure metrics would produce the same result as presented in the previous study. In order to calculate the metrics, we analysed the old network data in three different SNA tools: Gephi, R and UCINET. Our validation process showed that the reciprocity, average path length and small-world-ness index from the previous study contain incorrect values. The incorrect values for the average path length and small-world-ness were found to be caused by the fact that the relations (edges) in the Twitter follower network were interpreted as undirected. As explained before, undirected and directed networks use different formulas to compute the average path length. In addition, as the small-world-ness index is dependent on the average path length score, this score was incorrect as well. Finally, the incorrect values in the reciprocity score turned out to be caused by a possible bug in one of the Gephi modules (*Mutual_Degree_Range*). All in all, upon the findings of the validation process, we decided to recalculate the network structure metrics of the old network data to be able to have a valid comparison.

5.1 TWEETS

In order to analyse the tweets of all corporate accounts, a PHP script was developed which collected all the tweets up to 3200 tweets per Twitter account. The attributes collected are(1) the type of tweets (e.g. normal tweet, retweet or reply), (2) the amount of entities used in a tweet (e.g. hashtags, mentions and links) and,(3) the amount of received retweets. For the latter, a distinction is made between the amount of all retweets which includes the amount of retweets on tweets retweeted by the companies, and the amount of retweets on regular tweets. In this study, we use the terms normal and exclusive tweets interchangeably. Both terms refer to the tweets originating from the companies.

The results of the analysis on the tweets of all Twitter accounts are shown in Table 1 of Appendix A. Below in Table 4, a general summary of the findings is shown.

	Mean	Median
Total tweets	1455	1401
Total number of exclusive tweets	1112	1219
Total number of retweets	196	86
Total number of replies	147	61
Average number of hashtags per tweet	0.97	0.96
Average number of mentions per tweet	0.70	0.61
Average number of links per tweet	0.77	0.73
Average number of received retweets per tweet	2.13	1.66
Average number of received retweets per exclusive tweet	1.68	0.96

Table 4: General summary of the results from the analysis on the Tweets

The analysis on the tweets also provides an insight into the interaction rate of the companies with their followers. In this study, we define the interaction rate as the degree of replies[14] of the companies. Our findings show that only four of the 17 Twitter accounts have an interaction rate of 10% or higher. In addition, the following two accounts show an interaction rate of 0%: @cegidpresse[15] and @swiftcommunity[16].

5.2 NETWORK SIZE

During the first part of the analysis phase, the network size of the Twitter accounts were calculated. A summary of the changes of all Twitter accounts are shown in Table 5. As depicted in Table 5, the standard deviation (SD) is found to be very high. This can be explained as the number of followers of SAP (54536) is significantly higher than the second biggest and smallest Twitter accounts, 12146 and 141, respectively. A detailed version of the results and changes for each of the Twitter accounts provided can be found in Table 2, Table 3,and Table 4 of Appendix B.

	Old data	New data	Change
Followers - Mean	5613	11417	+103%
Followers - SD	13073	24723	+89%
Relations - Mean	85944	163655	+90%
Relations - SD	154989	295008	+90%

Table 5: General summary of changes in network size

The analysis on the network size has shown that the followers size of the Twitter accounts has increased substantially (103%) in one year's time. This increase was found to be even higher (123%) without the network size data of SAP which had the highest increase in terms of numbers. The percentual increase in the network size of these seventeen Twitter accounts was compared with the general percentual increase in Twitter's active user base. As Twitter's active user base has approximately increased by 75% during the same period between these two studies (Yarow, 2013) , we can say that the percentual increase of new followers for these seventeen accounts has been above average.

Using the results from the analysis performed above, we were able to see the numerical and percentual difference of the network size. However, what was missing is the amount of users that unfollowed the Twitter accounts and the amount of users that started following the Twitter accounts since the previous study. In order to calculate these values, we developed a PHP script which compared the data from the previous study with the current network data. A general summary of these changes is shown in Table 6.

[14] A reply is a tweet where a Twitter users replies to another user by mentioning his/her screen name.
[15] @cegidpresse is the Twitter account name which is used by CegidPresse
[16] @swiftcommunity is the Twitter account name which is used by SWIFT

	Amount	% difference in network size
Total number of unfollowers - Mean	1150	-18%
Total number of unfollowers, still existing - Mean	630	-9%
Total number of new followers - Mean	6954	+234%

Table 6: General summary of changes in followers size

The results in Table 6 show that on average 18% of the total followers of all Twitter accounts have unfollowed the companies since the previous study. Further analysis showed that most of these users do not have an account on Twitter anymore. However, even then there is still a percentual decrease of 9% followers. Even though the average percentual increase of new followers was much higher than the average decrease, we were still interested to see why some users had unfollowed the Twitter accounts.

DECREASE IN FOLLOWERS SIZE

In this part of the data analysis, the causes that lead to the decrease in follower size were investigated. Based on our assumptions and theory, we performed several correlation tests. The first correlation test concerned the association between the amount of posted tweets and the percentual decrease of users who have unfollowed the Twitter accounts. Previous study on this matter has shown that the decision to unfollow a Twitter account is in most cases related to the fact that users get annoyed by the amount of tweets they get to read from that specific Twitter account (Kwak, Chun, & Moon, 2011). To support this theory, we examined the association between the amount of posted tweets and the percentual decrease of users who have unfollowed the Twitter accounts.

A Pearson's correlation was run to determine the relation between the 17 Twitter account's PUSEU and ANTPD values (Table 15

		PUSEU
ANTPD	Pearson correlation (r)	0.462
	Statistical significance (p)	0.062
	Sample size (N)	17
Legend		
PUSEU	<u>% unfollowed and still existing users</u> Refers to the percentual amount of existing users that have unfollowed a Twitter account.	
ANTPD	<u>Average number of tweets per day</u> Refers to the average number of tweets (including retweets and replies) a Twitter account has shared per day.	

of Appendix I). This resulted in a nonsignificant correlation between PUSEU and ANTPD (r=0.462, N=17, p=n.s). However, we did identify an outlier in the scatterplot which we created (Figure 2 of Appendix I). The outlier in this test was caused by the data of @assecoesp[17], that had a 16% decrease of followers, which was substantially higher than the mean of the other accounts: 8%. Based on this

[17] @assecoesp is the Twitter account name which is used by Asseco

finding, we decided to exclude the data of @assecoesp and reanalyse the data. Another Pearson's correlation was run to determine the relationship between the 16 Twitter account's PUSEU and ANTPD values (Table 7). This time, it resulted in a strong, positive correlation between PUSEU and ANTPD (r=0.695, N=16, p<0.01).

		PUSEU
ANTPD	Pearson correlation (r)	0.695**
	Statistical significance (p)	0.006
	Sample size (N)	16
**. Correlation is significant at the 0.01 level (2-tailed)		
Legend		
PUSEU	<u>% unfollowed and still existing users</u> Refers to the percentual amount of existing users that have unfollowed a Twitter account.	
ANTPD	<u>Average number of tweets per day</u> Refers to the average number of tweets (including retweets and replies) a Twitter account has shared per day.	

Table 7: Results of the Pearson's correlation test to determine the relation between the Twitter account's PUSEU and ANTPD values (without @assecoesp)

The strong association can also be seen in the scatterplot as shown in Figure 10.

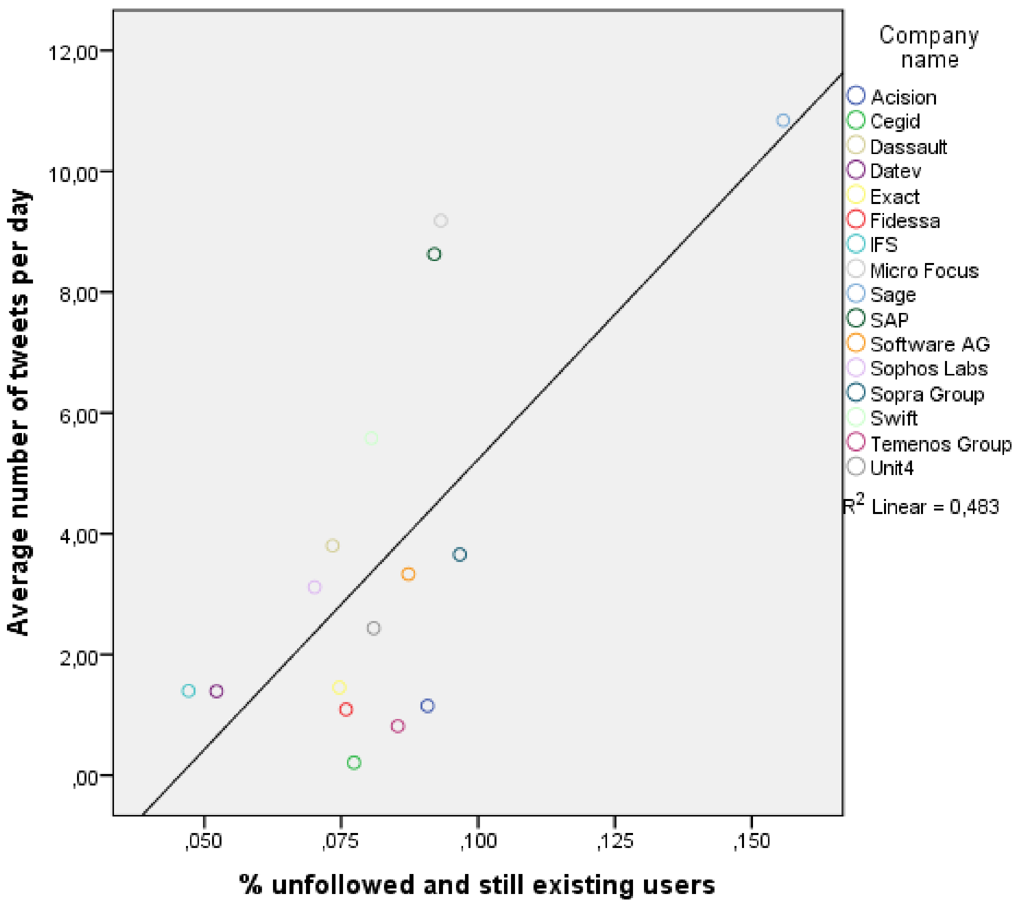

Figure 10: Scatterplot to show the strong, positive correlation PUSEU and ANTPD (without @assecoesp)

In addition to the previous test, we performed another test to analyse whether the percentual number of unfollowed users are associated with the different types of tweets (exclusive tweets, retweets, and replies) shared by the corporate Twitter accounts. A Pearson's correlation was run to determine the relation between the 16 Twitter account's PUSEU and ANEPD, ANRPD, and ANMPD values (Table 8). The test resulted in a strong, positive correlation between PUSEU and ANEPD ($r=0.574$, $N=16$, $p<0.05$), and PUSEU and ANRPD ($r=0.745$, $N=16$, $p<0.01$). However, there appears to be a nonsignificant correlation between PUSEU and ANRPD ($r=0.442$, $N=16$, p=n.s).

Based on the findings in this section, we are able to say that if we exclude the data of @assecoesp, we can conclude that the number of tweets indeed influences a Twitter user's decision to unfollow a Twitter account or not. As explained earlier, posting too many tweets in a short time might enable users to see the Twitter account as spam.

		ANEPD	ANRPD	ANMPD
PUSEU	Pearson correlation (r)	0.574*	0.442	0.745**
	Statistical significance (p)	0.020	0.086	0.001
	Sample size (N)	16	16	16

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Legend	
PUSEU	<u>% unfollowed and still existing users</u> Refers to the percentual amount of existing users that have unfollowed a Twitter account.
ANEPD	<u>Average number of exclusive tweets per day</u> Refers to the average number of exclusive tweets a Twitter account has shared per day. As explained earlier, exclusive tweets concerns the tweets that originate from the corporate Twitter accounts.
ANRPD	<u>Average number of retweets per day</u> Refers to the average number of times a corporate Twitter account has shared a tweet of someone else (retweeted) per day.
ANMPD	<u>Average number of replies per day</u> Refers to the average number of times a corporate Twitter account has replied to another Twitter account per day.

Table 8: Results of the Pearson's correlation test to determine the relation between the 16 Twitter account's PUSEU and ANEPD, ANRPD, and ANMPD values (without @assecoesp)

INCREASE IN FOLLOWERS SIZE

In the second part of the analysis on the network size, we wanted to gain deeper insight into the factors that influence the decision to follow a particular Twitter account. To achieve this goal, two Pearson's correlation tests were performed to analyse the association between the number of new followers and the use of Twitter by the corporate Twitter account's followers, and the use of Twitter by the corporate Twitter accounts.

Previous research on the use of Twitter has shown the importance of retweets on Twitter (Cha et al., 2010; Suh et al., 2010). As a result, we became interested to see whether retweets had any influence in the number of new followers in our study. Based on this, a Pearson's correlation was run to determine the relation between the 17 Twitter account's TNNF and TNRR, TNRRE, ANRT, and ANRET values (Table 9). As depicted from Table 9, the results in a strong, positive correlation between TNNF and TNRR (r=0.772, N=17, p<0.01), TNNF and TNRRE (r=0.833, N=17, p<0.01), TNNF and ANRT (r=0.650, N=17, p<0.01), and TNNF and ANRET (r=0.686, N=17, p<0.01).

		TNRR	TNRRE	ANRT	ANRET
TNNF	Pearson correlation (r)	0.772**	0.833**	0.650**	0.686**
	Statistical significance (p)	0.000	0.000	0.005	0.002
	Sample size (N)	17	17	17	17

**. Correlation is significant at the 0.01 level (2-tailed).

Legend	
TNNF	Total number of new followers Refers to the total number of new followers that have joined the network since the previous data collection.
TNRR	Total number of received retweets Refers to the total number of retweets a corporate Twitter account has received since the moment the data was collected in the previous study until it was collected in this study.
TNRRE	Total number of received retweets on exclusive tweets Refers to the total number of retweets a corporate Twitter account has received on exclusive tweets since the moment the data was collected in the previous study until it was collected in this study.
ANRT	Average number of retweets per tweet Refers to the average number of times a tweet of a corporate Twitter account has been retweeted.
ANRET	Average number of retweets per exclusive tweet Refers to the average number of times an exclusive tweet of a corporate Twitter account has been retweeted.

Table 9: Results of the Pearson's correlation test determine the relation between the 17 Twitter account's TNNF and TNRR, TNRRE, ANRT, and ANRET values

The strong correlation between TNNF and TNRRE is also illustrated in the scatterplot of Figure 11**.** Based on the results of the tests, we can conclude that the number of retweets is very important on Twitter. The results clearly indicate that Twitter accounts with a higher number of retweets tend to have a higher number of new followers.

