

Version 0.9
(Review version)

Influencing & measuring word of mouth on Twitter



INFLUENCING & MEASURING WORD OF MOUTH ON TWITTER

What Twitter strategies are most effective in influencing word of mouth diffusion on Twitter, and how should influence be measured?



Master Thesis Paper
April, 2011

Author:

Danny Oosterveer

Student Master of Business Administration - *Specialization Marketing*

Telephone: 06-36174838

Email: contact@dannyoosterveer.nl

Website: www.dannyoosterveer.nl

Twitter: [@DannyOosterveer](https://twitter.com/DannyOosterveer)

Principal:

Nijmegen School of Management

Radboud University Nijmegen

Thomas van Aquinostraat 3

6500 HC Nijmegen

The Netherlands

Primary supervisor:

Dr. Ir. N.G. Migchels

Secondary supervisor:

Dr. M.J.H. van Birgelen

Cover:

People connected in a network. Some people are more strongly connected than others. The people highlighted in yellow are spreading the word of mouth to the other people within the network.

Keywords:

WOM, word of mouth, eWOM, electronic word of mouth, viral marketing, buzz marketing, influence, word of mouth marketing, Twitter, microblogging, connected marketing, Web 2.0, retweets, retweet indegree, brand mentions, mentions indegree, brand sentiment analysis, opinion mining, text mining, number of followers, followers indegree, hashtag, communities, customer-relationship management, brand reputation management, customer service management, one to one communication, Radian6, social media, user-generated content, UGC, network theory, dissemination, communication, consumer to consumer, consumer to business, S-D logic of marketing, Tweetfeel Biz, engagement, big seed marketing, collaborative filtering, diffusion, influentials

ABSTRACT

During the last decade, the way consumers communicate has significantly changed. This change is facilitated by the World Wide Web as a platform whereby information is no longer produced by a small group of institutions. Instead, a rising number of consumers use the Web to express and disseminate their knowledge, experiences, and opinions about products and services. The transition from traditional broadcasting to "Web 2.0" has greatly expanded the opportunities for brands to use bidirectional communication.

Using over 250,000 tweets produced by brands and consumers during a 10 week research period, the effect of strategies as suggested by professional literature on a brand's influence on consumer tweets was investigated. As a social medium, Twitter is one of the 2.0 platforms which gained enormous popularity over the last years.

In this paper, empirical study is presented which is unique in its nature as it investigates the relationship between brands' Twitter strategies as condition, and its influence on consumer tweets. The study has shown that one to one communication, listening to consumers and community participation all significantly influence consumer word of mouth. One to one communication shows the largest effect across influence measures. This highlights the evidence of consumer engagement for brands.

The effect of community participation is also significant as well as listening to consumers, although its effect on retweet indegree and sentiment is negative. The amount of following has the greatest effect on followers indegree. This shows that relationships on Twitter may be more reciprocal than how relationships on Twitter are generally presented. Moreover, these results support the notion of the "million follower fallacy" which assumes people make use of etiquette to elevate their followers indegree.

Furthermore, various measures of influence were evaluated: mentions indegree, followers indegree, sentiment and retweet indegree. The study only partially supports the notion of the "million follower fallacy", since followers indegree shows strong correlations with the other measures of influence, while the theory suggests reciprocity distorts the relationship between followers and measures of influence.

The study's findings provide new insights for managers when developing a strategy for Twitter or other social media, aimed at increasing social influence. Furthermore, the results suggest concepts which can be used to measure a brand's influence on consumer's online word of mouth.

PREFACE

The interest for writing this paper started as a quest for measuring the spread of word of mouth on Twitter. The viral aspect of Twitter has gained much attention as one of the features which makes it the popular social medium it is today. Investigating *virality* turned out to be problematic, as one requires a relational multi-level dataset. Instead, the paper increasingly focused on predicting influence over word of mouth networks.

From both the professional and academic field yield a strong desire to understand and measure influence. Where it used to be problematic to measure the influence a brand has on consumers, the open ecosystem of Twitter allows for investigation of the word of mouth between consumers. Inspired by the challenge, and considering the gap in literature and its practical relevance, I chose to investigate prediction and measurement of influence over word of mouth networks. The paper thereby anticipates on the transition between traditional (sender-receiver) communication to the web 2.0 media landscape in which consumers increasingly disseminate their knowledge, experiences, and opinions with fellow consumers. While the paper focuses on the social medium Twitter, its findings may be applied to other social media.

Sources of inspiration for writing the final topic of the paper include the work of [Brian Solis](#). His book on the convergence between traditional and social media, [Engage](#), is a recommendation if it is your interest to better understand the deeply rooted forces behind this development. Other than it were the fascination for Twitter as a medium and the technology that makes it possible to investigate the massive amount of conversations in real time that attracted me to the subject.

I'm looking forward to the future work of professors [dr. J.M.M. Bloemer](#) and [dr. M.J.H. van Birgelen](#) who have shown interest in the work and data of this paper. For me their interest in this paper is both an acknowledgement and an honor for the energy and time I have invested in this study.

April 27, 2011 – Danny Oosterveer

ACKNOWLEDGEMENTS

I am grateful to many people for help, both direct and indirect, in writing this paper. I would like to acknowledge a few who have assisted me exceptionally. First, I would like to express my gratitude towards [Ludo Raedts](#). As manager of [The Webcare Company](#), Ludo has provided me with an account to the social media monitoring tool [Radian6](#). It is not that common that thesis papers are sponsored by products as great as Radian6. Without the account, I would neither have been able to get this rich and complete data, nor would I have been able to perform a sentiment analysis.

I have also received major help in handling the data. Without [Serhat Gülçiçek](#)'s efforts, it would have been very difficult if not impossible for me to execute this research. Not only has Serhat, Software Engineer at Logica, written the application that gathered the Twitter data in the pretest, he also wrote an extensive application which has transformed the dynamic data from Radian6 into a reliable datasheet, compliant for analysis with IBM SPSS. As a personal friend I would like to thank him from the bottom of my heart for all the efforts he put in.

Last but not least I want acknowledge thank the professors from the Radboud University. I want to thank my primary supervisor [Nanne Migchels](#) for guiding my all the way through the research process. Furthermore I would like to express my gratitude to methodology professors [Jörg Henseler](#) and [Paul Ligthart](#). Methodology has been a great challenge in writing this paper for the complexity of the data analysis strategy; therefore it was very helpful to receive assistance from methodology experts.

<Insert co-readers here>

Version 0.9
(Review version)

Table of contents

Abstract.....	i
Preface	ii
Acknowledgements	iii
Chapter 1: Introduction	1
1.1 Introduction.....	1
1.2 Research question	2
1.3 Theoretical relevance.....	3
1.4 Practical relevance.....	4
1.5 Structure of the report.....	4
Chapter 2: Conceptual framework	5
2.1 Impact of contemporary communication	5
2.2 Social media, Web 2.0 & User-generated content.....	7
2.3 Microblogging.....	7
2.4 Twitter.....	8
2.5 Twitter strategies.....	10
2.5.1 Listening to customers.....	10
2.5.2 One-to-one communication.....	10
2.5.3 Community participation.....	11
2.6 Influence.....	12
2.7 Word of mouth (WOM).....	13
2.8 Electronic word of mouth (eWOM)	13
2.9 Measuring influence	14
2.9.1 Followers indegree	14
2.9.2 Retweet indegree	15
2.9.3 Mentions indegree	15
2.9.4 sentiment Analysis.....	15
2.10 Conceptual model.....	16
Chapter 3: Methodology	17
3.1 Choice of research method	17
3.2 Research sample.....	17
3.3 Data processing	18
3.3.1 Data collection & Hardware setup	18
3.3.2 What data is gathered	18
3.4 Processing the data.....	19
3.5 Validity	20
3.6 Reliability.....	20
Chapter 4: Results.....	22

4.1 Assumptions	22
4.2 Analysis.....	22
Chapter 5: Conclusions	25
5.1 Conclusions.....	25
5.2 Limitations	25
5.3 Directions for future research	26
5.4 Implications	26
References.....	28
Appendix	32
Appendix 1: Amount of tweets per day, by Twitter Inc.	32
Appendix 2: Twitter downtime per month, by Pingdom	33
Appendix 3: Data alteration within SQLyog.....	34
Appendix 4: Tweet distribution	47
Appendix 5: Controlling activity gaps.....	48
Appendix 6: Data gathering process in the pretest	49
Appendix 7: Cell data contained by a single record.....	50
Appendix 8: Reliability analysis of the sentiment analysis.....	52

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

During the last decade, the way consumers communicate has significantly changed. Rather than being merely an audience, consumers increasingly interact with each other. This change in communication is facilitated by the new way in which internet is used. This is called Web 2.0, a term that is used to describe the new way in which consumers and developers started to utilize the World Wide Web; that is, as a platform whereby content and applications are no longer published by a small group of institutions, while the end-users are merely content consumers. Rather, content is a product of continuous active collaboration, participation and interaction of consumers. Web 2.0 succeeds the era of Web 1.0, more familiar as the dotcom era, which refers to the first period (approximately 1990-2001s) of the World Wide Web in which organizations rehashed old marketing strategies, turning the Web into another broadcast medium. But it became obvious that the internet was a participatory medium (Van Veelen et al, 2008). In this era applications such as Web pages and the idea of content publishing were already around, until the idea emerged that internet was a participatory medium. The rise of social media such as blogs, wikis and social network sites facilitated the transition from one-to-many to many-to-many communication. Consumers increasingly use the Web to express and disseminate knowledge, experiences and opinions. This intrinsic behavior can be reasoned back to the work of Freud's psycho-analysis in which he describes the identity of the self. In essence people want to be treated as individuals, whether it is by other people, governments or companies. In the large scaled world, this is practically impossible, unless companies know the individual, unless the individual reveals itself to them (Kelly, 2011). This makes transparency the cost for personalization, making privacy less crucial.