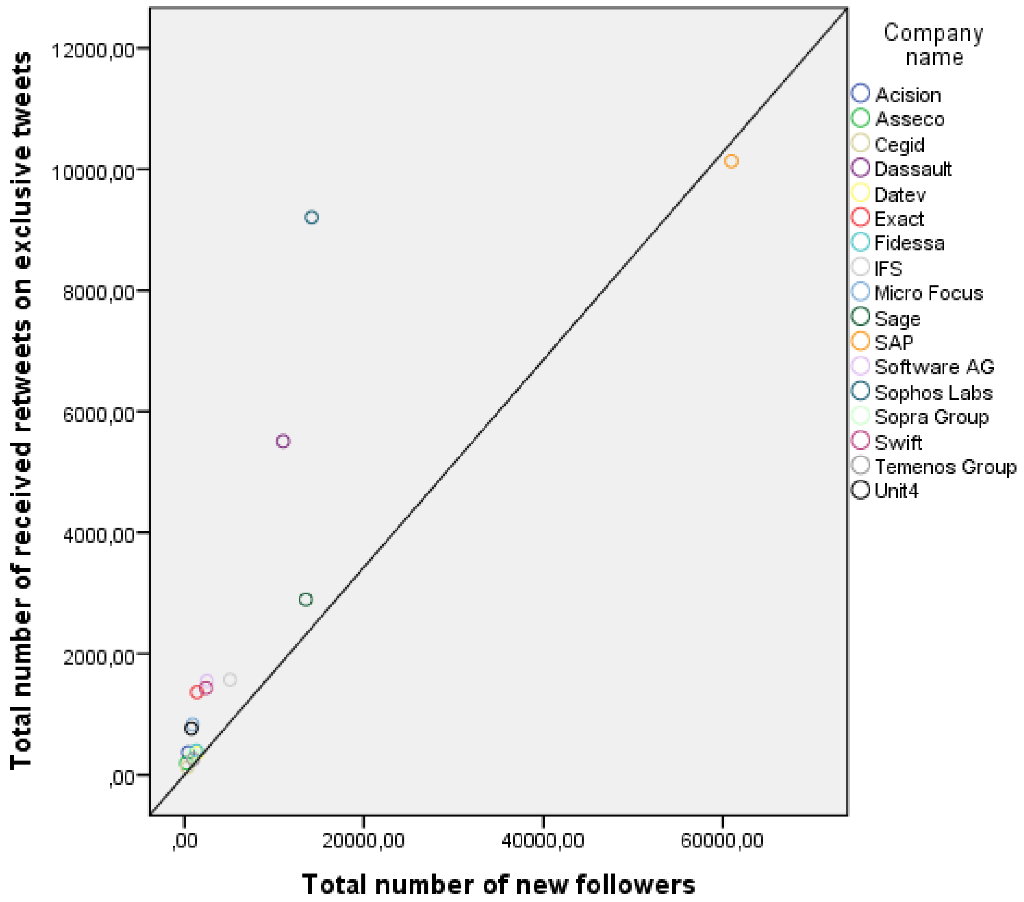

Figure 11: Scatterplot to show the strong, positive correlation between TNNF and TNRRE

As explained earlier, the second Pearson's correlation test concerned the analysis of the association between the growth of the network size and the use of Twitter by the corporate Twitter accounts. For this test we decided to see whether the usage of hashtags is associated with the growth of the network size. Previous research has indicated that tweets with more hashtags lead to more exposure in the search results of Twitter (Yardi, Romero, & Schoenebeck, 2009). In addition to that, tweets with hashtags seem to have a higher chance to get retweeted as well (Suh et al., 2010). Based on these researches, we assumed the usage of hashtags to have an impact on the network size of the Twitter accounts in our study. As a result, a Pearson's correlation was run to determine the relation between the 17 Twitter account's PFS and ANHT (Table 16 of Appendix I). Contrary to what we expected, the test resulted in a nonsignificant correlation between PFS and ANHT (r=0.462, N=17, p=n.s).

However, additional examination of the results resulted in the identification of an outlier. We noticed that the tweets which were shared by @softwareag[18], contained significantly more hashtags then the tweets shared by the other companies. On average, each tweet shared by @softwareag contained 13.96 hashtags whilst the second highest number was 2.43. Based on this finding, we decided to exclude the data of @softwareag and reanalyse the data. For a second time a Pearson's correlation was run to determine the relation between the 16 Twitter account's PFS and ANHT (Table 10). This

[18]@softwareag is the Twitter account name which is used by Software AG

time the test resulted in a moderate, positive correlation between PFS and ANHT ($r=0.520$, $N=16$, $p<0.05$).

		ANHT
PFS	Pearson correlation (r)	0.520*
	Statistical significance (p)	0.006
	Sample size (N)	16
*. Correlation is significant at the 0.05 level (2-tailed)		
Legend		
PFS	<u>% difference in follower size</u> Refers to the percentual difference in the follower size of a Twitter account in comparison with the follower size in the previous study.	
ANHT	<u>Average number of hashtags per tweet</u> Refers to the average number of hashtags a Twitter account has used per tweet.	

Table 10: Results of the Pearson's correlation test to determine the relation between the 16 Twitter account's PFS and ANHT values (without @softwareag)

The moderate correlation between PFS and ANHT is also illustrated in the scatterplot of Figure 12. Based on the test between the PFS and ANHT values, we were not able to draw any strong conclusion. A moderate, positive correlation was only found after the removal of the data of @softwareag. Still, we believe that hashtags are an important factor in the growth of a Twitter account as they lead to a higher exposure. However, we believe that the results of this study are too weak to strongly justify our assumption.

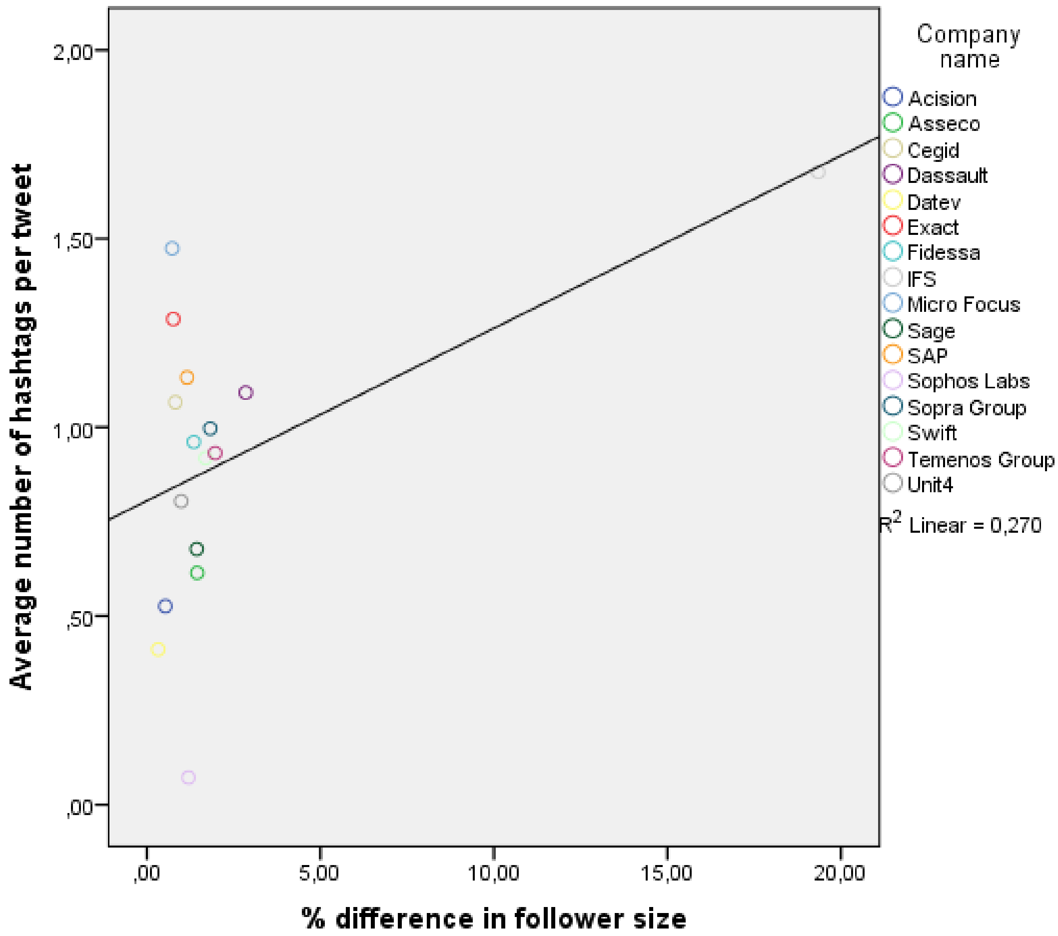

Figure 12: Scatterplot to show the moderate, positive correlation between PFS and ANHT (without @softwareag)

5.3 NETWORK STRUCTURE

In order to analyse the network structure, we employ the same metrics used in the previous study. These metrics are: reciprocity (arc and dyad), average degree, average path length, average clustering coefficient and small-world-ness. The network structure results of both studies together with the comparison of the two results can be found in Table 5, Table 6 and Table 7 of Appendix C, respectively.

The analysis of the network structure has been performed in both Gephi and R. As explained earlier, the network structure results from the previous consisted of incorrect values in some of the metrics. For this reason, and to be sure that the values in our analysis were correct, we decided to analyse the network structure in this study in two different SNA tools. To analyse the network structure with R, the following two packages were used: igraph[19] and SNA[20]. Several methods of these two packages were used simultaneously to calculate the SNA metrics. These methods are included in a customized R script which was also used for the analysis of the network structure (Appendix D).

---

[19] http://www.igraph.org/r/
[20] http://cran.r-project.org/web/packages/sna/index.html

SMALL-WORLD-NESS

In the previous study, the networks were found to have a high degree of small-world characteristics. We were interested to see whether this has changed or not. The results in this study were similar to the findings of the previous study. In addition, the networks of the companies seem to have an even higher degree of small-world characteristics as the small-world-ness index has increased by 22%. A general summary of the change in the small-world-ness index is shown in Table 11.

	Old data	New data	Comparison
Small-world-ness - Mean	13.47	16.46	+22.20%
Small-world-ness - SD	15.48	12.23	-20.96%

Table 11: General summary of the changes in the small-world-ness index

RECIPROCITY

The first metric we looked into was the reciprocity value of the networks. As explained in the theoretical background, there are two methods to calculate the reciprocity, i.e. the arc and dyad method (Hanneman, & Riddle , 2005). A general summary with the results of the reciprocity of all Twitter accounts is shown in Table 12. A recent study on the reciprocity level on Twitter has shown that the reciprocal relation on Twitter is relatively low (22.1%) compared with other SNSes such as Yahoo! 360 (84%) (Kwak et al., 2011). Even though the reciprocity of the Twitter follower networks has declined, it is still higher than the general reciprocal ratio (arc) on Twitter, evaluated to be 22% (Kwak et al., 2011).

	Old data	New data	Comparison
Reciprocity (arc) - Mean	0.50	0.46	-7.49%
Reciprocity (arc) - SD	0.15	0.15	+1.31%
Reciprocity (dyad) - Mean	0.35	0.31	-10.30%
Reciprocity (dyad) - SD	0.14	0.14	-4.03%

Table 12: General summary of the changes in the reciprocity rate

In the second part of the analysis on the reciprocity of the Twitter follower networks, we were interested to see whether there is a factor which influences the change in the reciprocity. Previous study on the reciprocity development on Twitter during a natural disaster has shown that the reciprocity declines as the network grows (Cao et al., 2012). A possible explanation for this phenomenon is that new nodes in a network are relatively unknown when they initially join a network. To support this statement, a Pearson's correlation was run to determine the relation between the 17 Twitter account's TNNF and PRA, and PRD (Table 17 of Appendix I). This resulted in a nonsignificant correlation between TNNF and PRA (r=0.287, N=17, p=n.s), and TNNF and PRD (r=0.228, N=17, p=n.s). However, based on the scatterplot (Figure 3 of Appendix I), we were able to identify an outlier which was caused by @sap. The outlier of @sap could be explained for the fact that the total number of new followers for @sap is five times higher than the second highest amount of new followers, respectively 60952 and 14171.

Based on the identification of the outlier of @sap, we decided to exclude the data of @sap and reanalyse the data. For a second time a Pearson's correlation was run to determine the relation between the 16 Twitter account's TNNF and PRA, and PRD (Table 13). This time the tests resulted in a moderately strong, negative correlation between TNNF and PRA ($r=0.606$, $N=16$, $p<0.05$), and TNNF and PRD ($r=0.576$, $N=16$, $p<0.05$).

		PRA	PRD	TNNF
PRA	Pearson correlation (r)	1	0.994**	0.606*
	Statistical significance (p)		0.000	0.013
	Sample size (N)	16	16	16
PRD	Pearson correlation (r)	0.994**	1	0.576*
	Statistical significance (p)	0.000		0.020
	Sample size (N)	16	16	16

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Legend	
PRA	<u>% difference in reciprocity (arc)</u> Refers to the percentual difference in the reciprocity (arc method) in comparison with the value from the previous study.
PRD	<u>% difference in reciprocity (dyad)</u> Refers to the percentual difference in the reciprocity (dyad method) in comparison with the value from the previous study.
TNNF	<u>Total number of new followers</u> Refers to the total number of new followers that have joined the network since the previous data collection.

Table 13: Results of the Pearson's correlation test to determine the relation between the 16 Twitter account's TNNF and PRA, and PRD values (without @sap)

Therefore, we can conclude that if we exclude the results of @sap, the reciprocity level indeed decreases with the growth of a network size as it was already the case in the work of Cao et al. (2012). The moderately strong correlation between the PRA and TNNF is also illustrated in the scatterplot of Figure 13.

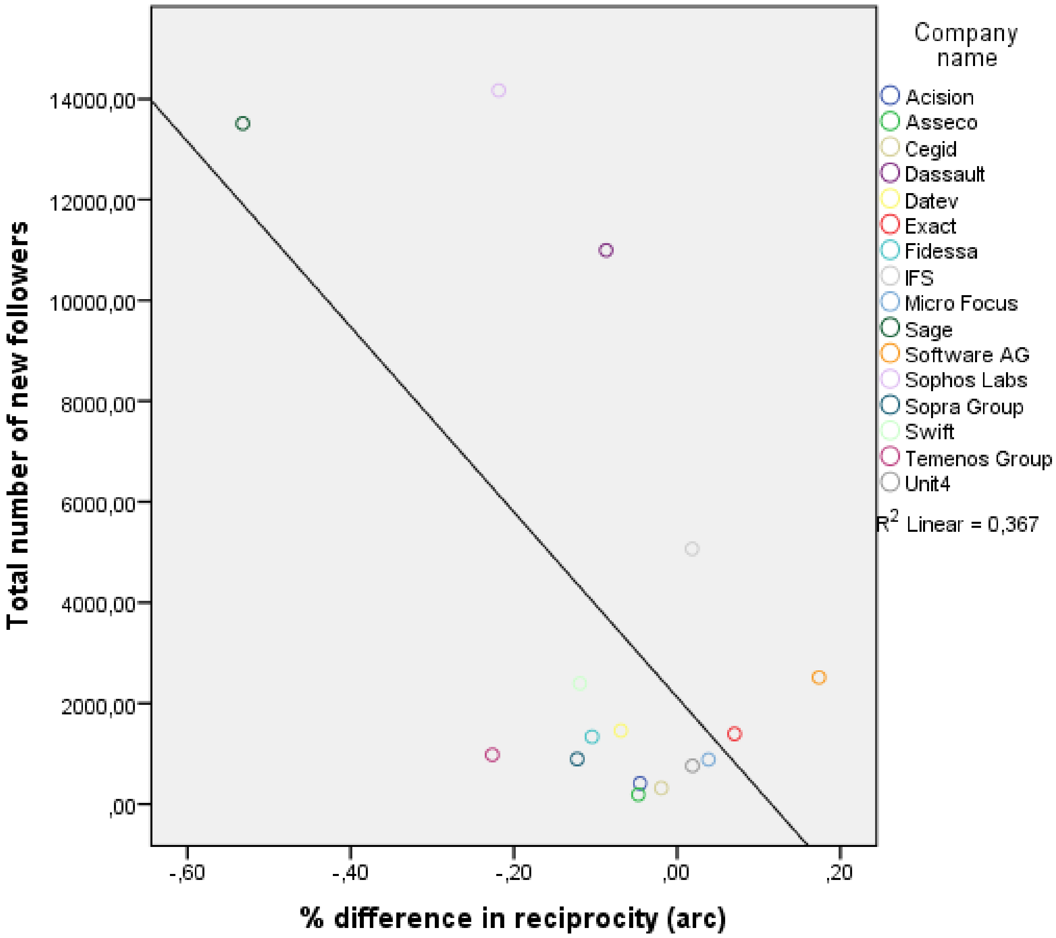

Figure 13: Scatterplot to show the moderately strong, negative correlation between PRA and TNNF (without @sap)

AVERAGE DEGREE

The second metric used to analyse the network structure is the average degree. As explained earlier, the degree of a node concerns the influence it has in a network. The results from the calculation on the average degree are shown in Table 14.

	Old data	New data	Comparison
Average degree - Mean	18.58	19.86	+6.86%
Average degree - SD	24.90	18.66	-25.06%

Table 14: General summary of the changes in the average degree

One would assume that the results of the average degree value would implicate that the nodes in the network have become more powerful. This is true, as the number of connected nodes to a particular node in the network has increased. However, it should not be forgotten that the number of total nodes has increased even more. Based on this, we decided to calculate the degree of the average degree in relation to the number of the network size. As depicted from Table 15, this number seems to have decreased. This indicates that the proportional degree of connectedness in a network decreases as the network size grows, which shows a characteristic of a real-life environment.

Finally, we decided to compare the mean of the average degree of all Twitter follower networks with the average degree of the European Twitter community, which was found to be 16.42 (Java et al., 2007). This indicates that the users in the Twitter follower networks of this study are densely interconnected. Furthermore, this also means an increase in the potential reach and exposure of tweets for the companies.

	Old data	New data	Comparison
Average degree in relation to network size - Mean	1.38%	0.89%	-25.64%
Average degree in relation to network size - SD	1.51%	1.08%	51.33%

Table 15: General summary of the changes in the proportional average degree

AVERAGE PATH LENGTH

The third metric we used to analyse the network data, is the average path length. As the average path length is an important metric to see how efficient the information flow in a network is (Cowan, & Jonard, 2004). The results from the calculation on the average path length are shown in Table 16.

	Old data	New data	Comparison
Average path length - Mean	3.18	3.27	+2.89%
Average path length - SD	0.48	0.63	+31.40%

Table 16: General summary of the changes in the average path length

While the average path length reveals how efficient the information flow in a network is, the increase in the average path length does not necessarily mean that the information flow has become less efficient. In fact, the opposite should be true, as the small-world-ness index has increased and thus the network has become more a small-world. Based on the small-network model as proposed by Newman (2003), the average path length grows as log(n), where n is the size of the network. This is also presented in Figure 14. To justify this and our assumption that the information flow has probably become more efficient, we calculated the log(n), as shown in Table 17. As depicted from Table 17, the increase of the log(n) is greater than the average path length. This means that the increase of the average path length is lower than expected, and thus the networks show indeed signs of becoming even more a small-world.

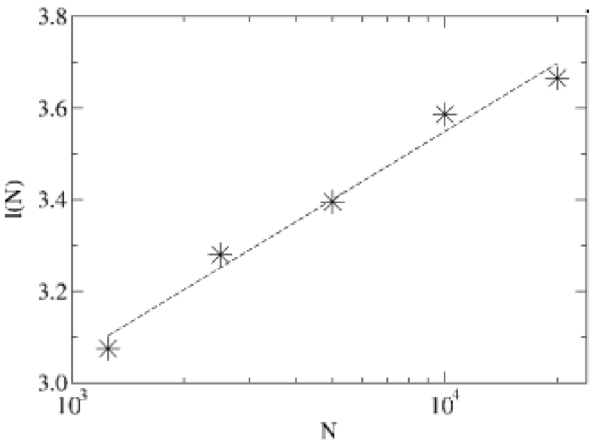

Figure 14: Scaling of the average path length with system size (Davidsen, Ebel, & Bornholdt, 2002)

	Old data	New data	Comparison
Log(n) - Mean	3.19	3.54	+10.97%
Log(n) - SD	0.66	0.64	-6.06%

Table 17: General summary of the changes in the log(n)

AVERAGE CLUSTERING COEFFICIENT

The metric that we used to analyse the network structure, is the average clustering coefficient. As explained before, the average clustering coefficient measures the degree to which nodes in a network tend to cluster together (Shavitt, & Weinsberg, 2009). The results from the calculation on the average clustering coefficient are shown in Table 18.

	Old data	New data	Comparison
Average clustering coefficient - Mean	0.45	0.46	+2.90%
Average clustering coefficient - SD	0.11	0.13	+19.03%

Table 18: General summary of the changes in the average clustering coefficient

5.4 NETWORK COMPOSITION

As explained earlier, it is important to identify your target audience in order to manage an online community effectively and successfully (Ang, 2011). For this reason, further data analysis was conducted to analyse the composition of the online communities of the Twitter accounts.

INTERNAL VERSUS EXTERNAL AUDIENCE

In the first part of the analysis of the composition of the online communities, we looked into the ratio of internal versus external audience. Internal audience in this context is seen as followers who are somehow affiliated with the company of the Twitter account they are following. In order to find the amount of internal audience, we applied text-mining on the account name and bio of the users. However, what is important to note is that the data from the previous study did not contain the bio of the accounts. For this reason, we only had the results of the internal audience based on their account name. As a result, in our calculations we made a distinctions between internal audience by @name (Twitter account name) and internal audience by @name and bio. The results of both studies together and the comparison of the two results can be found in Table 8 of Appendix E. A general summary of the results is shown in Table 19.

	Old data	New data	Comparison (nr)	Comparison (%)
Internal audience (@name) – Mean	79.53 (1.36%)	141.47 (1.83%)	+78%	+35%
Internal audience (@name & bio) – Mean	n/a	225.65 (2.13%)	n/a	n/a

Table 19: General summary of the changes in the internal vs external audience ratio

As shown in Table 19, we can see that the overall ratio of internal audience has increased by 35%. It is interesting to mention that the largest growth was seen in the network of @swiftcommunity where the internal audience based on the screen name, and internal audience based on the screen name and bio, had increased 457% and 850%, respectively. Furthermore, in the previous study, the networks of @ifsworld[21] and @unit4_group had the highest degree of internal audience (based on the screen name), with a degree of respectively, 5.72% and 3.66%. Due to the fact that @ifsworld had the largest growth in the network size (1694%), this number has decreased significantly to 0.69%. The degree of internal audience of @unit4_group[22] however has increased to 4.03% and they have now the highest degree of internal audience in their Twitter follower network.

Based on the growth of the internal audience, we performed a correlation test to see whether the growth could be associated with the change in the employee size. It is important to note that the number of employees were based on the employee size in the R&D (Research & Development) department of the companies as the list of Truffle[23] [24] does not provide the employee size from the other departments. A Pearson's correlation was run to determine the relation between the 17 Twitter account's PDIA and NDES, and PDES values (Table 20). As depicted from Table 20, the tests resulted in a nonsignificant correlation between PDIA and NDES (r=-0.063, N=17, p=n.s), and PDIA and PDES (r=-0.066, N=17, p=n.s).We believe that an association might be found if the degree of the internal audience is compared with data of the web care departments of the companies. Another possible explanation for the growth of the internal audience could be related to the fact that increasingly more companies tend to invest more in social media (Brotherton, 2012). As a result, it could be very possible that they also push their employees to have a presence on Twitter.

		NDES	**PDES**
PDIA	Pearson correlation (r)	-0.063	-0.066
	Statistical significance (p)	0.810	0.800
	Sample size (N)	17	17
Legend			
PDIA	<u>% difference in degree of internal audience</u> Refers to the percentual difference in the degree of internal audience of a Twitter account in comparison with the degree of internal audience in the previous study.		
NDES	<u>Numerical difference in employee size</u> Refers to the numerical difference in the number of R&D employees of the top 25 companies in comparison with the number in the previous study.		
PDES	<u>% difference in employee size</u> Refers to the percentual difference in the number of R&D employees of the top 25 companies in comparison with the number in the previous study		

Table 20: Results of the Pearson's correlation test to determine the relation between the 17 Twitter account's PDIA and NDES, and PDES values

[21] @ifsworld is the Twitter account name which is used by IFS AB
[22] @unit4_group is the Twitter account name which is used by Unit4
[23] http://www.truffle100.com/2011/ranking.php
[24] http://www.truffle100.com/2013/ranking.php

UNIQUE FOLLOWERS

Another interesting aspect of the composition of the online communities is to find the level of overlap between the followers of the Twitter accounts. Due to the geographical closeness of the companies and the fact that some of the companies might even be competitors, we would assume that in some cases followers from Twitter account A follow Twitter account B as well. These could possibly be marketers, suppliers, competitors or even potential customers. In order to determine the overlap, we will be using the Jaccard index, a measurement which is used to measure the similarity between two data sets (Bassecoulard&Zitt, 1999). The formula for the Jaccard index is defined as:

$$J\mu(A,B) = \frac{(A \cap B)}{(A \cup B)}$$

To explain the formula, we describe the calculation process as: (1) calculation of the total number of followers which are present in two data sets (A and B), which is denoted by ∩(intersection), (2) calculation of the number of distinct followers in the data sets, which denoted by ∪ (union), and (3) dividing the size of the intersection by the size of the union. The latter, is denoted by the coefficient of fraction, µ (mu).

Due to the large size of the matrix table of the Jaccard index values, only a small portion of the results is shown below. The results from the previous study are shown in Table 21, whilst the results from the current study are shown in Table 22. The full matrix table with all the results from the previous and current study and a comparison between the two are shown in Table 9, Table 10 and Table 11 of Appendix F.

	Misys	SoftwareAG	Swift	Temenos	Fidessa
Misys					
SoftwareAG	0.09%				
Swift	3.74%	0.33%			
Temenos	6.20%	0.33%	2.53%		
Fidessa	1.89%	0.21%	4.11%	1.39%	

Table 21: Jaccard index values for a selection of the Twitter accounts (old situation)

	Misys	SoftwareAG	Swift	Temenos	Fidessa
Misys					
SoftwareAG	0.65%				
Swift	3.13%	0.58%			
Temenos	4.37%	0.64%	3.15%		
Fidessa	3.02%	0.45%	4.47%	1.82%	

Table 22: Jaccard index values for a selection of the Twitter accounts (new situation)

A general summary of the values is presented in Table 23. On average, the network overlap had increased from 0.23% to 0.28%. We believe that the growth in the network overlap could be attributed to several factors which need to be further explored in future research. First of all, we believe the growth might be explained as Twitter might be reaching the maturity stage in its life cycle. While the user-base growth of Twitter is on the decline (Yarow, 2014), the current use of

Twitter is growing rapidly[25]. This could indicate that registered users are becoming more familiar with the platform, and are participating more than they used to do. Another possible explanation for the growth in the network overlap could be related to Twitter ads. Twitter ads is an advertising platform that was launched around the same time the data was collected in the previous study and provides companies the possibility to advertise on the Twitter platform. This enables companies to show advertisements to the followers of their competitors.

	Old data	New data	Comparison
Overlap - Mean	0.23%	0.28%	+22%
Combinations with no overlap	51 (0,24%)	1 (0,005%)	-98%

Table 23: General summary of the changes in the Jaccard index overlap

In addition to the network overlap, we also looked into the most frequent followers. In order to do that, a PHP script was developed to identify the Twitter accounts that were following six or more companies. The identified Twitter accounts were then categorized into four domains: news, company, individual and mixed. The results of this analysis, based on the data from the previous and current study, are presented in Table 13 and Table 14 of Appendix H. A general summary of the results is shown in Table 24. As depicted from Table 24, the number of frequent followers has doubled since the previous study. The most frequent follower in this study was found to be @twit_resrc, a Twitter account which was registered for research purposes in the previous study of Helms and Werder (2013). Furthermore, what was surprising was that only three companies have been added to the list of frequent followers. Due to the growing number of companies on Twitter (Coe, 2013), we expected this amount to be higher.

Domain	Twitter accounts (old data)	Twitter accounts (new data)
News	2	4
Company	5	8
Individual	3	9
Mixed	1	2
Total	11 (0.006%)	23 (0.013%)
Total (5 or more)	25 (0.026%)	83 (0.048%)

Table 24: General summary of the results of the most frequent followers

SOCIAL AUTHORITY

In order to calculate the influence of the followers of the Twitter accounts, we used the social authority metric, a score on a 1 to 100 scale, which shows the influence of an individual on Twitter (Bray, & Peters, 2013). The social authority metric was introduced in 2013 and is based on the following three components: (1) the retweet rate of the last few hundred regular tweets, (2), how recent those retweets are, and (3) other attributes for each Twitter account (e.g. follower count and friend count) that are optimized via a regression model trained to retweet rate.

[25] http://annenberg.usc.edu/News%20and%20Events/News/140203CDFFacebook.aspx

With the scores of the social authority, we were able to identify the most powerful individuals in the network composition. However, what is more interesting is that we were also able to identify the powerfulness of the individuals who had unfollowed the Twitter accounts, and see what their impact would have been had they not unfollowed the Twitter accounts. The results of the social authority are shown in Table 25.

	Mean
Average social authority in old network	9.37
Average social authority in new network	9.52
Percentual difference in social authority	+2%
Average social authority of unfollowed and still active users	10.24
Average social authority of joined users	9.38
Impact on the social authority if the unfollowed and active users had not left	+2%

Table 25: General summary of the social authority scores

As explained before, the measure of social authority is based on retweets. In order to validate this claim, a Pearson's r correlation was run to determine the relation between the 17 Twitter account's SA and ANRET values. This resulted in a strong, positive correlation between ASA and ANRET (r=0.700, N=17, p<0.001).

		ANRT
SA	Pearson correlation (r)	0.700**
	Statistical significance (p)	0.001
	Sample size (N)	17

**. Correlation is significant at the 0.01 level (2-tailed)

Legend	
SA	Social authority Refers to the social authority score of a corporate Twitter account.
ANRET	Average number of retweets per exclusive tweet Refers to the average number of times an exclusive tweet of a corporate Twitter account has been retweeted.

Table 26: Results of the Pearson's correlation test to determine the relation between the 17 Twitter account's ASA and ANRT values

The scatterplot of the strong correlation between the SA and ANRET values is also illustrated in the scatterplot of Figure 15. Based on these results, we can indeed justify the claim that the social authority score is based on the number of retweets.