And this takes place massively. Anno 2010 the Web counts 155 million blogs¹ as well as 17.6 million articles on Wikipedia². The largest social network site, Facebook, has over 600 billion members³ and they share 30 billion pieces of content a month⁴. 35 hours of video is uploaded to YouTube a minute, whereas 2 billion videos are being viewed a day⁵. Photo share site Flickr hosts over 5 billion photos, and every month another 130 million are uploaded⁶. Social media caused an enormous shift in Web usage; the top 20 of globally most visited English Websites now contains 9 social media, namely Facebook (2nd), YouTube (3rd), Wikipedia (6th), Blogger (7th), Twitter (9th), WordPress (11th), MySpace (15th), LinkedIn (16th) and Flickr (19th)⁷.

Organizations need to recognize the changing communication landscape. Mass marketing perspectives in which the organization was merely sender and the consumer merely receiver need to be deprecated. Organizations need to change their marketing and communication model in order to adapt to the changing environment. The S-D logic of marketing re-defines marketing as a process whereby there is a constant stream of communication between the firm and the customers to improve the quality of the value offer (Vargo & Lusch, 2004). Social media are particularly valid for organizations to connect with potential customers, but their great marketing potential has not been discovered by them, yet.

As one of the forms of social media, microblogging especially gained enormous popularity during recent years, mainly driven by the most popular service, Twitter. Within a year of its launch in October 2006, Twitter already hosted about 5.000 tweets (microblog postings) per day. As of

¹ BlogPulse, 2011. Retrieved from <http://www.blogpulse.com/>

² Wikipedia, 2011. Retrieved from <http://stats.wikimedia.org/EN/TablesWikipediaZZ.htm>

³ Business Insider, 2011. Retrieved from <http://www.businessinsider.com/facebook-has-more-than-600-mil...>

⁴ Facebook Inc., 2011. Retrieved from <http://www.facebook.com/press/info.php?statistics>

⁵ Google, 2010. Retrieved from <http://youtube-global.blogspot.com/2010/11/great-scott-over-35-hours-of-vi...>

⁶ Yahoo! Inc., 2010. Retrieved from <http://blog.flickr.net/en/2010/09/19/5000000000/>

⁷ Alexa Rankings: Top 500 Global Sites website. Retrieved November 2010 from <http://www.alexa.com/>

September 2010 Twitter hosts over 95.000.000 tweets per day⁹, that's an average of 1100 tweets per second. Other microblogging services include Jaiku, Pownce, Tumblr, Plurk and Yammer. Microblogs are short comments usually delivered to the poster's network of associates.

Apart from pull factors (i.e. the changing communication landscape), there are also factors pushing the use of social media, namely the diminishing effect of traditional media. Commercials and advertisements are everywhere, more than a human can consume (Simon, 1971). The overwhelming presence of advertising leads to information overload for consumers. The paradigm which focuses on the struggle organizations deal with in the constant battle for the consumer's attention is called the "attention economy". The attention economy considers attention a scarce resource (Davenport & Beck, 2002). As the effect of traditional push media diminishes, organizations look for new forms of ways to communicate with their target group.

The notion of influence plays a vital role in how businesses operate and how a society functions, for instance, how fashion spreads (Gladwell, 2002) and how people vote (Keller & Berry, 2003). A large body of research is dedicated to identifying antecedents of influence, such as source expertise, tie strength, demographic similarity and perceptual affinity. A modern view called collaborative filtering, de-emphasizes the role of traditional influentials and argues people in the new information age make choices based on the opinions of their peers (Domingos & Richardson, 2001). Research simulation has shown that influentials did not initiate all diffusions, moreover, in homogeneous networks, influentials were no more successful in initiating long cascades than ordinary users (Watts & Dodds, 2007; Watts et al., 2007). It is assumed that electronic, peer-to-peer communications are an effective means to transform (electronic) communication networks into influence networks, capturing recipients' attention, triggering interest, and eventually leading to adoption or sales (De Bruyn & Lilien, 2008).

Word of mouth diffusion is regarded as an important mechanism by which information can reach large populations, thereby influencing public opinion (Katz & Lazarsfeld, 1955) and consumer decisions for purchasing new products (e.g. Engel et al., 1969; Arndt, 1967; Richins, 1983; Richins & Root-Shaffer, 1988; Whyte, 1954; Bansal & Voyer, 2000). Word of mouth diffusion may be useful for marketers when they make use of consumers' network to spread the marketers' message. The function of WOM communication is based on social networking and trust: people rely on other people within their social network. WOM communication strategies are appealing because they combine the prospect of overcoming consumer resistance against traditional forms of communication, with significantly lower costs and fast delivery (Trusov et al., 2009). Positive WOM communication is considered a powerful marketing tool for influencing consumers.

eWOM is different from WOM in the sense that it occurs in an electronic or online environment. Moreover eWOM is many times anonymously or confidentially, as well as to provide geographical and temporal freedom. Similar to WOM, eWOM has been shown to significantly influence consumer decisions for purchasing new products (e.g. Dellarocas et al., 2004; Chevalier & Mayzlin, 2004). Moreover, research has shown that eWOM may have higher credibility, empathy and relevance to customers than marketer created sources of information on the Web (Bickart & Schindler, 2001). Research also indicates that people appear to trust seemingly disinterested opinions from people outside their immediate social network, such as online reviews (Duana, Gub & Whinston, 2008). eWOM may be less personal in that it is not face-to-face (or maybe just personal in a different way than in the past), but it is more powerful because it is immediate, has a significant reach, is credible by being in print, and is accessible by others (Hennig-Thurau et al., 2004).

1.2 RESEARCH QUESTION

As the communication landscape is changing, consumers increasingly act with each other via social media. For organizations it's no longer a choice whether or to put their brand online. Consumers are

⁹ Twitter Inc, 2011. Retrieved from <http://twitter.com/about>

already praising, discussing or criticizing the brand via social media, outside of the control of the organization. It is merely how the organization chooses to act upon these online discussions.

Moreover, social media in general and microblogging in particular still is a rather new phenomenon for many marketers. Professional, non-scientific management literature provides guidelines for marketers for how they should use Twitter strategically in order to gain influence.

There have been no scientific studies which have confirmed these strategies. Further insights in how organizations should utilize Twitter are desired. Moreover, organizations are facing difficulties when measuring the results of their social media efforts. Organizations require more knowledge about how to measure their online influence. The current study evaluates several concepts for measuring influence on Twitter.

The research question to be answered in this study is:

“ What Twitter strategies are most effective in influencing word of mouth diffusion on Twitter, and how should influence be measured? ”

1.3 THEORETICAL RELEVANCE

Though microblogging is a very powerful online social medium, it is relatively unexplored in marketing literature compared to other online social phenomena such as social network sites (e.g. Facebook, LinkedIn), online review Websites (e.g. reviews on Amazon) and online communities (e.g. Flickr, YouTube). Prior researches which focused on microblogging were mostly descriptive in the sense they investigated the nature of tweets, sentiment (positive/negative loading), tweet types (Comments/Sentiment/Information seeking/Information providing), the type of interaction (digital exhibitionism/voyeurism or information seeking/sharing), and word usage (Jansen et al., 2009; Costa et al., 2008; Java et al., 2007). Others have checked its validity for online discussion and collaboration (Honeycutt & Herring, 2009; Ebner & Schiefner, 2008).

Influence has long been studied in the fields of sociology, communication, marketing, and political science (Rogers 1962; Katz & Lazarsfeld 1955). The notion of influence plays a vital role in how businesses operate and how a society functions, for instance, see observations on how fashion spreads (Gladwell, 2002) and how people vote (Keller & Berry, 2003). A large body of research is dedicated to identifying sources of influence as antecedents of WOM influence, such as source expertise, tie strength, demographic similarity and perceptual affinity. Studying influence patterns, however, has been difficult. Huberman et al. (2008) investigated the quality of the links between people active on Twitter, showing what connections are actually valuable. Huberman et al. (2008) separated friends from followers, and their value difference. Nardi et al. (2004) investigated what caused people to express themselves online. Previous studies have quantified influence in terms of network metrics (e.g. Page Rank). More recently studies used more alike WOM measures, such as the size of the entire diffusion tree associated to quantify influence. Diffusion trees have been studied for gesture exchanges within Second Life (Bakshy et al., 2009), adoption of a mobile phone application over the Yahoo! messenger network (Aral et al., 2009), Fan Pages on Facebook (Sun et al., 2009) and identifying leaders for Yahoo! Movie data (Goyal et al., 2010).

Most closely related to the current research is a series of recent papers which were published only during the execution of this paper, which that examined influence through eWOM on Twitter specifically. Kwak et al. (2010) compared three different measures of influence: the number of followers, Page-Rank, and number of retweets. Cha et al. (2010) also compared three different measures of influence: the number of followers, the number of retweets, and the number of mentions. Finally, Weng et al. (2010) compared number of followers with a modified Page Rank measure that accounted for topic.