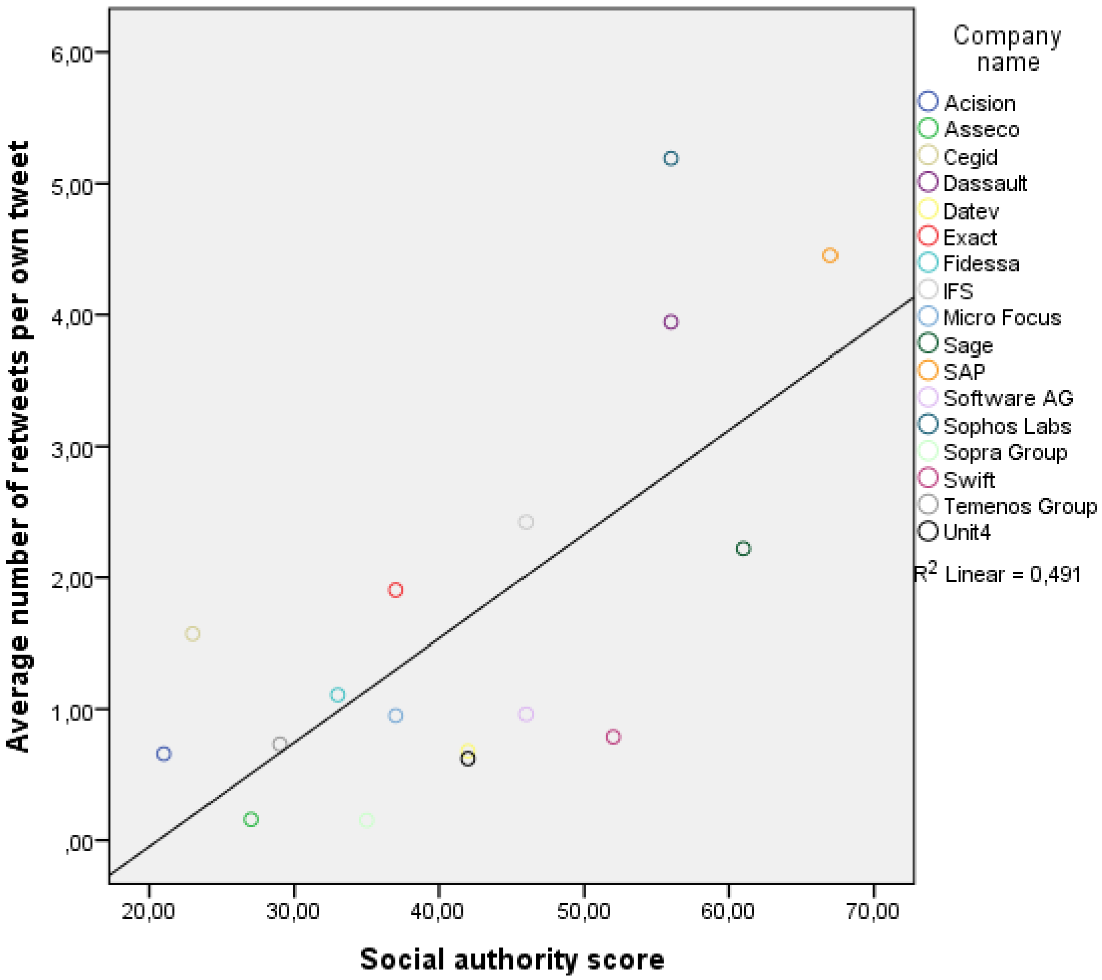

Figure 15: Scatterplot to show the strong, positive correlation between SA and ANRET

COMPETITORS

Competitive intelligence is very important for the success of organizations competing at one industry (Trim, 2004). In this study, we were interested to find the number of Twitter accounts from the Truffle 100 companies that are following the top 25 product software companies. In order to find the number of Twitter accounts that belong to the companies from the Truffle 100 companies, we first identified all Twitter accounts that were related to one of these companies in our data set. We manually removed some accounts which had nothing to do with the companies. For each of the top 25 product software companies, we used their list of followers (denoted by A), together with the competitors' Twitter accounts (denoted by B) to calculate the ratio of competitors present in their composition. The formula that was performed for this calculation is shown below:

$$C\mu(A, B) = \frac{(A \cap B)}{A}$$

The full list of the calculation for each of the top 25 product software companies can be found in Table 12 of Appendix G. A general summary of the results is presented in Table 27. Interesting results are SAP's presence on Twitter, having 27 accounts which follow 12 of the other 20 companies with a Twitter account. They were followed by Sage, who have 13 accounts that follow 10 of the other 20 companies. Furthermore, what is also interesting to see is that Sophos, who operates in the IT security industry is followed by 14 accounts which are related to other antivirus companies. A possible explanation could be the importance for antivirus companies to be up to date with the latest viruses

	Competitors	% Competitors
Mean	8.19	0.21%
St dev	9.66	0.19%
Legend		
Competitors	Refers to the total number of Twitter accounts that belong to one of the Truffle 100 companies and follow the Twitter account of one of the top 25 product software companies.	
% Competitors	Refers to the degree of competitors in relation to the total network size of one of the top 25 product software companies.	

Table 27: General summary of the competitors analysis

ACTIVITY

The final section of the data analysis part concerned the analysis on the tweeting strategy of the corporate Twitter accounts. In this part, a third-party plugin on Followerwonk enabled us to analyse the activity of the Twitter account's followers. Based on this data, the plugin was then able to compute the most active moment of all followers and suggest the best possible time for a company to share content on Twitter. We compared this data with the most active moment a corporate Twitter accounts shares content on Twitter. Our findings that none of the 21 corporate Twitter accounts has shared most of their content on Twitter during the most active hour of their followers (Table 28). These findings show that while companies have been using Twitter for years, they still lack behind in the identification of their followers.

Company	MAMF	ISMTIA
Acision	14:47	No
Asseco	11:46	No
CegidPresse	11:23	No
Dassault Systems	18:11	No
Datev	11:35	No
Exact Software	16:48	No
Fidessa	17:48	No
HP Autonomy	16:27	No
IFS World	22:51	No
Invensys	16:12	No
Micro Focus	17:24	No
Misys	16:07	No
Sage	17:20	No
SAP	17:36	No
SoftwareAG	16:32	No
Sophos	16:16	No

Sopra Group	14:26	No
SWIFT	16:58	No
Swisslog	15:04	No
Temenos	17:53	No
UNIT4	13:01	No
Legend		
MAMF	Refers to most active moment of the corporate Twitter account’s followers in terms of time (hh:mm).	
ISMTIA	Indicates whether the corporate Twitter accounts have taken the MAMF into account when sharing content on Twitter.	

Table 28: Most active moment of the followers of the corporate Twitter accounts

6 CONLUSION & DISCUSSION

In this chapter, we refer back to the introduction of this study where the research questions were defined. The first section, gives an overview of the answered research question. In the second section, we will discuss the limitations of the research and possibilities for future research.

6.1 CONCLUSION

The main research question that was formulated for this research was:

> *"What is the structure and composition of online Twitter communities and how does it evolve?"*

In order to answer the research question, we first give an overview of the answered sub questions.

> *"What are online communities?"*

The first sub-question of this study concerned the creation of a theoretical background. The theoretical background was based on the following topics: online communities, research on Twitter and social network analysis (SNA).

In the first part of the theoretical background, we looked into the development of online communities. Based on that, we found that in the beginning of the nineties, three types of online communities were used frequently: E-mail lists, chat rooms, and discussion boards. This pattern changed in the late nineties when a new type of online community was introduced: social networking sites (SNSes). In the beginning of the SNSes life-cycle, the key difference between SNSes and previous online communities was the fact that SNSes were mainly organized around individuals, and not interests. However, due to the continuous development of SNSes, this pattern has changed. Nowadays, SNSes are not only organized around individuals but also interests. The latter is also evident in our research as we are researching the Twitter accounts of product software companies.

The theoretical background was continued by looking into literature regarding previous research on Twitter. We found a large number of researches on topics such as the adoption of Twitter, the value of Twitter for businesses, the distribution of content on Twitter, the characteristics of the Twitter users, how Twitter is used for making predictions, and the study of social networks on Twitter. Based on the latter, we continued with the final section of the theoretical background where we looked into the SNA method and possible tools which we could use to perform our analysis.

> *"What particular structure do these networks of the top 25 product software companies in Europe have? (e.g. small network)"*

In our second sub-question, we wanted to explore the structural characteristics of the network and see whether the networks still have small-network characteristics. First, we looked into the degree of the reciprocal relations between the nodes in the Twitter follower networks. The mean of all reciprocity (arc) values was found to be declined from 0.49 to 0.46. The decline of the reciprocity was expected as previous research has shown that the reciprocity value declines upon a significant increase in the network size (Cao et al., 2012). In the next step, we looked into the average degree of the Twitter follower networks. The average degree of the nodes is increased from 18.58 to 19.86, which is higher than the average degree of the European Twitter community, 16.42 (Java et al., 2007). This indicates that the nodes in the Twitter follower networks are densely interconnected. In addition to that, this also means an increase in the potential reach and exposure of tweets for the companies.

Finally, we were interested to see the form of the network. In particular, we were interested to see whether the Twitter follower networks still show signs of being a small-world network. In order to do that, we used the small-world-ness measure where a network is found to be small-network if the value of (S) > 1. In the previous study, the Twitter follower networks were found to be small-network with a small-world-ness index of 13.47. In this study, the small-world-ness index was found to be 16.46, which clearly shows that the networks have not only kept their small-world characteristics but even exhibit stronger small-world characteristics. This pattern was also identified in calculation of the average path length. In a small-world graph, the average path length is expected to be the same or smaller than the log(n) of the network, where n is the network size. In addition, the average path length is expected to be growing proportionally with the growth of the log(n) in a small-network graph. Based on that theory, we compared the growth of the average path length the log(n). The results showed an increase of 2.89% in the average path length value, and an increase of 10.97% in the log(n) value. This clearly indicates the network to exhibit even stronger small-world characteristics now. Stronger small-world characteristics, indicate an increasing number of hubs which facilitate a faster spread of information. Based on this, we can conclude that the information flow in the Twitter follower networks has become more efficient and effective.

> *"Who are the followers of these product software companies?"*

For the third sub-question, we wanted to explore the network composition and identify the followers of the corporate Twitter accounts. In addition to that, we wanted to see whether companies have identified their target audience themselves.

In the first part of this analysis, we explored the general attributes of the networks, which were also analysed in the previous study. In order to identify the number of followers that are related to the corporate Twitter accounts, we calculated the degree of the internal audience. Our analysis shows an increase of 78% in the degree of the internal audience of all Twitter accounts. In order to explain the growth, we performed a correlation test to see whether it is association with the change in the employee size of the R&D department. We found no strong association there, but we believe that an association might be found if the degree of internal audience is compared with data about the web care departments of the companies. Finally, we believed that another possible explanation for the growth could be could be related to the fact that increasingly more companies tend to invest more on social media (Brotherton, 2012). As a result, it could be very possible that they also push their employees to have a presence on Twitter.

The following part concerned the analysis of the network overlap. First, we used the Jaccard index, to measure the degree of users which follow company A and B. In comparison with the results of the previous study, we found that the network overlap was increased by 22%. In addition, we also saw an increase in the most frequent followers. The ratio of nodes who follow six or more companies doubled, from 0.006% to 0.013%. We believe that a possible explanation for the increase in the network overlap and the most frequent followers could be related to the growth of Twitter, and the clients or competitors who use SNSes more and more these days to keep up to date with companies from the same industry. Based on the latter, we also performed a competitors analysis to find the total amount of Twitter accounts related to the companies of the Truffle 100 list which follow one of the top 25 product software companies in Europe. Our analysis shows that companies such as SAP and SageUK, have a very strong presence of Twitter. We found, 24 Twitter accounts which were each related to different business units of SAP, and followed 12 of the other 20 companies with a Twitter account.

Finally, we looked into the Twitter strategy of the companies, regarding the activity of their account. Using Followerwonk, we were able to identify the highest peak time of the followers of each of the top 25 product software companies. We compared this data, with the activity of the companies. Our findings showed that none of the companies has posted most of their tweets during the most active hour of their followers. This clearly indicates, that the information (tweets and retweets) shared by the companies, could receive more exposure, if the companies consider changing their tweeting strategy.

"To what extend does the composition and structure evolve over time?"

After identifying the network structure and composition, and thus answering the first part of the main research question of this study, we performed several correlations tests to identify patterns related to the changes in the network structure and composition. We first looked into the changes in the network structure. The results of the tests, showed that companies that had posted more tweets per day, had faced a higher number of unfollowed users. A possible explanation for this phenomenon could be that followers who see too much tweets of one account could receive a feeling of spam. For the opposite, the increase of followers size, we found that the amount of retweets and hashtags have a positive impact on the increase in network size. This phenomenon can be explained by the fact that the number of retweets and hashtags increases the visibility of the corporate Twitter accounts on Twitter. Finally, we performed several tests related to the SNA metrics and the network composition and our most important conclusion is that the Twitter follower networks of the corporate Twitter accounts are becoming more small-world. This was also shown in the correlation tests where we found a strong association between the increase in small-world-ness and in terms of percentage increase of the network size.

6.1 DISCUSSION & FUTURE RESEARCH

In this study, we gave an explorative view of how the Twitter follower networks of the corporate Twitter accounts of the top 25 product software companies change over time. However, for more concrete conclusions regarding the changes, we believe that additional research might be useful.

Regarding the changes in the network structure, we believe that stronger conclusions could be drawn if this study is repeated over multiple time periods. In addition to that, analysis of the data from multiple periods of time would enable us to get a better insight in the evolvement of the Twitter follower networks. Furthermore, it is important to note that in this study we looked into the centrality using the average degree of the nodes. We believe that with future research, focusing on the centrality of the nodes, could provide more concrete conclusions regarding the changes in the powerfulness characteristics of the users in the network. For example, it would be interesting to see what type of an user a node is. We believe that there could be users in the network that have a low degree but are important as they operate as a broker and thus, connect other groups of individuals. Finally, it would also be interesting to see whether the changes in the networks are related to a specific domain in the industry. For example, do the small and medium enterprises change differently than the large companies?

Regarding the research on the composition, we believe that future research which emphasizes on the composition could help companies to propose a strategic plan for their Twitter presence. In particular, our results on the activity of the corporate Twitter accounts, in relation to the activity of their followers showed that this is an area where companies need to work on. Finally, in this study, we performed several correlation tests to see the relation between the changes in the network structure and network size. However, additional research could also emphasize on the relation between the structural and compositional information and see how the two can be related.

7 BIBLIOGRAPHY

Abel, F., Gao, Q., Houben, G. J., & Tao, K. (2013). Twitter-based user modeling for news recommendations. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence* (pp. 2962-2966). AAAI Press.

Ahmad, A. (2011). Social Network Sites and Its Popularity. *International Journal of Research & Reviews in Computer Science*, *2*(2).

Ang, L. (2011). Community relationship management and social media. *Journal of Database Marketing & Customer Strategy Management*, *18*(1), 31-38.

Ariyasu, K., Fujisawa, H., & Sunasaki, S. (2013). Twitter analysis algorithms for Intelligence Circulation System. *Multimedia systems*, *19*(6), 477-491.

Ayyadurai, V. S. (2013). The Email Revolution: How to Build Brands and Create Real Connections. Skyhorse Publishing, Inc..

Bader, J. S., Chaudhuri, A., Rothberg, J. M., & Chant, J. (2004). Gaining confidence in high-throughput protein interaction networks. *Nature biotechnology*, *22*(1), 78-85.

Bassecoulard, E., & Zitt, M. (1999). Indicators in a research institute: A multi-level classification of scientific journals. *Scientometrics*, *44*(3), 323-345.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1-8.

Borgatti, S.P., Everett, M.G., & Freeman, L.C. (2002). Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Bray, P., & Peters, M. (2013). *Social Authority: Our Measure of Twitter Influence*. Retrieved May20, 2013, from http://www.moz.com/blog/social-authority

Breslin, J. G., Harth, A., Bojars, U., & Decker, S. (2005). Towards semantically-interlinked online communities. In *The Semantic Web: Research and Applications* (pp. 500-514). Springer Berlin Heidelberg.

Brinkkemper, S. (1996). Method engineering: engineering of information systems development and tools. *Elsevier, Information and software technology*, *38*(4), 275-280.

Brotherton, P. (2012). Social media and referrals are best sources for talent: a new survey shows that companies are investing more and more of their recruitment resources in social media networks and seeing it pay off. *T+ D*,*66*(1), 24.

Burton, S., & Soboleva, A. (2011). Interactive or reactive? Marketing with Twitter. *Journal of Consumer Marketing*, *28*(7), 491-499.

Butts, C. T. (2007). Software Manual for the R sna Package. *R package version*, *1*.

Butts, C. T. (2008). network: a Package for Managing Relational Data in R. *Journal of Statistical Software*, *24*(2), 1-36.

Cao, Y., Wan, Q., Lu, Y., Quan, J., & Chen, J. (2012). The dynamics of a virtual community during a natural disaster: a network analysis. *International Journal of Internet and Enterprise Management*, *8*(1), 1-15.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. *ICWSM*, *10*, 10-17.

Cheong, F., & Cheong, C. (2011). Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods. In *PACIS* (p. 46).

Cherven, K. (2013). Network Graph Analysis and Visualization with Gephi. Packt Publishing Ltd.

Coe, D. (2013). Companies and shareholder engagement Making the most of social media. *Keeping Good Companies*, *65*(5), 307.

Cowan, R., & Jonard, N. (2004). Network structure and the diffusion of knowledge. *Journal of economic dynamics and control*, *28*(8), 1557-1575.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5).

Culnan, M. J., McHugh, P. J., & Zubillaga, J. I. (2010). How large US companies can use Twitter and other social media to gain business value. *MIS Quarterly Executive*, *9*(4), 243-259.

Date, C. J., & Darwen, H. (1987). *A Guide to the SQL Standard* (Vol. 3). New York: Addison-Wesley.

Davidsen, J., Ebel, H., & Bornholdt, S. (2002). Emergence of a small world from local interactions: Modeling acquaintance networks. *Physical Review Letters*, *88*(12), 128701.

Donath, J., & Boyd, D. (2004). Public displays of connection. *bt technology Journal*, *22*(4), 71-82.

Driskell, R. B., & Lyon, L. (2002). Are virtual communities true communities? Examining the environments and elements of community. *City & Community*,*1*(4), 373-390.

Ellis, D., Oldridge, R., & Vasconcelos, A. (2004). Community and virtual community. *Annual review of information science and technology*, *38*, 145-188.

Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210-230.

FossoWamba, S., & Carter, L. (2013, January). Twitter adoption and use by SMEs: An empirical study. In *The 46th Hawaii International Conferences on System Sciences (HICSS), Maui, Hawaii*.

Freeman, L. C. (2011). The Development of Social Network Analysis—with an Emphasis on Recent Events. *The SAGE handbook of social network analysis*, 26-54.

Gruzd, A., Wellman, B., & Takhteyev, Y. (2011). Imagining Twitter as an imagined community. *American Behavioral Scientist*, *55*(10), 1294-1318.

Govani, T., & Pashley, H. (2005). Student awareness of the privacy implications when using Facebook. unpublished paper presented at the “Privacy Poster Fair” at the Carnegie Mellon University School of Library and Information Science, 9.

Guare, J. (1992). *Six degrees of separation*. Dramatists Play Service, Inc..

Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods. University of California, Riverside. Published in digital form at http://faculty.ucr.edu/~hanneman

Hargittai, E., & Litt, E. (2012). Becoming a tweep: How prior online experiences influence twitter use. *Information, Communication & Society*, *15*(5), 680-702.

Hassan, A., Abbasi, A., & Zeng, D. (2013, September). Twitter Sentiment Analysis: A Bootstrap Ensemble Framework. In *Social Computing (SocialCom), 2013 International Conference on* (pp. 357-364). IEEE.

Helms, R. W., & Werder, K. (2013). Who Reads Corporate Tweets? Network Analysis of Follower Communities.

Heymann, S., & Le Grand, B. (2013). Visual Analysis of Complex Networks for Business Intelligence with Gephi. In Proceedings of the 1st International Symposium on Visualisation and Business Intelligence, in conjunction with the 17th International Conference Information Visualisation.

Hosseini, S., Unankard, S., Zhou, X., & Sadiq, S. (2014, January). Location Oriented Phrase Detection in Microblogs. In *Database Systems for Advanced Applications* (pp. 495-509). Springer International Publishing.

Huberman, B. A., Romero, D. M., & Wu, F. (2008). Social networks that matter: Twitter under the microscope. *arXiv preprint arXiv:0812.1045*.

Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, *6*(3), 248-260.

Humphries, M. D., & Gurney, K. (2008). Network ‘small-world-ness’: a quantitative method for determining canonical network equivalence. *PloS one*,*3*(4), e0002051.

Hutchings, C. (2012). Commercial use of Facebook and Twitter–risks and rewards. *Computer Fraud & Security*, *2012*(6), 19-20.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, *60*(11), 2169-2188.

Java, A., Song, X., Finin, T., & Tseng, B. (2007, August). Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 56-65). ACM.

Jin, E. M., Girvan, M., & Newman, M. E. (2001). Structure of growing social networks. *Physical review E*, *64*(4), 046132.

Kafai, Y. B., Fields, D. A., & Burke, W. Q. (2010). Entering the clubhouse: Case studies of young programmers joining the online Scratch communities.*Journal of Organizational and End User Computing (JOEUC)*, *22*(2), 21-35.

Kaplan, A. M., & Haenlein, M. (2011). The early bird catches the news: Nine things you should know about micro-blogging. *Business Horizons*, *54*(2), 105-113.

Kierkegaard, S. (2010). Twitter thou doeth?. *Computer Law & Security Review*,*26*(6), 577-594.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, *54*(3), 241-251.

Kim, J., Lee, E., Choi, J., Bae, Y., Ko, M., & Kim, P. (2013). Monitoring social relationship among Twitter users by using NodeXL. In *Proceedings of the 2013 Research in Adaptive and Convergent Systems* (pp. 107-110). ACM.

Kim, Y., & Srivastava, J. (2007, August). Impact of social influence in e-commerce decision making. In *Proceedings of the ninth international conference on Electronic commerce* (pp. 293-302). ACM.

Kottke, J. (2005). Tumblelogs. Retrieved May 10, 2013, fromhttp://www.kottke.org/05/10/tumblelogs

Krentler, K. A., & Willis-Flurry, L. A. (2005). Does technology enhance actual student learning? The case of online discussion boards. *Journal of Education for Business*, *80*(6), 316-321.

Kwak, H., Chun, H., & Moon, S. (2011). Fragile online relationship: a first look at unfollow dynamics in twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1091-1100). ACM.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media?. In *Proceedings of the 19th international conference on World wide web* (pp. 591-600). ACM.

Lee, R., Wakamiya, S., & Sumiya, K. (2011). Discovery of unusual regional social activities using geo-tagged microblogs. *World Wide Web*, *14*(4), 321-349.

Meola, M., & Stormont, S. (2000). Real-time reference service for the remote user: from the telephone and electronic mail to Internet chat, instant messaging, and collaborative software. *The Reference Librarian*, *32*(67-68), 29-40.

McWilliam, G. (2012). Building stronger brands through online communities. *Sloan management review*, *41*(3).

Mulvaney, C. (2012). How to Increase Firm Revenue Using Internet Marketing.*PRESIDENT'S PERSPECTIVE*, 20.

Latha, R. H., & Sathiyakumari, K. (2012). Predicting Link Strength In Online Social Networks. *International Journal of Engineering Research and Applications (IJERA)*, *2*(6), 703-707.

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: International Journal of an Emerging Transdiscipline*, *9*, 181-212.

Lunden, I. (2012). *Analyst: Twitter Passed 500M Users In June 2012, 140M Of Them In US; Jakarta 'Biggest Tweeting' City*. Retrieved October 10, 2013, from http://techcrunch.com/2012/07/30/analyst-twitter-passed-500m-users-in-june-2012-140m-of-them-in-us-jakarta-biggest-tweeting-city/.

Namayandeh, R., Didehvar, F., & Shojaei, Z. (2013). Clustering validity based on the most similarity. *arXiv preprint arXiv:1302.3956*.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, *45*(2), 167-256.

Ojeda-Zapata, J. (2008). Twitter means business: How microblogging can help or hurt your company. Happy About.

Preece, J. (2000). *Online communities: Designing usability, supporting sociability*. Chichester, UK: John Wiley & Sons.

Rogers, I. (2002). The google pagerank algorithm and how itworks. Retrieved April , 2014 from http://www.iprcom.com/papers/pagerank/

Shavitt, Y., & Weinsberg, U. (2009, April). Quantifying the importance of vantage points distribution in internet topology measurements. In *INFOCOM 2009, IEEE* (pp. 792-800). IEEE.

Shafranovich, Y. (2005). Common format and MIME type for Comma-Separated Values (CSV) files.

Siles, I. (2012). The rise of blogging: Articulation as a dynamic of technological stabilization. *New Media & Society*, *14*(5), 781-797.

Sippey, M. (2012). *Changes coming in Version 1.1 of the Twitter API*. Retrieved November 26, 2013, from https://dev.twitter.com/blog/changes-coming-to-twitter-api.

Sloan, B. (2006). Electronic discussion lists. *Journal of library administration*,*44*(3-4), 203-225.

Stocker, K. (2013). *Battle of the social network climbers*. Retrieved February10, 2014, from http://www.drapersonline.com/digital/social-media/battle-of-the-social-network-climbers/5052344.article

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010, August). Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Social computing (socialcom), 2010 ieee second international conference on* (pp. 177-184). IEEE.

Taubman, G. (2002). Keeping out the Internet? Non-democratic legitimacy and access to the Web. *First Monday*, *7*(9).

Team, R. C. (2005). R: A language and environment for statistical computing. *R foundation for Statistical Computing*.

Thijs, V., Lemmens, R., &Fieuws, S. (2008). Network meta-analysis: simultaneous meta-analysis of common antiplatelet regimens after transient ischaemic attack or stroke. *European heart journal*, *29*(9), 1086-1092.

Trim, P. R. (2004). The strategic corporate intelligence and transformational marketing model. *Marketing Intelligence & Planning*, *22*(2), 240-256.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM*, *10*, 178-185.

Ullrich, C., Borau, K., & Stepanyan, K. (2010). Who students interact with? a social network analysis perspective on the use of twitter in language learning. In *Sustaining TEL: From Innovation to Learning and Practice* (pp. 432-437). Springer Berlin Heidelberg.

Wasserman, S. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *nature*, *393*(6684), 440-442.

Webster, J., & Watson, R. T. (2002). ANALYSING THE PAST TO PREPARE FOR THE FUTURE: WRITING A. *MIS quarterly*, *26*(2).

Weng, J., Lim, E. P., Jiang, J., & He, Q. (2010, February). Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining* (pp. 261-270). ACM.

Wilson, H. J., Guinan, P. J., Parise, S., & Weinberg, B. D. (2011). What's your social media strategy. *Harvard Business Review*, *89*(7/8), 23-25.

Yardi, S., Romero, D., & Schoenebeck, G. (2009). Detecting spam in a twitter network. *First Monday*, *15*(1).

Yarow, J. (2014). *Check Out Twitter's Growth Versus Facebook*. Retrieved April28, 2013, from http://www.businessinsider.com/chart-facebook-twitter-growth-2013-12

Yep, J., & Shulman, J. (2014). Analysing the library's Twitter network Using NodeXL to visualize impact. *College & Research Libraries News*, *75*(4), 177-186.

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences*, *26*, 55-62.

Zhou, L., Zhang, P., & Zimmerman, H. D. (2011). Call for Papers for a Series of Special Issues: Social Commerce. *Electronic Commerce Research and Applications*.

APPENDIX A

Company	Date 1	Date 2	Days diff	Tot Tweets	Regular Tweets	Retweets	Replies	Total hashtags	Total mentions	Total links	Total of received retweets	Received retweets on regular tweets
Acision	08-05-2012	19-12-2013	590	679	563	83	33	357	416	585	539	370
Asseco	08-05-2012	12-12-2013	583	1330	1219	86	25	817	252	1112	4643	194
CegidPresse	08-05-2012	30-11-2013	571	120	79	41	0	128	47	87	254	124
Dassault Systems	17-05-2012	02-02-2014	626	2379	1396	630	353	2598	3521	1509	13619	5507
Datev	19-05-2012	07-01-2014	598	833	529	108	196	343	435	535	563	359
Exact Software	10-05-2012	23-12-2013	592	861	717	74	70	1108	659	526	1570	1365
Fidessa	09-05-2012	02-12-2013	572	624	360	243	21	600	826	393	1034	399
IFS World	08-05-2012	08-12-2013	579	810	649	105	56	1358	385	582	2150	1572
Micro Focus	13-06-2013	19-12-2013	188	1726	876	761	89	2544	2956	1357	2481	832
Sage	31-07-2013	19-03-2014	230	2494	1305	451	738	1690	2199	1701	4430	2896
SAP	11-05-2013	06-03-2014	298	2571	2276	249	46	2911	1333	2438	14266	10135
SoftwareAG	12-05-2012	29-12-2013	596	1985	1616	308	61	3517	1357	1663	2319	1552
Sophos	13-05-2012	08-01-2014	605	1885	1773	32	80	135	246	1791	9654	9206
Sopra Group	09-05-2012	12-12-2013	582	2127	2126	0	1	2121	3	2126	325	325
SWIFT	26-09-2012	09-12-2013	438	2444	1822	76	546	2244	2148	1995	1689	1433
Temenos	11-05-2012	09-12-2013	577	470	357	69	44	438	339	313	397	261
UNIT4	08-05-2012	04-12-2013	575	1401	1234	20	147	1127	771	973	804	766

Table 1: Analysis of Tweets

APPENDIX B

Rank	Company	HQ	SW Revenue*	Employees	Edges	Nodes (followers)	2nd Level Followers
1	**SAP**	DE	12 336.7	14 991	499,175	54,536	39,013,657
2	**Dassault Systems**	FR	1 563.8	3 700	105,821	4,289	2,512,420
3	**Sage**	UK	1 542.9	*2 076	432,296	10,178	25,948,372
4	**Software AG**	DE	919.2	850	12,738	1,856	5,159,745
5	**Datev**	DE	684.6	*1 250	353,258	3,600	32,175,260
6	**HP Autonomy**	UK	657.0	*563	347,189	2,199	2,678,872
7	**Asseco**	PL	516.4	*2 047	17,134	141	1,975,229
8	**Swift**	BE	511.1	*452	24,622	1,515	2,162,271
9	**Wincor Nixdorf**	DE	461.6	*372	613	398	215,028
10	**Misys**	UK	431.2	*1 102	2,298	332	96,661
11	**Unit4**	NL	421.7	1 150	6,933	766	505,657
12	**Sopra Group**	FR	354.7	1 000	984	538	183,732
13	**Temenos Group**	CH	338.2	*617	1,920	539	352,338
14	**Swisslog**	CH	324.9	*511	53,429	1,449	18,701,506
15	**Micro Focus**	UK	322.7	*300	5,144	1,148	5,321,962
16	**Compugroup Holding**	DE	312.4	*900	n/a	n/a	n/a
17	**Murex**	FR	310.0	255	2	9	1,739
18	**Invensys**	UK	279.2	1 328	1,542	2,616	116,301
19	**NIS**	UK	269.0	*760	n/a	n/a	n/a
20	**IFS**	SE	264.0	524	1,062	297	216,149
21	**Acision**	UK	260.4	486	2,685	683	269,599
22	**Sophos Labs**	UK	259.4	*600	251,321	12,146	5,733,759
23	**Fidessa**	UK	228.8	*300	19,369	1,041	1,626,394
24	**Exact**	NL	228.2	456	17,109	1,768	2,517,519
25	**Cegid**	FR	218.0	537	3,987	375	134,051