This study aims to fill three gaps in research in this field. Since its novelty, not many studies have examined influence via word of mouth on social media, and Twitter in particular. Whereas influence used to be difficult to quantify, the open system of social media, but that of Twitter in particular, now allow for investigation of the flow of conversations among consumers. The huge amount of data available on social media is like never witnessed before. As put by Nicholas Christakis, a Harvard sociology professor: "We're on the cusp of a new way of doing social science. Our predecessors could only dream of the kind of data we now have." (Rosenbloom, 2007).

Second, whereas previous studies have been largely of descriptive nature, the current study predicts brands' influence by considering the brands' Twitter strategy as predictor. Professional management literature suggests a variety of strategies marketers could use to gain influence on Twitter. There is limited scientific confirmation that these strategies actually have the proposed effect.

Whereas previous studies have examined user influence, the current study is unique in its nature since it investigates electronics word of mouth which relates back to the brands' actions on Twitter.

1.4 PRACTICAL RELEVANCE

Web 2.0 has changed the way in which consumers communicate for good. It allows consumers to participate, share and collaborate on the World Wide Web. Rather than being mere recipients of the information that is disseminated by marketers, a rising number of consumers use the Web 2.0 to express and disseminate their knowledge, experiences, and opinions about products and services. The transition from traditional word of mouth networks to digital networks has greatly expanded the opportunities for bidirectional communication (Dellarocas, 2003). Consequently, electronic word of mouth has become a significant market force that influences consumer decision-making. Marketers have to deal with consumers who increasingly interact with each other through social media.

Social medium Twitter offers organizations a platform to communicate with customers in almost real time. Professional literature suggests Twitter can be effectively used to create ties with consumers, to increase loyalty and commitment. Several strategies are suggested which should achieve a greater brand influence. For marketers it would be very interesting to know how effective these strategies are. By means of these findings it is possible to draw conclusions and managerial implications considering how eWOM should be managed. Although other studies which were released during this study measured influence, there haven't been any studies investigating the effect of certain strategies on increasing influence.

Also, the results of this research provide an evaluation of various performance measures of influence through eWOM. By means of these findings it is possible to draw conclusions and managerial implications considering how eWOM should be measured.

1.5 STRUCTURE OF THE REPORT

Chapter 2 contains the theoretical framework of the study. Chapter 3 elaborates on the methodology and the processing of the data which is used in the study. Chapter 4 contains the analysis of the findings. Chapter 5 contains the conclusions and discussion of the findings.

CHAPTER 2: CONCEPTUAL FRAMEWORK

2.1 IMPACT OF CONTEMPORARY COMMUNICATION

For traditional media, communication corresponds to the mass marketing or broadcasting principle. This model is often referred to as the transmission model or standard view of communication. In this model the information is sent from the organization (the sender) to the consumer (the receiver). This common conception of communication views communication as a means of sending and receiving information. The strengths of this model are simplicity, linearity, generality, and quantifiability.

This form of communication was modeled by Shannon & Weaver (1949). Their transmission model is usually described along a few major dimensions:

- An information sender, which produces a message.
- The encoding of the message into signals
- A channel, to which signals are adapted for transmission
- The decoding of the message, which interprets the message from the signal.
- An information receiver.

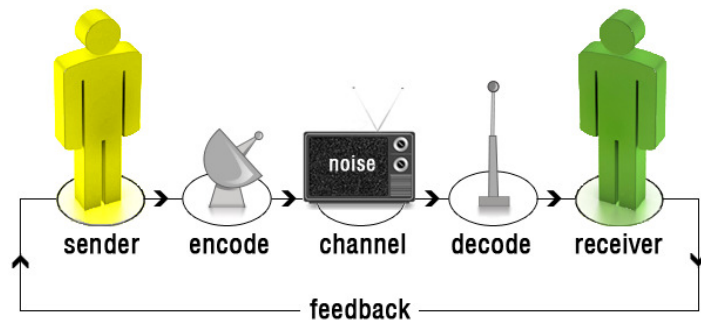


Figure 1: Transmission Model of Communication (Shannon & Weaver, 1949)

The encoded message reaches recipients, through advertising or salespeople (channel), who then decode and absorb the information either fully or partially. The quality of the transmission can be distorted by ‘noise’ occurring because the receiver does not interpret the message in the way the source intended, due to e.g. cultural background or cognitive dissonance.

This transmission model of communication is still common within many organizations. It may be applied to various channels, e.g. advertising, sponsoring or personal sales. It is expected that once the information is processed by the target audience, it will translate into certain communication goals, such as brand awareness, brand knowledge, brand attitudes, behavior intention or return on investment (ROI). The organization’s job is to optimize the message to be communicated, in order to minimize the noise. Minimizing noise reduces dissonance between sender and receiver, thereby optimizing the results.

The transmission model is also incorporated in the Cluetrain Manifesto Model (Derksen, 2011). The model is based on the Cluetrain Manifesto Theses (Locke et al., 2000). Basically it displays marketing's paradigm shift in its transition from the 1.0 to the 3.0 media landscape. The Cluetrain Manifesto regards the Web as a novel set of media where old rules of marketing communication and information exchange do not apply. In favor, it advocates the concept of conversational exchanges. In creating conversational knowledge, individuals and institutes create and share knowledge through open dialog, rather than one-dimensional monologues.

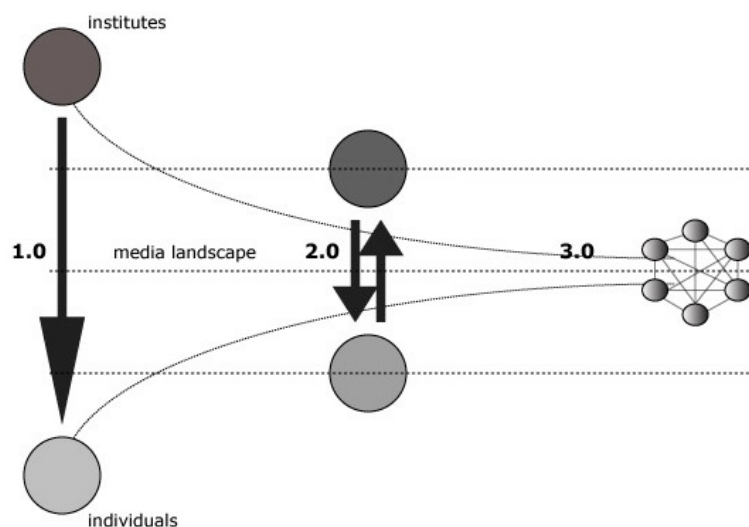


Figure 2: The Cluetrain Manifesto Revival Model

The Cluetrain Manifesto Revival Model displays the changing roles of institutes and individuals within the media landscape. In the 1.0 era institutes broadcast their message via traditional media to the individuals. Here the power relies with the institutes. In era 2.0, individuals increasingly gain possibilities to share their own message with their environment. Rather than being mere recipient, individuals engage in dialogue, increasing the influence of the individuals (“Power to the crowd”). In the 3.0 era, the hierarchical relation between institutes and individuals disappears and is superseded by cooperation within networks. Social media and other technological developments support increasingly intelligent and smoother cooperation.

Elaborating further on the 2.0 paradigm of communication, technologies 2.0 empower consumers to interact with one another, making them active participants in the communication process rather than being merely an audience. In other words, the sender and receiver are no longer static. An increasing number of consumers uses the Web to express and disseminate their knowledge, experiences, and opinions, thereby influencing fellow consumers (Constantinides & Fountain, 2007; Mangold & Faulds, 2009). Marketers have to deal with consumers who increasingly interact with each other via social media, with its latest being microblogging.

Organizations should let go of their traditional media mindset in which they assume full control of the communication process. The content, timing and frequency of the conversations among consumers via social media are outside the organization’s direct control. Consumers’ ability to communicate with one another limits the amount of control companies have over the content and dissemination of information (Mangold & Faulds, 2009). Organizations have to adapt and identify their role in the changing communication environment. An often initial start is listening to customers. A subsequent step may involve communicating or engaging with consumers. Due to the Web’s properties, social media require organizations to change their way of communicating. When a brand participates as sender in the social media process, it should make sure that its message has value (is interesting, unique, exclusive, etc) for the consumers, so that consumers actually want to listen. They might even replicate it to their network. This is significantly different from the media landscape 1.0 where a brand would simply broadcast their message to a large audience. In relation to social media, traditional media have some disadvantages: they are usually expensive, characterized by a high percentage of waste and the level of consumer involvement is low. Social media in contrast are usually inexpensive, are characterized by a low percentage of waste (recipients voluntarily receive the message) and a high level of consumer involvement. In order to influence customers, brands should engage with their customers (e.g. start a blog, join social network sites or participate in online discussion for a). Apart from communicational purposes, organizations may use user-generated content as input for innovations and co-creation or market research.