Table 2: Network size of all companies (old situation)

Rank	Company	HQ	SW Revenue*	Employees	Edges	Nodes (followers)	2nd Level Followers
1	**SAP**	DE	15 930.0		1,200,556	103,096	83,514,005
2	**Dassault Systems**	FR	1 853.4		285,198	14,641	14,408,373
3	**Sage**	UK	1 591.4		513,894	21,227	52,953,307
4	**Software AG**	DE	922.2		52,751	4,061	21,512,147
5	**Datev**	DE	736.7		353,258	4,607	38,018,153
6	**HP Autonomy**	UK	n/a		68,196	4,687	15,449,641
7	**Asseco**	PL	1002.1		3,036	294	3,244,813
8	**Swift**	BE	594.9		77,435	3,674	4,988,234
9	**Wincor Nixdorf**	DE	1 257.3		n/a	n/a	n/a
10	**Misys**	UK	454.3		19,087	1,914	9,270,462
11	**Unit4**	NL	469.8		15,684	1,390	2,847,718
12	**Sopra Group**	FR	354.7		4,144	1,329	487,217
13	**Temenos Group**	CH	350.4		12,370	1,430	691,944
14	**Swisslog**	CH	403.7		1,493	159	621,544
15	**Micro Focus**	UK	323.1		16,345	1,830	7,516,233
16	**Compugroup Holding**	DE	324.6		n/a	n/a	n/a
17	**Murex**	FR	318.0		n/a	n/a	n/a
18	**Invensys**	UK	304.1		66,635	2,751	3,147,271
19	**NIS**	UK	389.7		n/a	n/a	n/a
20	**IFS**	SE	307.2		17,818	5,328	14,424,151
21	**Acision**	UK	516.4		6,215	989	1,445,720
22	**Sophos Labs**	UK	315.2		605,123	24,460	20,819,756
23	**Fidessa**	UK	343.4		61,007	2,188	3,696,151
24	**Exact**	NL	217.1		47,081	2,918	12,274,888
25	**Cegid**	FR	226.0		9,438	626	305,069

Table 3: Network size of all companies (new situation)

Rank	Company	Nodes left	Nodes left (inactive)	Nodes joined	Old Avg SA	Old Avg SA (active accounts)	New Avg SA
1	**SAP**	23%	9.20%	60952	6.28	6.40	6.49
2	**Dassault Systems**	15%	7.34%	10997	8.77	8.55	8.72
3	**Sage**	24%	15.57%	13511	9.62	9.50	9.46
4	**Software AG**	17%	8.73%	2513	8.52	7.43	10.34
5	**Datev**	23%	5.22%	1455	12.72	9.22	11.70
6	**HP Autonomy**	21%	12.37%	2946	9.55	9.58	9.27
7	**Asseco**	25%	16.31%	188	17.69	16.00	19.73
8	**Swift**	16%	8.05%	2395	9.85	9.18	9.03
9	**Wincor Nixdorf**	n/a	n/a	n/a	n/a	n/a	n/a
10	**Misys**	84%	71.08%	1860	7.13	6.02	8.80
11	**Unit4**	17%	8.09%	756	8.88	5.97	9.16
12	**Sopra Group**	19%	9.67%	894	7.63	8.20	7.74
13	**Temenos Group**	16%	8.53%	977	9.32	8.35	7.31
14	**Swisslog**	100%	91.30%	154	14.31	14.33	9.08
15	**Micro Focus**	18%	9.32%	884	8.44	7.48	9.01
16	**Compugroup Holding**	n/a	n/a	n/a	n/a	n/a	n/a
17	**Murex**	n/a	n/a	n/a	n/a	n/a	n/a
18	**Invensys**	73%	65.26%	2587	6.18	6.39	6.32
19	**NIS**	n/a	n/a	n/a	n/a	n/a	n/a
20	**IFS**	12%	4.71%	5068	7.17	6.15	8.63
21	**Acision**	15%	9.08%	411	7.39	7.92	7.78
22	**Sophos Labs**	15%	7.01%	14171	8.75	7.82	8.03
23	**Fidessa**	18%	7.59%	1337	9.31	8.70	8.83
24	**Exact**	14%	7.47%	1394	7.00	6.53	7.58
25	**Cegid**	18%	7.73%	315	11.95	9.97	12.35

Table 4: Changes in network size

APPENDIX C

Company	Reciprocity (arc)	Reciprocity (dyad)	Avg degree	Avg path length	Avg clustering coefficient	Small-world-ness	Log(n)
SAP	0.15	0.08	3.08	4.31	0.27	70.18	4.74
Dassault Systems	0.46	0.3	25.78	2.98	0.505	14.52	3.63
Sage	0.62	0.45	45.1	2.98	0.346	19.83	4.01
SoftwareAG	0.46	0.3	8.4	3.37	0.461	16.03	3.27
Datev	0.87	0.77	105.69	2.49	0.418	4.48	3.56
HP Autonomy	0.55	0.38	9.24	3.61	0.445	17.92	3.34
Asseco	0.63	0.46	8.73	2.51	0.556	2.73	2.15
SWIFT	0.42	0.27	18.54	2.86	0.521	7.63	3.18
MISYS	0.41	0.26	7.79	3.21	0.623	4.95	2.52
UNIT4	0.53	0.36	10.14	3.13	0.567	8.34	2.88
Sopra Group	0.49	0.33	3.59	3.66	0.357	8.6	2.73
Temenos	0.53	0.36	5.16	3.52	0.36	8.6	2.73
Swisslog	0.87	0.77	39.43	2.61	0.366	3.97	3.16
Micro Focus	0.52	0.35	5.47	3.61	0.371	8.82	3.06
Invensys	0.53	0.36	4.28	3.87	0.593	10.99	2.79
IFS World	0.54	0.37	4.87	3.24	0.554	5.9	2.47
Acision	0.44	0.28	5.59	3.51	0.258	10.83	2.83
Sophos	0.32	0.19	22.55	2.9	0.514	18.78	4.08
Fidessa	0.48	0.31	20.85	2.76	0.488	5.3	3.02
Exact Software	0.57	0.4	11.44	3.53	0.453	14.04	3.25
CegidPresse	0.51	0.34	10.92	2.66	0.659	4.38	2.57
Mean	0.52	0.37	17.94	3.21	0.46	12.71	3.14
Stddev	0.16	0.16	23.19	0.48	0.11	14.16	0.61

Table 5: Network structure values (old situation)

Company	Reciprocity (arc)	Reciprocity (dyad)	Avg degree	Avg path length	Avg clustering coefficient	Small-world-ness	Log(n)
SAP	0.19	0.1	14.11	5.06	0.171	44.18	5.01
Dassault Systems	0.42	0.26	22.32	3.21	0.495	40.09	4.17
Sage	0.29	0.17	27.36	3.84	0.2	28.62	4.33
SoftwareAG	0.54	0.37	14.87	3.12	0.441	17.12	3.61
Datev	0.81	0.68	85.66	2.61	0.433	6.36	3.66
HP Autonomy	0.51	0.34	15.86	3.34	0.519	21.1	3.67
Asseco	0.6	0.42	12.24	2.69	0.466	3.54	2.47
SWIFT	0.37	0.22	23.09	3.01	0.533	11.76	3.57
MISYS	0.53	0.36	13.44	3.54	0.425	12.02	3.28
UNIT4	0.54	0.37	12.33	3.05	0.547	10.41	3.14
Sopra Group	0.43	0.27	4.16	3.91	0.379	9.93	3.12
Temenos	0.41	0.26	8.64	2.81	0.738	7.79	3.16
Swisslog	0.58	0.41	9.33	2.42	0.73	3.15	2.20
Micro Focus	0.54	0.37	9.73	3.20	0.503	14.23	3.26
Invensys	0.41	0.26	24.29	2.87	0.666	9.94	3.44
IFS World	0.55	0.38	4.88	4.04	0.423	31.27	3.73
Acision	0.42	0.26	8.33	3.46	0.452	11.08	3.00
Sophos	0.25	0.14	26.1	3.04	0.554	17.63	4.39
Fidessa	0.43	0.27	30.07	2.72	0.526	6.83	3.34
Exact Software	0.61	0.44	17.89	3.12	0.446	13.3	3.47
CegidPresse	0.5	0.33	15.78	2.69	0.573	5.69	2.80
Mean	0.47	0.32	19.07	3.23	0.49	15.53	3.47
Stddev	0.14	0.12	16.95	0.61	0.14	11.48	0.65

Table 6: Network structure values (new situation)

Company	Reciprocity (arc)	Reciprocity (dyad)	Avg degree	Avg path length	Avg clustering coefficient	Small-world-ness	Log(n)
SAP	26.67%	25.00%	358.12%	17.40%	-36.67%	-37.05%	6%
Dassault Systems	-8.70%	-13.33%	-13.42%	7.72%	-1.98%	176.10%	15%
Sage	-53.23%	-62.22%	-39.33%	28.86%	-42.20%	44.33%	8%
SoftwareAG	17.39%	23.33%	77.02%	-7.42%	-4.34%	6.80%	10%
Datev	-6.90%	-11.69%	-18.95%	4.82%	3.59%	41.96%	3%
HP Autonomy	-7.27%	-10.53%	71.65%	-7.48%	16.63%	17.75%	10%
Asseco	-4.76%	-8.70%	40.21%	7.17%	-16.19%	29.67%	15%
SWIFT	-11.90%	-18.52%	24.54%	5.24%	2.30%	54.13%	12%
MISYS	29.27%	38.46%	72.53%	10.28%	-31.78%	142.83%	30%
UNIT4	1.89%	2.78%	21.60%	-2.56%	-3.53%	24.82%	9%
Sopra Group	-12.24%	-18.18%	15.88%	6.83%	6.16%	15.47%	14%
Temenos	-22.64%	-27.78%	67.44%	-20.17%	105.00%	-9.42%	16%
Swisslog	-33.33%	-46.75%	-76.34%	-7.28%	99.45%	-20.65%	-30%
Micro Focus	3.85%	5.71%	77.88%	-11.36%	35.58%	61.34%	7%
Invensys	-22.64%	-27.78%	467.52%	-25.84%	12.31%	-9.55%	23%
IFS World	1.85%	2.70%	0.21%	24.69%	-23.65%	430.00%	51%
Acision	-4.55%	-7.14%	49.02%	-1.42%	75.19%	2.31%	6%
Sophos	-21.88%	-26.32%	15.74%	4.83%	7.78%	-6.12%	7%
Fidessa	-10.42%	-12.90%	44.22%	-1.45%	7.79%	28.87%	11%
Exact Software	7.02%	10.00%	56.38%	-11.61%	-1.55%	-5.27%	7%
CegidPresse	-1.96%	-2.94%	44.51%	1.13%	-13.05%	29.91%	9%
Mean	-5.91%	-8.25%	48.30%	3.10%	5.90%	52.23%	11.29%

Table 7: Percentual difference in network structure

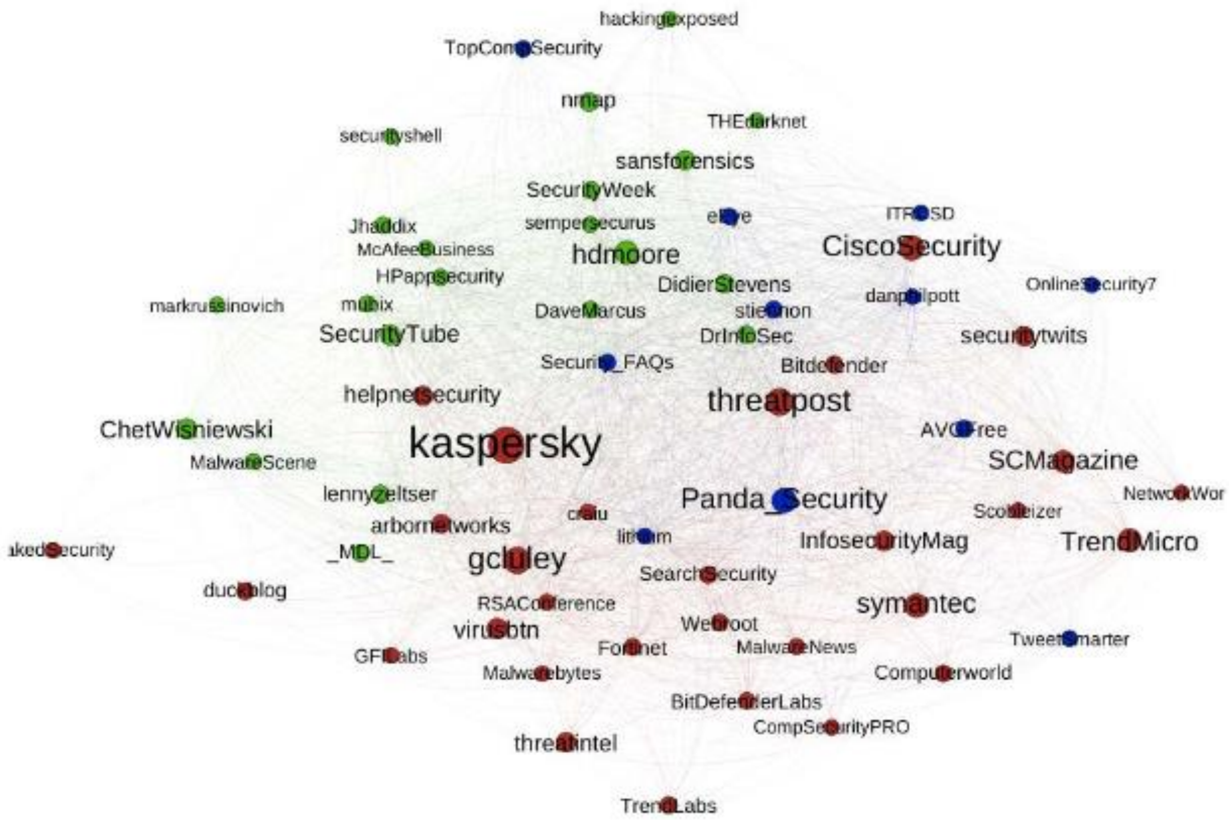

Figure 1: Pattern in the follower network with more than 1000 degree of @sophoslabs based on vendors' domain in security. Size of the nodes is based on degree. Color-coding is based on modularity measure. (Helms, &Werder, 2013)