Web 2.0 aligns with an emerging, dominant logic in marketing which argues that value is defined by collaborating and co-creation with- and learning from customers (Vargo & Lusch, 2004). This approach is called the service-dominant (S-D) logic of marketing. It argues all organizations are in the business of providing services where those that produce goods only do so as a means of “transmitting” their services to the customer (Maglana, 2007). The S-D logic of marketing redefines the relationship between the organization and the customer where the latter has been promoted to a co-producer (rather than primarily a recipient) of value. Marketing is a process whereby there is a constant stream of communication between the firm and the customers to improve the quality of the value offer (Vargo & Lusch, 2004). The service-centered view of marketing is customer-centric and market-driven, what goes beyond being consumer oriented; it means collaborating with- and learning from customers and being adaptive to their individual and dynamic needs. A service-centered dominant logic implies that value is defined by- and co-created with the consumer, who is able to provide direct or indirect feedback regarding what he thinks of the offering. This is referred to as a “sense-and-respond” strategy as opposed to a “make-and-sell” strategy (Haeckel, 1999; as cited by Vargo & Lusch, 2004) which essentially redefines a firm’s objective from merely “making the sale” to maintaining an ongoing relationship with the customer.

2.2 SOCIAL MEDIA, WEB 2.0 & USER-GENERATED CONTENT IN CONTEXT

Before elaborating on microblogging, this paragraph discusses terminology. Defining social media requires the explanation of two related concepts that are frequently named in conjunction with social media: Web 2.0 and User-Generated Content (UGC).

As previously mentioned, the term Web 2.0 is used to describe the new way in which consumers and developers started to utilize the World Wide Web; that is, as a platform whereby content and applications are no longer published by a small group of institutions, while the end-users are merely content consumers. Rather content is a product of continuous active collaboration, participation and interaction of consumers.

While Web 2.0 represents the ideological and technological foundation, User-Generated Content (UGC) can be seen as the sum of all ways in which people make use of social media. The term is usually applied to describe the various forms of media content that is created and published by end-users and which is publicly available. While UGC existed prior to Web 2.0, the combination of technological drivers (e.g., increased availability of broadband), economic drivers (e.g., increased availability of tools for the creation of UGC) and social drivers (e.g., rise of a generation with substantial technical knowledge and willingness to engage online) make UGC nowadays fundamentally different.

Based on these clarifications of Web 2.0 and UGC, social media can be defined as “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User-Generated Content” (Kaplan & Haenlein, 2010). Within the social media landscape, there are various types of social media. Kaplan & Haenlein (2010) developed a classification scheme based on two dimensions: social presence/media richness and self-presentation/self-disclosure.

		Social presence / Media richness			
		Low	Medium	High	
Self presentation / self-disclosure	High	Blogs (e.g. Blogspot)	Microblogs (e.g. Twitter)	Social networking sites (e.g. Facebook)	Virtual social worlds (e.g. Second Life)
	Low	Collaborative projects (e.g. Wikipedia)	Content communities (e.g. YouTube)	Virtual game worlds (e.g. World of Warcraft)	

Figure 3: Classification of social media by social presence/media richness and self-presentation/self-disclosure (Kaplan & Haenlein, 2010)

On the social presence & media richness dimension, the social media differ based on the amount of information they allow to be transmitted, and the extent to which acoustic, visual, and physical contact can be achieved between two communication partners. On the self-disclosure & self-presentation dimension, the social media differ based on the extent to which personal information revealed. On the continuum of social media classification, microblogging is characterized by a relatively high self-presentation/disclosure, and a low to medium level of social presence/media richness.

2.3 MICROBLOGGING

In order to define microblogging, the concept of blogging needs to be elaborated on first. The term blogging is a contraction of the words web and log, and it refers to a type of website in the form of a public journal where one or multiple authors publish articles about their personal experiences or regarding a specific topic. Blogs are interactive, allowing visitors to leave comments and message

each other. It is this interactivity that distinguishes them from other static websites. Most blogs are primarily textual, nevertheless blogs may combine text, video, images, and links to other websites.

Microblogging is a form of blogging that enables users to compose brief text updates (usually less than 200 characters) which are delivered to the user's network of associates. A microblog differs from a traditional blog in that its content is typically smaller in file size and length. Typically, microblogs are optimized for smartphone usage, although it's not intrinsically in its definition. This enables users regardless of physical location and device to access and update their microblog.

The success of microblogging can be identified by three factors (Kaplan & Haenlein, 2011). The first reason relates back to the Greek philosopher Aristotle who coined that sometimes the whole is greater than the sum of its parts. Similarly, different microblog postings sent out over time can paint a very accurate picture of a person's activities. This concept is called ambient awareness. Just as physical proximity allows one's mood to be interpreted through a series of little behaviors (e.g., body language, sighs and stray comments), several microblog postings together can generate a strong feeling of closeness and intimacy. The second reason behind the popularity of microblogs is the unique type of communication they allow. The unique combination of push-push-pull; communication from sender to followers, from followers to followers and from receivers to external information sources. The last factor is the fact that microblogging creates the perfect environment for virtual exhibitionism and voyeurism. Generally, microblog postings are public by default. Additionally, microblog postings become public knowledge within minutes of its publication as they are incorporated by search engines. Empirical proof for this statement can be found by considering that people are more likely to watch reality programs when they have a higher voyeuristic tendency caused by factors such as the disclosure of personal information, gossip, and private emotions (Baruh, 2009; as cited by Kaplan & Haenlein, 2011).

2.4 TWITTER

Twitter is currently by far the most popular microblogging service. Twitter was created by a San Francisco-based 10-person start-up called Obvious Corp., and launched in October 2006. Twitter enables its users to share microblog postings, called 'tweets', limited to 140 characters, to a Web interface, where they are publicly available. Twitter is optimized to be used anytime anywhere. Tweets may be posted via twitter.com, text messaging, via Twitter's mobile Website m.twitter.com, or using third party clients. By March 2010, the company recorded over 70.000 registered applications¹¹, ranging from mobile and desktop Twitter clients to tools make use of the data generated on Twitter. Even in situations where there's no internet, such as during the uproar in Egypt in early 2011, services such as Google's Voice-to-Tweet enabled offline Egyptians to tweet by leaving a voicemail on an international phone number¹². The Twitter ecosystem is extensive as Twitter makes an API available for developers. The API is an interface which enables others to develop software which have access to the Twitter ecosystem, i.e. they have access to the stream of Twitter data. The character limit allows tweets to be produced, consumed and shared at minimal effort, allowing a fast paced conversational environment to emerge. The central feature of Twitter, which users see when they log in, is a real-time stream of tweets posted the user's network of associates, listed in reverse chronological order. Like social network sites, profiles are connected through an underlying articulated network. Users declare the people they are interested in following. A user who is being followed by another user does not necessarily have to reciprocate by following them back, which makes the links of the Twitter social network directed.

Within a year of its launch, Twitter already hosted about 5.000 tweets per day. As of September 2010 Twitter hosts over 95.000.000 tweets per day¹³, that's an average of 1100 tweets per second. As of September 2010, Twitter has over 175 million registered users. Twitter had nearly

¹¹ New Statesman, 2010. <http://www.newstatesman.com/digital/2010/03/twitter-registered-created>

¹² The New York Times, 2011. Retrieved from <http://www.nytimes.com/2011/02/02/world/middleeast/02t...>

¹³ Twitter Inc, 2011. Retrieved from <http://twitter.com/about>

96 million unique visitors in August 2010, up 76% from the same period last year¹⁴. Half of the tweets are in English, reflecting its high penetration rate in English-speaking countries and the tendency of Twitter users that are non-native English speakers to tweet in English. The sixth to eight languages on Twitter are major European languages, namely Italian, Dutch and German, each accounting for about 1% to 2% of total messages, thereby surpassing i.a. French, Chinese and all Scandinavian and East-European languages (Guyot, 2010). However, the distribution of English tweets of this study shows Twitter is still dominated by American users ([figure 17](#)).

Mentions & Conversations

Direct posts and mentions are used when a user aims to refer to a specific person, whereas the regular updates are not directed or referred to a specific user. These mentions emerged after they were introduced by the early adopters of Twitter. They used the '@' sign followed by the ID of the user to direct a messages to that user, as a form of addressivity, which originated from Internet Relay Chat (IRC) (Werry, 1996). Werry (1996) noted that a high degree of addressivity is required in multi-participant public environments such as IRC, where mentions function as attention-seeking; it is a specifically intended to alert the mentioned person that they are being talked about.

Conversations are not marked as such within the Twitter ecosystem; nevertheless it is how this study will refer to it. Conversations are those tweets where the mention is located at the beginning. That way, the tweet will only be addressed to that specific user. It is this where conversations differ from mentions; as tweets mentioning individuals or organizations are published to the entire network of associates. Around 25.4% of all posts are directed, which shows that this feature is widely used among Twitter users (Huberman et al., 2008). Similarly, Mischaud found that in his sample, "many postings often read like fragments of virtual conversation" (p.30). In the sample of Honeycutt & Herring (2009), 15.7% of the tweets were found to be directed to a specific person, while Boyd, Golder et al. (2010) and Kong et al. (2009) report significantly higher percentages, 36% and 35% respectively.

Retweets

Retweeting is the act of reposting content. A retweet is a message which is replicated by another user to its network of associates. A retweet follows the 'RT @userID: message' syntax. When performing a retweet, the original tweet as well as the user ID is replicated to all of the retweeter's followers. During Q4 2009, Twitter rolled out its new retweet functionality, whereby rather than the syntax, a retweet is a tweet which directly references back to the original tweet. Both new and old style retweets are widely used. To an inferior degree, the "via @userID" syntax has also emerged as a way of reporting content.

Structurally, retweeting is the Twitter-equivalent of email forwarding where users post messages originally posted by others. While retweeting can simply be seen as the act of copying and rebroadcasting, the practice contributes to a conversational system in which information is diffused, and where this information is validated and engaged with by other users, still with the original author in mind (Boyd et al., 2010). This convention serves various purposes, for instance, showing sympathy, or acknowledgement on a value of a certain tweet rather than the user of the original tweet. For this reason, some of the most visible Twitter participants retweet others and look to be retweeted.

Hashtags

Like retweeting and mentioning, the usage of hashtags has also emerged during the early years. The usage of the hashtag functionality allows individuals to automatically co-construct a resourceful site where the active participation of a micro-network on a given topic is aggregated through a special hashtag (#), followed by a keyword identifying the topic. It allows the dispersed network to come together into one single topic almost in an instant way. The practice of using the '# + keyword' syntax to label tweets most likely parallels the use of "tags" to freely categorize Web

¹⁴ comScore Media Metrix Ranks Top 50 U.S. Web Properties. Retrieved from http://comscore.com/Press_Ev...

content, which gained visibility in social bookmarking. The practice of using hashtags may stem from a history among computer programmers of prefacing specialized or variable words with punctuation marks, such as \$ and * and # (Huberman & Golder, 2006; as cited by Boyd et al. 2010). Hashtags have become a quite successful way of connecting the remote network to a given event and also creating a collaborative resource based on spontaneous reaction and unpremeditated story-telling (Costa et al., 2008). Boyd et al. (2010) found that 5% of tweets contained a hashtag.

Whereas mentioning, retweeting and the usage of hashtags have emerged thanks to early adopters of Twitter, the functionalities are nowadays incorporated into the Twitter ecosystem.

2.5 TWITTER STRATEGIES

2.5.1 LISTENING TO CUSTOMERS

One to one marketing is an important underlying construct for Twitter strategies as proposed by professional literature. The purpose of one to one marketing (also called relationship marketing or customer-relationship management) is to understand each customer well and foster high customer loyalty (Pine, Peppers, & Rogers, 1995). The more customers teach the company, the more solid the knowledge database of the company will be, which will make the company more capable of adjusting its value offer to market demands. Historically, marketing researchers have always struggled to integrate customers into their decision-making processes. At the same time, the concept of customer knowledge as a source of competitive advantage has become increasingly prominent in the academic literature (Vargo & Lusch, 2004). The more a company is taught by its customers, the larger the competitive advantage.

The essence of one to one marketing is knowing what customers and potential customers are saying about the brand. Microblogs allows for investigation of what customers really feel about the brand and its competitors in real time (Jansen & Zhang, 2009; Comm, 2009; Kaplan & Haenlein, 2010). As opposed to other social media, Twitter is much more communal. On Twitter by default all profiles and tweets are publicly available, as opposed by, for example, Facebook which requires relations to be reciprocal before profile information is accessible. Professional literature suggests companies should follow their target group in order to monitor them, to see what drives them, what's keeps them busy, i.e. what is going on in their world. Twitter is a great outlet to share quick thoughts and information which makes it an interesting platform to acquire customer knowledge.

Listening to customers and their brand statements is generally considered as the first step towards managing online brand reputation and/or customer service. There's a wide range of applications available which makes companies able to track their brand mentions in real time. Listening to customers on Twitter may also serve the purposes of input for market research, future innovations or co-creation (Kaplan & Haenlein, 2011). Online brand management and customer service are typically classified as outside to outside strategies, whilst market research and co-creation are considered outside-in (Kerkhofs et al., 2010).

2.5.2 ONE-TO-ONE COMMUNICATION

One-to-one marketing is a paradigm which focuses on developing a marketing strategy to interact with individual consumers. This approach first focuses on gradually understanding consumers, and then it customizes the value offer to the consumer's needs (Pine, Peppers & Dorf, 1999). This is different from traditional marketing in terms of the broadness of consumer contact, as traditional marketing mainly focuses on marketing mass consumers. One-to-one marketing focuses on customer satisfaction and is customer oriented.

One-to-one marketing is executed as one-to-one communication through 'conversations' on Twitter ([paragraph 2.4](#)). For Twitter strategy, one to one communication is an essential component (Comm, 2009; Blom, 2009; Raman, 2010). This is what makes Twitter an interesting communication tool, as it provides a platform to connect directly with customers and join the conversation. Though,

this will require a different strategy from more traditional, more push media. Organizations will need to actually add relevant value for their target group, or the message won't reach its intended effect (e.g. read, replication). Professional literature suggests this one to one communication is very useful for especially customer service and online brand management.

While research has shown the desirable effects of positive feedback and recommendations (Reichheld, 2003), it has also underlined the devastating impact which negative customer comments can have on an organization (Goldenberg et al., 2007; Richins, 1983). This stresses the importance of managing dissatisfied- and complaining consumers. Moreover, engaging with dissatisfied consumers quickly, organizations can avoid issues growing out of proportion and evolving into organized forms of consumer protest, e.g. consumer boycotts (Garrett, 1987) or complaint websites (Ward & Ostrom, 2006).

An adequate customer service on Twitter will prevent consumer frustration and may solve issues or influence the current discussion. Adequate brand monitoring on Twitter influences the current debate about its brand thereby positively affecting the mindset of the direct participants about the brand. Indirectly it also influences potential customers and other stakeholders. Word of mouth messages are archival in the sense that they permanently exist and are searchable via Web search engines and other services (Gelb & Sundaram, 2002; Kiecker & Cowles, 2001). The broad reach of eWOM therefore influences brand image and perceptions (Reynolds, 2006; Urban, 2005). As such, eWOM is increasingly important for organizations concerned with reputation management. The challenge for the brands is to influence this online appearance in a positive way so consumers' brand image and perceptions are positively affected by it. Also, by positively helping consumers, brands may win their hearts and minds. Possibly they may create brand advocates; consumers who have favorable perceptions about a brand and recommend it to their network. Brand advocacy is directly correlated to business growth, as found by Reichheld (2003). In accordance, another study by Keller (2005) showed that 91% of people would be likely to use a brand recommended by someone who has used it themselves.

2.5.3 COMMUNITY PARTICIPATION

Tweets and conversations on Twitter are dispersed around the network. Hundreds of users could be talking about the same topic within their own personal network, while neither of the conversations crosses each other. In order to aggregate the dispersed network into a single topic, the hashtag was introduced by early Twitter adaptors. Including a hashtag acts as a way of creating categories, groups or topics for tweets so they are more easily found by the people interested in a particular topic. Hashtags have become a quite successful way of connecting the remote network to a given event and also creating a collaborative resource based on spontaneous reaction and unpremeditated story-telling (Costa et al, 2008). Using hashtags, topic, event or brand oriented micro-communities are created. Communities on Twitter are very dynamic and can emerge and fade away on a daily basis (Blom, 2009). The hashtags function is particularly valid to setup a temporary or more permanent community on Twitter (Lacy, 2010). Creating a community around the brand may likewise be beneficial. Membership in even trivial or minimal groups has been shown to produce social identification which, in turn, produces measurable in-group bias (e.g., Tajfel, 1970; Diehl, 1990, as cited by Thompson & Sinha, 2008). As a result, members tend to evaluate the in-group more favorably while evaluating the out-group more negatively (Hogg & Abrams, 2003; as cited by Thompson & Sinha, 2008).

Temporary or more permanent communities also enable organizations to interact with a greater number of users than it would normally reach when addressing its personal network. As such, tweets are found by relevant audiences who are not in the organization's first-degree network.

Hashtag usage may also be beneficial when it is used for brand monitoring, i.e. when monitoring company or product sentiment. Furthermore, the trending (hot) topics are listed on the homepage of Twitter. This leads to even more exposure to the hashtag.

Altogether, community participation leads to a greater word of mouth influence, as it enables organizations to interact not only with first degree network, but also with target audiences outside of the direct network.

2.6 INFLUENCE

Influence has long been studied in the fields of sociology, communication, marketing and political science (Rogers, 1962; Katz & Lazarsfeld, 1955). The notion of influence plays a vital role in how businesses operate and how a society functions, for instance, how fashion spreads (Gladwell, 2002) and how people vote (Keller & Berry, 2003). A large body of research is dedicated to identifying sources of influence as antecedents of influence, such as source expertise, tie strength, demographic similarity and perceptual affinity.

Studying influence patterns, however, has been difficult. This is because such a study does not lend itself to readily available quantification, and essential components like human choices and the ways our societies function cannot be reproduced within the confines of the lab. Previous studies assessing influence through the word of mouth mechanism have quantified influence in terms of diffusion tree size or network metrics. Measuring diffusion trees, social influence was shown to significantly affect adoption rates, and this occurred more rapidly among friends than among strangers (Bakshy et al., 2009). Other research showed that diffusion of Facebook fan pages can be predicted with the user's demographics or Facebook usage characteristics (Sun et al., 2009). Aral et al. (2009) showed adoption of a mobile phone application over the Yahoo! messenger network could be predicted by homophily (Aral et al., 2009), whereas another study was able to identify leaders for Yahoo! Movie user actions (Goyal et al., 2010). More closely related to the current research is a series of recent papers all published during the execution of this paper, which have quantified influence in terms of both network metrics and diffusion on Twitter specifically. Kwak et al. (2010) compared three different measures of influence: the number of followers, Page-Rank and number of retweets. Cha et al. (2010) also compared three different measures of influence: the number of followers, the number of retweets and the number of mentions. Finally, Weng et al. (2010) compared number of followers with a modified Page Rank measure that accounts for the topic.