APPENDIX D

```
graph.analyseCompany<- function (filename, company) {
	# ANALYSE KARL'S DATA
	d <- read.csv(paste(filename, "_karl.csv", sep=""), header=TRUE,
sep=";")
	g <- graph.data.frame(d, directed=TRUE, vertices=NULL)
	cat("Old data Of: ", company, "\n")
	graph.analyseData(g)
	cat("----------------")
	cat("\n")

	# ANALYSE CURRENT DATA
	d <- read.csv(paste(filename, ".csv", sep=""), header=TRUE, sep=";")
	g <- graph.data.frame(d, directed=TRUE, vertices=NULL)
	cat("New data Of: ", company, "\n")
	graph.analyseData(g)
	cat("--------------------------------------------------------------")
	cat("\n\n")
}

graph.analyseData<- function(g) {
	# GET NR OF NODES AND EDGES
	n <- vcount(g)
	e <- ecount(g)
	cat("	Nodes: ", n, "\n")
	cat("	Edges: ", e, "\n")

	# RECIPROCITY
	arc <- reciprocity(g, mode = c("default"))
	dyad<- reciprocity(g, mode = c("ratio"))
	cat("	Reciprocity (ARC): ", arc, "\n")
	cat("	Reciprocity (DYAD): ", dyad, "\n")

# AVERAGE DEGREE
	deg<- degree(g, mode=c("out"))
	adeg<- mean(deg)
	cat("	Average Degree: ", adeg, "\n")

	# AVERAGE CLUSTERING COEFFICIENT
	aclu<- transitivity(g, type="average")
	cat("	Average Clustering Coefficient: ", aclu, "\n")

	# AVERAGE PATH LENGTH
	avdist<- average.path.length(g, directed=TRUE, unconnected=TRUE)
	cat("	Average Path Length", avdist, "\n")

	# CLUSTERING COEFFICIENT
	clu<- transitivity(g, type="global")
	cat("	Clustering Coefficient: ", clu ,"\n")

	# CREATE ER RANDOM GRAPH
	er<-erdos.renyi.game(n,e, type="gnm",directed=TRUE)

	# ER CLUSTERING COEFFICIENT
	eclu<-transitivity(er, type="global")
	cat("	ER Clustering Coefficient: ", eclu ,"\n")

	# ER AVERAGE PATH LENGTH
	disthist_er<- path.length.hist(er, directed=TRUE)
	diameter_er<- length(disthist_er)
	eavdist<- weighted.mean(1:diameter_er, disthist_er)
	cat("	ER Average Path Length", eavdist, "\n")

	# CALCULATE SMALL WORLD INDEX
	sindex<- (clu/eclu) / (avdist/eavdist)
	cat("	Small-world-index", sindex, "(",clu," / ",eclu,") /
(",avdist,"/",eavdist,")	\n")
}
```

APPENDIX E

Rank	Company	Old IA - @name	Old IA - @name (%)	New IA - @name & bio	New IA - @name & bio (%)	New IA - @name	New IA - @name (%)
1	SAP	1027	1.88%	2721	2.64%	1867	1.81%
2	Dassault Systems	3	0.07%	58	0.40%	4	0.03%
3	Sage	176	1.73%	365	1.72%	226	1.06%
4	Software AG	8	0.43%	52	1.28%	18	0.44%
5	Datev	4	0.11%	17	0.37%	6	0.13%
6	HP Autonomy	6	0.27%	13	0.28%	1	0.02%
7	Asseco	3	2.13%	7	2.38%	6	2.04%
8	Swift	14	0.92%	133	3.62%	78	2.12%
9	Wincor Nixdorf	n/a	n/a	n/a	n/a	n/a	n/a
10	Misys	3	0.90%	34	1.78%	9	0.47%
11	Unit4	28	3.66%	120	8.63%	56	4.03%
12	Sopra Group	4	0.74%	30	2.26%	13	0.98%
13	Temenos Group	3	0.56%	21	1.47%	7	0.49%
14	Swisslog	2	0.14%	4	2.52%	3	1.89%
15	Micro Focus	7	0.61%	39	2.13%	12	0.66%
16	Compugroup Holding	n/a	n/a	n/a	n/a	n/a	n/a
17	Murex	n/a	n/a	n/a	n/a	n/a	n/a
18	Invensys	5	0.81%	38	1.38%	19	0.69%
19	NIS	n/a	n/a	n/a	n/a	n/a	n/a
20	IFS	17	5.72%	74	1.39%	27	0.51%
21	Acision	2	0.29%	7	0.71%	2	0.20%
22	Sophos Labs	16	0.13%	45	0.18%	29	0.12%
23	Fidessa	6	0.58%	18	0.82%	11	0.50%
24	Exact	26	1.47%	115	3.94%	34	1.17%
25	Cegid	8	2.13%	14	2.24%	9	1.44%
	Mean	65	1.20%	187	2.01%	116	0.99%

Table 8: Internal audience in the Twitter follower network

APPENDIX F

	acision	assecoesp	cegidpresse	dassault3ds	datev	exactsoftware	fidessa	hpautonomy	ifsworld	invensysopsmgmt	microfocus	misysfs	sageuk	SAP	softwareag	sophoslabs	soprarh	swiftcommunity	swissloginspire	Temenos	unit4_group
acision																					
assecoesp	0.00																				
cegidpresse	0.00	0.00																			
dassault3ds	0.08	0.00	0.11																		
datev	0.00	0.00	0.03	0.08																	
exactsoftware	0.04	0.00	0.09	0.10	0.09																
fidessa	0.29	0.00	0.00	0.02	0.02	0.04															
hpautonomy	0.24	0.04	0.04	0.20	0.00	0.08	0.31														
ifsworld	0.00	0.00	0.00	0.07	0.00	0.29	0.07	0.20													
invensysopsmgmt	0.15	0.00	0.00	0.06	0.00	0.04	0.06	0.18	0.11												
microfocus	0.38	0.08	0.07	0.15	0.06	0.14	0.32	0.45	0.21	0.34											
misysfs	0.20	0.00	0.28	0.00	0.00	0.00	1.89	0.16	0.00	0.00	0.27										
sageuk	0.06	0.01	0.20	0.08	0.23	0.41	0.08	0.40	0.11	0.08	0.26	0.03									
SAP	0.02	0.01	0.04	0.12	0.12	0.14	0.04	0.23	0.05	0.03	0.09	0.03	0.46								
softwareag	0.16	0.05	0.13	0.13	0.84	0.50	0.21	1.16	0.05	0.08	0.73	0.09	0.52	0.42							
sophoslabs	0.05	0.01	0.00	0.09	0.10	0.04	0.02	0.15	0.01	0.03	0.11	0.02	0.34	0.24	0.19						
soprarh	0.08	0.15	1.64	0.17	0.00	0.04	0.13	0.04	0.00	0.00	0.06	0.46	0.03	0.05	0.25	0.01					
swiftcommunity	0.00	0.00	0.21	0.03	0.08	0.09	4.11	0.13	0.00	0.05	0.30	3.74	0.18	0.09	0.33	0.04	0.15				
swissloginspire	0.00	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.02	0.03	0.01	0.00	0.00			
Temenos	0.08	0.00	0.00	0.10	0.05	0.04	1.39	0.15	0.00	0.09	0.30	6.20	0.06	0.07	0.33	0.02	0.56	2.53	0.05		
unit4_group	0.00	0.11	0.00	0.04	0.00	2.09	0.11	0.37	0.75	0.14	0.26	0.00	0.51	0.12	0.65	0.04	0.23	0.13	0.05	0.00	

Table 9: Jaccard index matrix table with percentual values (old situation)

	acision	assecoesp	cegidpresse	dassault3ds	datev	exactsoftware	fidessa	hpautonomy	ifsworld	invensysopsmgmt	microfocus	misysfs	sageuk	SAP	softwareag	sophoslabs	soprarh	swiftcommunity	swissloginspire	Temenos	unit4_group
acision																					
assecoesp	0.16																				
cegidpresse	0.06	0.11																			
dassault3ds	0.04	0.01	0.12																		
datev	0.02	0.02	0.04	0.04																	
exactsoftware	0.05	0.03	0.08	0.09	0.21																
fidessa	0.19	0.04	0.04	0.04	0.03	0.04															
hpautonomy	0.23	0.06	0.04	0.24	0.04	0.26	0.31														
ifsworld	0.09	0.02	0.07	0.13	0.11	0.32	0.12	0.23													
invensysopsmgmt	0.03	0.03	0.03	0.39	0.04	0.21	0.08	0.15	0.09												
microfocus	0.39	0.14	0.08	0.15	0.09	0.15	0.42	0.58	0.15	0.22											
misysfs	0.24	0.09	0.08	0.07	0.03	0.19	3.02	0.18	0.17	0.04	0.37										
sageuk	0.07	0.01	0.12	0.16	0.14	0.36	0.12	0.41	0.28	0.09	0.20	0.10									
SAP	0.04	0.01	0.04	0.32	0.14	0.18	0.05	0.35	0.22	0.10	0.11	0.07	0.55								
softwareag	0.20	0.05	0.09	0.22	1.00	0.92	0.45	1.13	0.44	0.21	1.05	0.65	0.51	0.60							
sophoslabs	0.06	0.02	0.00	0.11	0.12	0.05	0.05	0.23	0.10	0.06	0.14	0.04	0.32	0.39	0.25						
soprarh	0.09	0.31	1.99	0.23	0.02	0.05	0.14	0.03	0.05	0.02	0.16	0.15	0.08	0.07	0.26	0.02					
swiftcommunity	0.02	0.03	0.12	0.07	0.06	0.11	4.47	0.19	0.11	0.05	0.25	3.13	0.15	0.15	0.58	0.06	0.24				
swissloginspire	0.09	0.22	0.13	0.01	0.02	0.03	0.04	0.02	0.02	0.10	0.05	0.10	0.04	0.02	0.12	0.01	0.07	0.08			
Temenos	0.25	0.12	0.05	0.11	0.03	0.07	1.82	0.18	0.16	0.05	0.31	4.37	0.10	0.12	0.64	0.06	0.69	3.15	0.06		
unit4_group	0.08	0.12	0.20	0.09	0.08	2.67	0.14	0.38	0.70	0.12	0.43	0.15	0.43	0.17	0.81	0.06	0.22	0.12	0.06	0.18	

Table 10: Jaccard index matrix table with percentual values (new situation)

	acision	assecoesp	cegidpresse	dassault3ds	datev	exactsoftware	fidessa	hpautonomy	ifsworld	invensysopsmgmt	microfocus	misysfs	sageuk	SAP	softwareag	sophoslabs	soprarh	swiftcommunity	swissloginspire	Temenos	unit4_group
acision																					
assecoesp																					
cegidpresse																					
dassault3ds	-0.50		0.09																		
datev			0.33	-0.50																	
exactsoftware	0.25		-0.11	-0.10	1.33																
fidessa	-0.34			1.00	0.50	0.00															
hpautonomy	-0.04	0.50	0.00	0.20		2.25	0.00														
ifsworld				0.86		0.10	0.71	0.15													
invensysopsmgmt	-0.80			5.50		4.25	0.33	-0.17	-0.18												
microfocus	0.03	0.75	0.14	0.00	0.50	0.07	0.31	0.29	-0.29	-0.35											
misysfs	0.20		-0.71				0.60	0.13			0.37										
sageuk	0.17	0.00	-0.40	1.00	-0.39	-0.12	0.50	0.03	1.55	0.13	-0.23	2.33									
SAP	1.00	0.00	0.00	1.67	0.17	0.29	0.25	0.52	3.40	2.33	0.22	1.33	0.20								
softwareag	0.25	0.00	-0.31	0.69	0.19	0.84	1.14	-0.03	7.80	1.63	0.44	6.22	-0.02	0.43							
sophoslabs	0.20	1.00		0.22	0.20	0.25	1.50	0.53	9.00	1.00	0.27	1.00	-0.06	0.63	0.32						
soprarh	0.13	1.07	0.21	0.35		0.25	0.08	-0.25			1.67	-0.67	1.67	0.40	0.04	1.00					
swiftcommunity			-0.43	1.33	-0.25	0.22	0.09	0.46		0.00	-0.17	-0.16	-0.17	0.67	0.76	0.50	0.60				
swissloginspire					-0.93								-0.50	0.00	3.00	0.00					
Temenos	2.13			0.10	-0.40	0.75	0.31	0.20		-0.44	0.03	-0.30	0.67	0.71	0.94	2.00	0.23	0.25	0.20		
unit4_group		0.09		1.25		0.28	0.27	0.03	-0.07	-0.14	0.65		-0.16	0.42	0.25	0.50	-0.04	-0.08	0.20		

Table 11: Percentual difference (*100) in Jaccard index values

APPENDIX G

Rank	Company	Nodes (followers)	Followers that are competitors	Followers that are competitors %	Competitors
1	**SAP**	103096	29	0,03%	AcsisInc, aditroonline, AffectoAcademy, Affecto_Denmark, BetaSystems, CegedimRM, Comarch_IT, ElcaIT_fr, ExactTarget, hexagoninfosoft, MicroFocusDEE, microfocusITA, QlikTech_NAM, ReadSoft, READSOFTde, ReadSoftUS, ReadSoftZA, SageAccsSols, SageCRM, sagegroupplc, SageIT, softwareag, SoftwareAGUK, SoftwareAG_NA, sophosbanking, SopraBanking, SopraBelux, UNIT4_DCarbone, UNIT4_UK
2	**Dassault Systems**	14641	12	0,08%	avanquestuk_b2b, AVEVAGroup, aVg, cad2shop, CAD4MAC, cadaddict, cad_it, delcamartcam, ESIgroup, SAPAerospace, SAPAutomotive, SDL
3	**Sage**	21227	37	0,17%	acs100, acsacsltd, acsapt, ACSRecruitment, ACS_Building, CadairViewLodge, cadascumbria, CaddySCFC, cadenzaass, CADicksonCo, CADS_Sheff, CentricProject, EsInfoas, ExactIT, eXactOrder, GFISoftware, Hexagonfans, hexagonsoftware, IFSEC_Tristan, IRISSoftware, MamutSoftwareUK, MISysinccom, MISysWebGuy, sapcafe, SAPCrystalGo, SAPNorthAmerica, SapphireAccount, SapphireSigns, SapphireVilla, Sapphire_Venue, SAPSouthRegion, SDLNDTACADEMY, SDLsocial, swiftdavid1949, Swiftpage, SwiftTick, Swift_Therapy
4	**Software AG**	4061	19	0,47%	Invensys_Skelta, Kewill_Germany, Kofax, microfocusITA, Sage_Germany, SAPBPMMentor, SAPByDesign, SAPChnlPartnEna, SAPCloud, SAPEMEAPartners, SAPMillMining, SAPNorthAmerica, SAPPartnerEdge, SAPSocialOD, SAPTalentCloud, SAP_BPM, swiftcommunity, SWIFTSrvcBureau, SWIFT_Partners
5	**Datev**	4607	7	0,15%	BaswareGmbH, Lumesse_DE, SageHR, Sage_Germany, SAPde, SAPStore, SAP_PSD
6	**HP Autonomy**	4687	5	0,11%	GFI_Informatica, Kofax, sageuk, SAPAnalytics, sapoem
7	**Asseco**	294	2	0,68%	PandaComunica, SageSpain
8	**Swift**	3674	11	0,30%	Linedata, misysbanking, misysfs, misys_europe, SAGE_Prospero, sapcrm, SAPforBanking, SmartStream_STP, SopraBanking, Temenos, TemenosCareers
9	**Wincor Nixdorf**	n/a	n/a	n/a	
10	**Misys**	1914	6	0,31%	Kofax, sophosbanking, SopraBanking, SWIFTSrvcBureau, SWIFT_Partners, Temenos
11	**Unit4**	1390	2	0,14%	hexagonsoftware, SageDespProf
12	**Sopra Group**	1329	5	0,38%	CegidPublic, CegidSIRH, gfiinformatique, SageFrance, SAPCrystalGo,
13	**Temenos Group**	1430	7	0,49%	fidessa, misysbanking, misysfs, SopraBanking, swiftcommunity, SWIFTSrvcBureau, SWIFT_Partners
14	**Swisslog**	159	0	0,00%	
15	**Micro Focus**	1830	0	0,00%	