Nevertheless, there have been important theoretical studies on the diffusion of influence, albeit with opposing results. The traditional view assumes that seeding a piece of information using a minority of members whose connectivity of position in the society allows them to trigger a disproportionately large amount of the population. They are generally described as being informed, respected and well-connected; they are called the opinion leaders in the two-step flow theory (Katz & Lazarsfeld, 1955), influentials or influencers in marketing literature (Keller & Berry, 2003) and hubs, connectors, or mavens from network perspective (Gladwell, 2002). By targeting influentials in the network, marketers may achieve a large-scale chain-reaction driven by word of mouth at minimal marketing expenses (Katz & Lazarsfeld, 1955). The theory of influential has gained huge popularity in the field of marketing (Chan & Misra, 1990; Coulter et al., 2002; Myers & Robertson, 1972; Van den Bulte & Joshi, 2007; Vernet, 2004; as cited by Watts & Dodds, 2007).

A more modern view, in contrast, de-emphasizes the role of influentials (Bakshy et al., 2009; Watts & Dodds, 2007; Watts, 2007). People in the new information age make choices based on the opinions of their peers, rather than by influentials (Domingos & Richardson, 2001). This modern view of influence leads is called collaborative filtering. It argues the theory of influential intuitively compelling, but its models do not explain how information actually spread (Watts & Dodds, 2007). Moreover, it is argued that influentials have little impact on social epidemics. Researchers argue that direct marketing through influentials would not be as profitable as using network based approaches like collaborative filtering. The theory of influentials is criticized because it does not take into account the role of ordinary users. Research simulation has shown that influentials did not initiate all diffusions, moreover, in homogeneous networks, influentials were no more successful in initiating long cascades than ordinary users (Watts & Dodds, 2007; Watts et al., 2007). This means that a trend's success depends not on the person who starts it, but on how susceptible the society is to the

trend. The theory proposes that electronic word of mouth communications are an effective way to transform consumer networks into influence networks, capturing consumers' attention, triggering interest and eventually leading to adoption or sales (De Bruyn & Lilien, 2008).

2.7 WORD OF MOUTH (WOM)

Word of mouth diffusion is regarded as an important mechanism by which information can reach large populations, thereby influencing public opinion (Katz & Lazarsfeld, 1955), consumer decisions for purchasing new products (e.g. Engel et al., 1969; Arndt, 1967; Richins, 1983; Richins & Root-Shaffer, 1988; Whyte, 1954; Bansal & Voyer, 2000), but also shape consumer expectations (Anderson & Salisbury, 2003; Zeithaml & Bitner, 1996), pre-usage attitudes (Herr, Kardes & Kim, 1991), as well as post-purchase product perceptions (Bone, 1995; Burzynski & Bayer, 1977) and risk reduction associated with buying decisions (Murray, 1994; Godes & Mayzlin 2004). Word of mouth diffusion may be useful for marketers when they make use of consumers' network to spread the marketers' message. WOM communication strategies are appealing because they combine the prospect of overcoming consumer resistance against traditional forms of communication, with significantly lower costs and fast delivery (Trusov et al., 2009). Positive WOM communication is considered a powerful marketing tool for influencing consumers. Customers may spread the marketing message because they are pleased with a brand (positive WOM) or because they are dissatisfied with it (negative WOM). Both positive and negative WOM have different motivations behind it (Anderson, 1998). The major incentive for people to spread positive WOM is to gain social or self-approval. WOM sentiment has shown asynchronous effects in the sense that the impact of negative WOM was stronger than the impact of positive WOM (Chevalier & Mayzlin, 2004; Anderson, 1998). Additionally, altruistic behavior of sharing expertise with others has also been shown to motivate positive WOM (Fehr & Falk, 2002; Richins, 1983). Hostility (Jung, 1959; Kimmel, 2004) and vengeance (Richins, 1983) motivates dissatisfied consumers to engage in negative eWOM.

Whereas business to consumer is perceived as subjective, word of mouth is perceived as more reliable, credible and trustworthy. WOM has shown to be more effective in situations than personal selling and various types of advertising (Katz & Lazarfeld, 1955; Engel et al., 1969; Feldman & Spencer, 1965). Furthermore, WOM has a greater impact on product judgments than printed information (Herr, Kardes & Kim, 1991). Customers acquired through WOM improved sales & market share (Danaher & Rust, 1996) and add more long-term value to the organization than customers acquired through traditional marketing (Villanueva et al., 2008).

Interest in WOM communication has been revitalized in marketing practice through its proposed role in fashion and other diffusion processes (Gladwell, 2002), as well as through its role in virtual communities (Hagel & Armstrong, 1997). Furthermore, traditional forms of communication appear to be losing effectiveness (Nail, 2005). More specifically, the Internet has emerged as a source and an outlet for electronic word of mouth communication for customers (Hennig-Thurau et al., 2004).

2.8 ELECTRONIC WORD OF MOUTH (EWOM)

Web 2.0 has revolutionized the speed and the scope of word of mouth communication. The emergence of Web 2.0 has revitalized marketer's interest in word of mouth diffusion for its proposed power in creating viral effects (Watts, 2002). Although similar to earlier forms of word of mouth, eWOM differs significantly from traditional WOM. eWOM is many times more anonymous and confidential. Moreover it provides geographical and temporal freedom. Compared to WOM, in an eWOM context there is much less social context, such as verbal nuances (e.g. gaze, body language), physical context (e.g. meeting sites, seating arrangements) and observable social characteristics (e.g. age, gender, race). Combined with the high level of anonymity, this can cause high levels of insecurity and uncertainty (Daft & Lengel, 1986), and possibly anti-social and aggressive behavior (Kiesler et al., 1985; Dubrovsky et al., 1986, as cited by Brown et al., 2007).

However, this is inconsistent with the contemporary growth of electronic word of mouth communication. Similar to WOM, eWOM has been shown to significantly influence consumer decisions for purchasing new products (e.g. Dellarcas et al., 2004, Chevalier & Mayzlin, 2004). Research has shown that eWOM may have higher credibility, empathy and relevance to customers than marketer created sources of information (Bickart & Schindler, 2001). Research also indicates that people appear to trust seemingly disinterested opinions from people outside their immediate social network, such as online reviews (Duana, Gub & Whinston, 2008). eWOM may be less personal in that it is not face-to-face, but it is more powerful because it is immediate, has a significant reach, is credible and publicly available (Hennig-Thurau et al., 2004). People are willing to accept information equally in either WOM or eWOM context (De Rooy, 2009). Walther (1992) found that online communities showed positive, socially rich, relational behavior and both friendly and romantic relationships developed (Walther, 1992, 1996; Tidwell & Walther, 2002; Kraut et al., 1998; Utz, 2000; as cited by Brown, 2007). Altogether, it is clear that social resources such as emotional support, companionship and a sense of belonging are visibly exchanged between online individuals (Haythornthwaite, 1999). Credibility within eWOM relies heavily on the altruistic nature of the sender as opposed to the observable attributes by which a sender is judged in non-electronic WOM context (Steffes & Burgee, 2009).

Online messages are also archival in the sense that they permanently exist and are searchable via search engines and other online services (Gelb & Sundaram, 2002; Kiecker & Cowles, 2001). Because of its broad reach and ability to influence consumer opinion and (actual) purchase decisions (Chen & Xie, 2008; Davis & Khazanchi, 2008), eWOM is considered increasingly important by organizations. Word of mouth on Twitter is also asynchronous noninvasive, since one can choose who to receive updates from, and sender and receiver are separated in space. eWOM can occur very near the purchase decision or even during the purchase process (Barton, 2006).

2.9 MEASURING INFLUENCE

Whereas measuring influence used to be difficult, word of mouth on Twitter is observable and influence may be directly compared across brands and consumers. The operational definition of influence for the current study is somewhat narrow, since it focuses on the brand's ability to diffuse information through Twitter's social graph. The concept is quantified using the following indicators: followers indegree, retweet indegree, mentions indegree and sentiment.

2.9.1 FOLLOWERS INDEGREE

The evidence for the follower/following principle is supported by permission marketing, a theory proposed by Seth Godin. Permission marketing is the privilege (not the right) of delivering anticipated, personal and relevant messages to people who actually want to receive them (Godin, 1999). It recognizes the power of consumers to ignore marketing and that treating people with respect is the best way to earn their attention.

Twitter is asynchronous noninvasive: social relations are not necessarily reciprocated, i.e. directed, nor modulated and are mostly focused on the exchange of information. It is through the voluntary process of deliberately following a brand that permission based marketing is achieved. Since consumers have just as much of a choice to follow a brand as not follow it, it is guaranteed that the message will only reach those consumers who opt-in.

Another theory supporting the evidence of the followers indegree is that of big seed marketing, a model proposed by Watts, Frumin & Peretti (2007). It is a theory which combines the power of traditional advertising with viral propagation. It argues that viral marketing campaigns rarely reach exponential growth. Nevertheless, by combining the viral capabilities with large initial



Figure 6: Followers counter has a central position in the layout of Twitter

seeding, campaigns can succeed in reaching large populations. It is argued to be relatively reliable as compared to pure viral theories, moreover, it is straightforward to implement. It overcomes the unpredictability and difficulty of reaching large audiences using purely viral techniques.

Previous studies showed that individuals consume more content from network associates than from people outside their direct network (Kerman, 2007; Kerman & Jones, 2007; Sun et al., 2009). Likewise, bloggers are more likely to join a group that many of their associates joined [2]. The fact that individuals act like their network of associates is in line with collaborative filtering theories.

The follower indegree is the first degree network who receives the brand's message. The assumption is that the larger the number of initial followers, the further the message will spread across the entire social graph. The follower indegree is widely used measure in social media monitoring and previous studies (Cha et al., 2010; Kwak et al., 2010; Weng et al., 2010). It is also well covered by professional literature (Comm, 2009; Raman, 2009; Lacy, 2009). Moreover, the amount of followers has a prominent position within the layout and system of Twitter.

2.9.2 RETWEET INDEGREE

The second measure is the retweet indegree. The retweet indegree indicates the amount of users who replicated the brand's message to their followers. It is argued to be the highest degree of content approval; entailing the tweet was so valuable that the user was willing to share it with their network, thereby putting his or her own reputation on the line. Retweets measures the actual spreading of an eWOM message between Twitter users. This measure has been used in previous studies (Kwak et al., 2010; Cha et al., 2010). Although conversational, the retweet indegree does not represent all of the times content is reported on Twitter. Instead, it are only those tweets which explicitly attribute the original user. Unless a URL is present (which allows for statistics), it is generally infeasible however to include all instances in which content was reproduced. On contrast, an advantage of using retweet indegree is that it doesn't incorrectly attribute influence to what in reality are independent events.

2.9.3 MENTIONS INDEGREE

The third performance measure is the mentions indegree. Because Twitter is like a giant open chat, the more people who reply to tweets, the more influential the tweets are. It's a great sign that people are interested in what a brand has to say and want to take part in the discussion. That's also the theory behind TwitterRank, a service that uses the number of incoming replies to give each Twitter user a score that supposed to represent their popularity. Moreover, mentions indegree has been used in previous research (Cha et al., 2010). The theory is similar to Google's Page Rank, which rates the importance of Web sites based, among other things, on the number and quality of incoming links the site receives. In the end the number of incoming mentions represents the brands's ability to engage others in conversation. Previous experimental research by Weng et al. (2010) showed that their model using the TwitterRank Page Rank algorithm outperformed the node in-degree in the network, i.e., the number of followers and other related algorithms, including the original PageRank and Topic-sensitive PageRank.

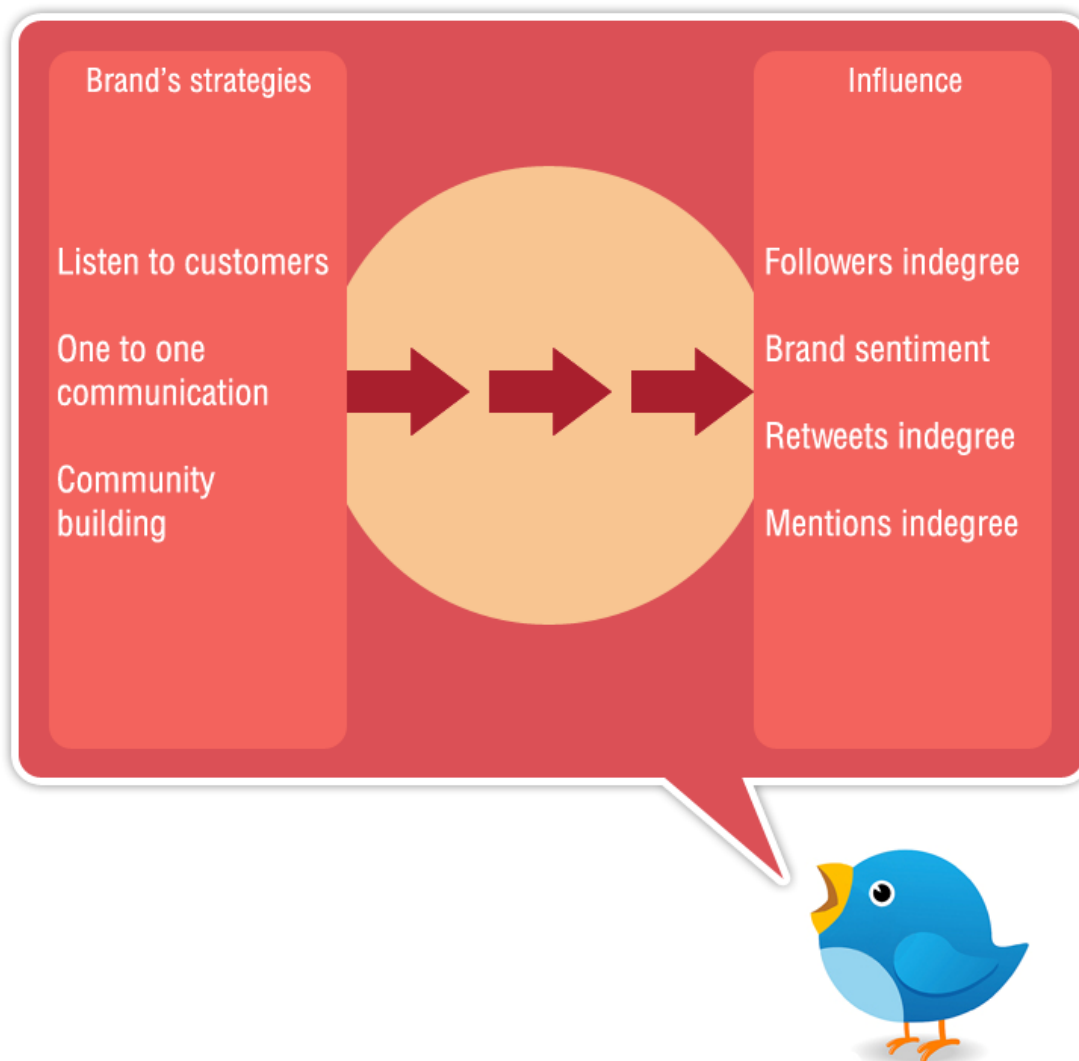
2.9.4 SENTIMENT ANALYSIS

The rise of social media and Web 2.0 has revitalized interest in sentiment analysis. Sentiment analysis, or opinion mining, deals with the computational treatment of opinion, sentiment and subjectivity in text (Pang & Lee, 2008). The sentiment is determined by a complex algorithm. The algorithm is trained to determine how people use adjectives in online utterances. Text is flagged either neutral, positive or negative, based on words which imply positivity or negativity, in relation to its context. The algorithm is constantly trained to cope with cultural factors, linguistic nuances (e.g. humor, irony) and differing contexts. Earlier research conducted by Liu, Hu & Cheng (2005) resulted

in the development of an application for analyzing and comparing consumer opinions for a set of competing products. Archak, Ghose & Ipeirotis (2007) examined online product reviews in order to identify specific product characteristics and then weight each in terms of importance to customers. Wijaya & Bressan (2008) leveraged the Page Rank algorithm to predict box office numbers based on peer reviews. The study showed that their model was a strong predictor of box office rankings.

Jansen & Zhang were one of the first researchers to apply sentiment analysis to Twitter. They analyzed microblog postings and showed that 19% contained the mention of a brand. Of those, more than 50% were positive and 33% were critical about the brand or product. Jansen & Zhang also compared automated and manual coding, showing no significant difference between the two approaches. They also investigated the change in sentiment. Huberman & Asur (2010) assessed sentiment on Twitter, showing its ability to predict future box office revenue. Huberman & Asur found that the rate at which people produce tweets combined with the sentiment they express can accurately forecast the box office revenue of the film. And the predictions from tweets are more accurate than any other method of forecasting. Bollen et al. (2010) found that collective moods derived from Twitter can be used to predict the stock market. With an accuracy of 87.6% the researchers were able to predict the daily up and down changes in the closing values of the Dow Jones Industrial Average.

2.10 CONCEPTUAL MODEL



CHAPTER 3: METHODOLOGY

3.1 CHOICE OF RESEARCH METHOD

Historically, measuring influence over word of mouth networks has been difficult. Influence generally is intangible and difficult to measure, especially in case of large networks with multiple levels. Digital social media, Twitter in particular, are promising because they allow for a detailed investigating of flows of influence over word of mouth, thereby overcoming the issues of unobservability and sampling difficulties.

A method well suited for the informational needs of this study are observation methods. Such a real life observational research method is favorable because of its ultimate level of external validity. Nevertheless, researchers often settle for a laboratory observation study. Real life observation is in many cases infeasible, complex and expensive. However, the nature of the Twitter ecosystem lends itself perfectly for gathering tweets over a period of time using the Twitter API. This actually makes it perfectly feasible to execute a real life observation study using this mechanical form of observation. Using actual Twitter data is favorable as its reliability does not reside in the mind of the respondents trying to recall a huge amount of data, rather, actions are being directly observed.