16	**Compugroup Holding**	n/a	n/a	n/a	
17	**Murex**	n/a	n/a	n/a	
18	**Invensys**	2751	1	0,04%	SAPartnersUSA
19	**NIS**	n/a	n/a	n/a	
20	**IFS**	5328	4	0,08%	Centric_Wimvb, sapexplore, SapphireLoungeX, UNIT4_UK
21	**Acision**	989	2	0,20%	sapoem, SAP_Telco
22	**Sophos Labs**	24460	14	0,06%	AVGFree, GFISoftware, pandacilla, PandaSecurityCA, PandaSecurityFR, pandasecuritytr, PandaSecurityUK, PandaSecurityUS, pandasuisse, Panda_Arg, Panda_Security, sage_guru, sapphiredotnet, SAPPublicSector
23	**Fidessa**	2188	4	0,18%	Kofax, SAPforBanking, SimCorp, Temenos
24	**Exact**	2918	2	0,07%	UNIT4_Multivers, VismaHRMnieuws
25	**Cegid**	626	3	0,48%	Comarch_IT, gfiinformatique, sage_DrouotS
	Mean	6085	8.19	0.21%	AcsisInc, aditroonline, AffectoAcademy, Affecto_Denmark, BetaSystems, CegedimRM, Comarch_IT, ElcaIT_fr, ExactTarget, hexagoninfosoft, MicroFocusDEE, microfocusITA, QlikTech_NAM, ReadSoft, READSOFTde, ReadSoftUS, ReadSoftZA, SageAccsSols, SageCRM, sagegroupplc, SageIT, softwareag, SoftwareAGUK, SoftwareAG_NA, sophosbanking, SopraBanking, SopraBelux, UNIT4_DCarbone, UNIT4_UK

Table 12: Competitors in the Twitter follower network

APPENDIX H

Account	Frequency	Domain	Description based on bio of Twitter use
MarqitNL	9	News	News platform for IT professional
MartrainLtd	7	Company	Tech marketing agency
Transacting	6	Mixed	Banking technology professional; follow me for news in finance, tips in technology and the latest in best practice in banking systems
rainmakerfiles	6	Company	Steer technology rainmakers towards high growth opportunities
Persistentcom	6	Company	Specialize in the acquisition and development of Premium Domain Names
Johnpeterking	6	Individual	
ITDF	6	News	IT Directors Forum
IBSIntelligence	6	Company	Independent research, news and analysis of financial technology and core banking systems
econique_group	6	Company	CxO relations and offer communication channels for chief executives with high level networking and cutting-edge conferences
CarlatStar	6	Individual	Business Development Manager working for Star
AWVance	6	Individual	B2B IT Lead Generation Campaign Manager for Enterprise IT Vendors

Table 13: Most frequent followers (old situation)

Account	Frequency	Domain	Description based on bio of Twitter use
twit_resrc	20	Individual	
Tsics	12	Individual	Enterprise tech voice in Libya: banking & finance and utilities. Board member of Canadian Academy of Libya. Better education and healthcare a priority!
Hootsuite	8	Mixed	Updates about the social media management tool which helps teams to securely engage audiences & measure results. See also: @HootSuite_Help @HootWatch& more.
econique_GmbH	8	Company	High Level Networking
ictberichten	8	News	Actuele persberichten van IT-, Telecom-, Internet- en Officebedrijven.
MarqitNL	8	News	Nieuws | Whitepapers | Leveranciers | Evenementen | en veel meer over de IT markt in Nederland.
Dlainsa	7	Individual	Economics, Finance and Technology
AWVance	7	Individual	Campaigns Manager at Elation Sales - B2B Lead Generation Campaigns, Sales Strategy & Sales Training.
WinnTechnology	7	Company	Global, integrated #B2B #marketing solutions for the #technology industry. Specializing in #DemandGen, #Channel Marketing, #Event Promo, #Data and #Social.
MartrainLtd	7	Company	B2B Tech Marketing agency Martrain Ltd. Delivering IT Lead Generation Campaigns for global IT Vendors like SAP & VMware.
Iteuropa	7	News	Respected pan-European #newsletter on Europe's IT #channels, #resellers, #integrators, #vendors with #ITnews, #research, #data for sale http://t.co/8CAy4QhE

IBSIntelligence	6	News	Independent research, news and analysis of financial technology and core banking systems.
hk_andre	6	Individual	Executive in IT with many years experience in both North America and Asia. Really loves to share information.
XLR8Marketing	6	Company	XLR8 Marketing provides joined up marketing services for its technology sector clients - accelerating business performance and delivering measurable ROI.
rwang0	6	Individual	Constellation Research Analyst. Provocateur, Keynote Speaker, Disruptive Tech, Innovation, Author, Strategist, Contract Negotiator. Chairman & Founder. Club DJ.
CruiseLineFocus	6	Company	Our mission is to help make distinctive improvements in Product Sales, Delivery Support & Industry Consulting @DouglasDiggle @AcrossOceansGrp @CruiseLineSOLAS
jobone4u	6	Company	BUSINESS CONSULTING & HR RECRUITMENT, ADVISORY SERVICES FOR CORPORATES & HNI Recruitment/business consulting/Reality /executive search @one place
Ultraerp	6	Company	Independent consulting firm; ERP selection & implementation, business intelligence, business process improvement. #ERP #CIO #IT #EnSw #BI #manufacturing
EnterApps	6	Company	Discover how today’s IT mega trends of Mobile, Social, Big Data and Cloud have transformed the enterprises business software landscape
CXP	6	Company	L'actu du CXP, cabinet européen indépendant d’analyse et de conseil, spécialisé dans les solutions logicielles
Pjtec	6	Individual	Following the enterprise apps market for over a decade. APICS buff, former Baan ERP expert. TEC veteran & principal analyst
TFConsult	6	Individual	Premier #Consulting: #BusinessDevelopment, #Innovation #Strategy, #LifeCycleManagement, #Transformation & #Change, #Sportive Passion 4 #HighPerformanceTeams
JonahLupton	6	Individual	@Nutraspire @LuptonMedia @GreatestPitch @Strivved @bStrongClothing @FullForceLabs @Cauzly @ParabolicVC @iCapGroup @TruRides @LuptonGroup http://t.co/WwJlCnvti8

Table 14: Most frequent followers (new situation)

APPENDIX I

		PUSEU
ANTPD	Pearson correlation (r)	0.462
	Statistical significance (p)	0.062
	Sample size (N)	17
Legend		
PUSEU	% unfollowed and still existing users Refers to the percentual amount of existing users that have unfollowed a Twitter account.	
ANTPD	Average number of tweets per day Refers to the average number of tweets (including retweets and replies) a Twitter account has shared per day.	

Table 15: Results of the Pearson's correlation test to determine the relation between the 17 Twitter account's PUSEU and ANTPD values

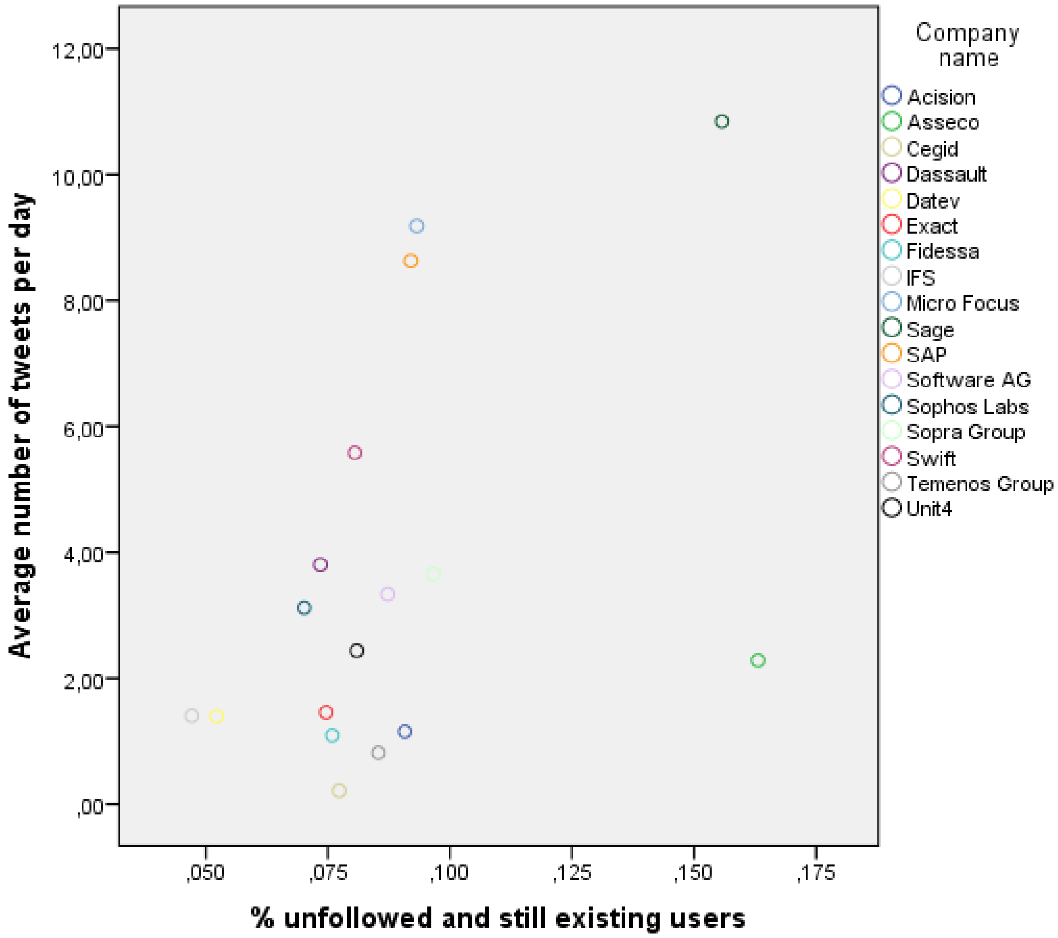

Figure 2: Scatterplot of the association between PUSEU and ANTPD

		ANHT
PFS	Pearson correlation (r)	0.462
	Statistical significance (p)	0.006
	Sample size (N)	17
Legend		
PFS	% difference in follower size Refers to the percentual difference in the follower size of a Twitter account in comparison with the follower size in the previous study.	
ANHT	Average number of hashtags per tweet Refers to the average number of hashtags a Twitter account has used per tweet.	

Table 16: Results of the Pearson's correlation test to determine the relation between the 17 Twitter account's PFS and ANHT values

		PRA	PRD	TNNF
PRA	Pearson correlation (r)	1	0.993**	0.287
	Statistical significance (p)		0.000	0.264
	Sample size (N)	17	17	17
PRD	Pearson correlation (r)	0.993**	1	0.228
	Statistical significance (p)	0.000		0.378
	Sample size (N)	17	17	17
**. Correlation is significant at the 0.01 level (2-tailed).				
Legend				
PRA	% difference in reciprocity (arc) Refers to the percentual difference in the reciprocity (arc method) in comparison with the value from the previous study.			
PRD	% difference in reciprocity (dyad) Refers to the percentual difference in the reciprocity (dyad method) in comparison with the value from the previous study.			
TNNF	Total number of new followers Refers to the total number of new followers that have joined the network since the previous data collection.			

Table 17: Results of the Pearson's correlation test to determine the relation between the 17 Twitter account's TNNF and PRA, and PRD values

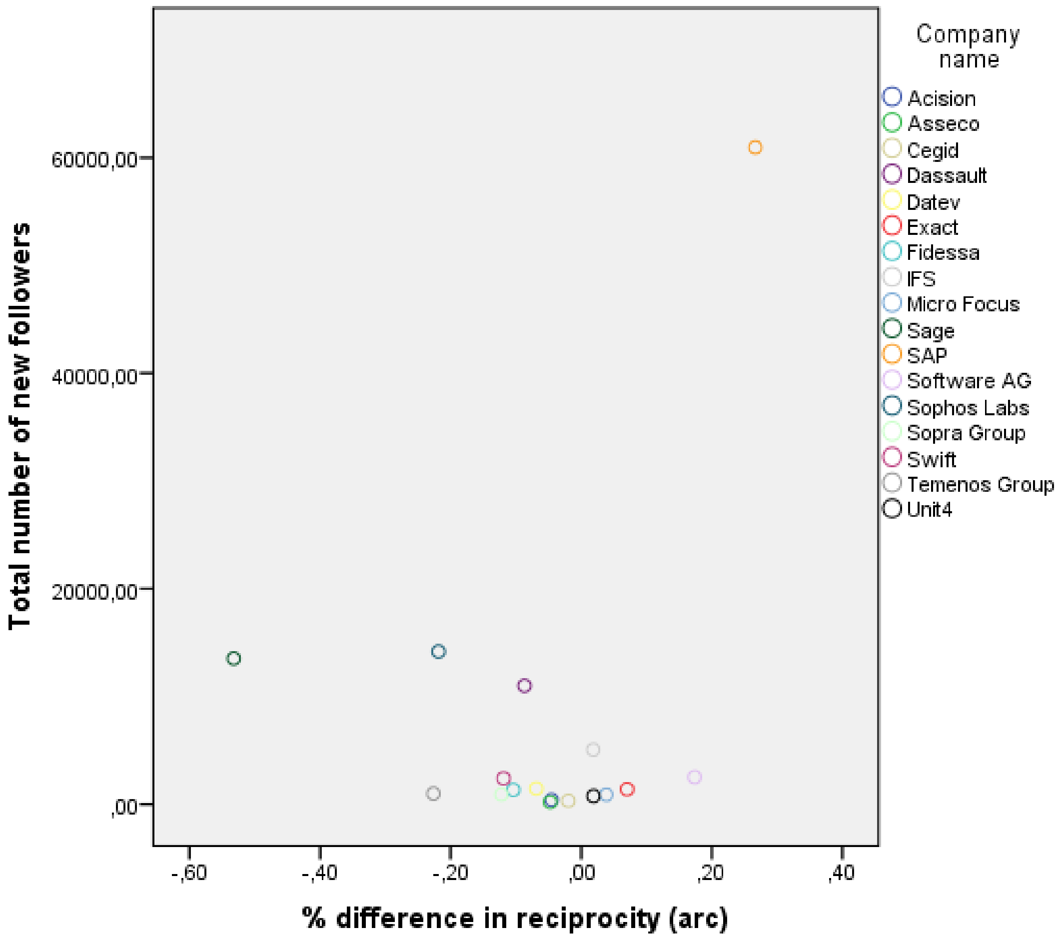

Figure 3: Scatterplot of association between the PRA and TNNF