Such a research method, however, requires a fair amount of technological development. Thankfully, the current study has received assistance for migrating and measuring data, data calculation and application development. During a 9 week pretest, which lasted from 30-07-2010 till 30-09-2010, the application was tested to control for any problems possibly encountered. The actual 10-week lasting data collection took place between the 12th of November 2010 and the 20th of January 2011.

3.2 RESEARCH SAMPLE

Considering the fact that this study is a master thesis and the hypothesis are of exploratory nature, a sample of 30 brands will suffice. The brands are selected from different industries, in order to prevent industry-specific observations.

Figure 7: Brands selected for sample

Beverage Industry			
Soft drinks	Twitter account	Carbonated & Energy drinks	Twitter account
Coca Cola	@CocaCola	AMP Energy	@AMPEnergy
Coke Zero	@CokeZero	Gatorade	@Gatorade
Dr. Pepper	@DrPepper	Monster Energy	@monsterenergy
Pepsi	@pepsi	Mountain Dew	@mtn_dew
Tropicana	@tropicanaoj	Red Bull	@redbull
Electronics Industry			
Consumer electronics	Twitter account	Smartphones	Twitter account
Canon	@canon_camera	Blackberry	@Blackberry
Hewlett-Packard	@HPnews	HTC	@HTC
Logitech	@Logitech	Motorola	@MotoMobile
Microsoft	@microsoft	Nokia	@nokia
Sony	@SonyElectronics	Sony Ericsson	@SonyEricsson
Travel Industry			
Transportation	Twitter account	Leisure	Twitter account
British Airways	@British_Airways	Carnival Cruises	@CarnivalCruise

Delta Airlines	@Delta	Disney Resorts	@disneyparks
Hertz	@ConnectByHertz	Expedia	@Expedia
JetBlue	@JetBlue	Holland America Line	@HALcruises
Southwest Airlines	@SouthwestAir	Marriott	@MarriottIntl

3.3 DATA PROCESSING

3.3.1 DATA COLLECTION & HARDWARE SETUP

In order to capture tweets from the Twitter API in the pretest, an application which subtracts Twitter data was written. The application has been written by the author and [Serhat Gülçicek](#), a Software Engineer at Logica. The application is written in Java, an object-oriented programming language developed by Sun Microsystems Inc. The application runs simultaneously on two servers (ANNE & ELISA, both running Linux Debian 5.0.3 (stable)). Both servers are connected to the web via a 35/35 Mbps fiber connection in The Netherlands. Sentiment data was gathered using Tweetfeel Biz, a sentiment analysis tool by Conversation. To avoid loss of data in case of problems, two servers were used simultaneously. In case one server goes down (e.g. power failure, hardware failure, disconnected from internet), there is a second server to back up the data gathering process. In addition, both servers are not at the same physical location, in order to control for local failures causing both servers to go down. Both servers run an independent MySQL database, and are synchronized only after the data collection is finished. For a detailed description of how data was gathered in the pretest, see [appendix 6](#).

At the time Conversation provided their tool Tweetfeel Biz. Due to some limitations of Tweetfeel Biz (not possible to monitor Twitter account names, skipping of neutral and ambiguous (e.g. slang, irony) tweets), and because the study was offered to work with Radian6, it was decided to drop the custom application as well as Tweetfeel Biz in favor of Radian6. Social media monitoring tool Radian6 was offered to the study by The Webcare Company (see figure 8).

Using Radian6 for data collection offers several advantages compared to the pretest. By default, Twitter doesn't make all of the Twitter data available to developers, unless you have "firehose" access. To date, this has only been granted to very large companies such as Google. Nevertheless, Radian6 also has a full firehose contract and therefore has access to 100% of the tweets. During the actual data collection, 22,4% more tweets were gathered than in the pretest. This is not solely accountable to the full firehose contract, the body of tweets increases over time which therefore increases the number of tweets collected. Moreover, Radian6 identifies the sentiment of all the tweets it gathers, thus no secondary tool is required. During the pretest, the data from Tweetfeel Biz inevitably differed from that of the research application. Using Radian6, also it became unnecessary to cover for hardware or connection failures, and the data was more precise (number of following/followers is checked for each tweet, not every two hours). Finally, Radian6 also incorporates new style retweets, whereas in the pretest this was not possible.



Figure 8: About The Webcare Company and Radian6

3.3.2 WHAT DATA IS GATHERED

During the pretest, the application was able to crawl Twitter for brand mentions and retweets. A second script hourly crawled the brand Twitter profiles and inserted the data in the database. The sentiment was determined by Tweetfeel Biz.

For the actual research period, Radian6 replaced the research application and Tweetfeel Biz. Radian6 collected all data required for the study. The tool collects and archives all data which it subtracts from Twitter. Within Radian6, search sets tell Radian6 what data to pull from its archive. For the study, all tweets posted by the brands in the research sample as well as all tweets which mentioned the brands' Twitter account during the research period were pulled from the database.

A single record (tweet) in the database contains the following cells and information:

Figure 9: Cell data contained by a single record

Cells	Cell data
ARTICLE_ID	The unique ID of the tweet in the Radian6 database
AUTHOR	The accountname of the user who posted the tweet
CONTENT	The actual tweet message
ARTICLE_URL	The URL to the individual tweet on Twitter.com
PUBLISH_DATE	The date-time of the tweet when posted. Follows the format: MMM DD, YYYY HH:MM TT
FOLLOWING	The amount of users the tweet posters follows
FOLLOWERS	The amount of users who follow the tweet poster
UPDATES	The cumulative number of the poster's tweets
BLOG_POST_SENTIMENT	The classified sentiment

3.4 PROCESSING THE DATA

When the data collection is finished, the data from Radian6 is exported to .CSV files (raw data separated by commas). Using a Java application, the raw data is inserted into a MySQL database version 5.0.51a (-24+ lenny4). The MySQL database is monitored using SQLyog Community GUI v8.6 RC2. MySQL data is more suitable for alternation and calculation, because it allows for regular expressions and macro-commands. This is desirable as the data needs calculation and editing before the analysis can be performed in statistics software. The MySQL database is present on the servers ANNE and ELISA, and is synchronized with one another at the end of the day.

In SQLyog a script adds the tweetid, referred brand, brand industry and data extraction variables. All records are flagged true/false for whether or not it's a retweet, mention or conversation and whether it contains a hashtag. Moreover it calculates daily ratios, incrementals and absolute sums of conversations, hashtags and following. All of the functions and the scripts used to alter the data are included in [appendix 3](#). The complete cell data overview after these modifications and additions can be found in [appendix 7](#).

For the type of data that this study produces, a multi-level analysis design would be most appropriate, as two levels: company tweets (and the effect on) consumer tweets, and possibly a third (company characteristics) can be identified. A design to fit the data and the purpose of this research is the GLLAMM analysis. GLLAMM stands for Generalized Linear Latent And Mixel Models and allows for a multilevel dataset. GLLAMM is developed by Rabe-Hesketh, Skondral & Pickles as an extension for StatCorp Stata (version 6 or greater). After focusing on StatCorp Stata 10 Corporate Edition, the GLLAMM extension and the GLLAMM manual, it was decided in accordance with Methodology Professor P.E.M. Ligthart to drop the GLLAMM analysis as it is outside of the scope of the master thesis objectives.

Instead, the MANOVA (Multivariate Analysis Of Variance) is adopted. The disadvantage of this design is that it does not support multi-level analysis. This means that the company and consumer data has to be aggregated into a single datasheet (i.e. level). The consumer tweet records are kept, and the company tweet data is added as extra columns. Doing so, the standard error is underestimated as those values are blown out of proportions. This led to decision to opt for a large level of confidence of 99% (i.e. p-value of 0.01). This reduces that chance of rejecting the null hypothesis when it is actually true (Type I error). The MANOVA analysis can be performed in both StatCorp Stata and IBM SPSS Statistics. Since IBM SPSS Statistics 19 allows for more data and time

transformation options and because of the researcher's familiarity with the program, IBM SPSS Statistics 19 rather than StatCorp Stata 10 CE is used in this study.

In order to prepare the data so that it fits the program and the MANOVA design, the MySQL database is first converted to a static datasheet since IBM SPSS Statistics 19 can't handle relational databases. The datasheet is then imported to IBM SPSS Statistics 19, ready for further analysis.

3.5 VALIDITY

Internal validity is the extent to which a relationship between two variables isn't explained by extraneous inferences. The selected brands are to have a minimum required level of activity on Twitter. On average, the brands should post at least 10 tweets a week. The advantage of using mechanical real life observational research is that there is no interference from researcher bias. In addition, there are also no respondent biases such as social desirability and strategic answering, as the respondents are unaware of being subject of a research. However, when a real-life situation is observed, the results may be affected by all sorts of extraneous influences. First a control variable is added to control for the growth of the volume of tweets over time. Looking at the development of the body of tweets over time, an incremental growth is visible (see [appendix 1](#)). To control for this growth, a time-based control variable is added. Assuming exponential growth, the study's data shows the body of tweets grew by 27,7% during the research (see [figure 15](#)). Second, a control variable is added to control for external events and news. Influential external events, e.g. CES 2011 or the announcements of the cooperation between Google and Logitech, are tagged over time and inserted in one control variable.

External validity is the extent to which the results of this study can be generalized to the population. Real life observational research is considered to have ultimate level of external validity. In the research setting, respondents are unaware of being subject to a study. Therefore, no biases are present, making the research results highly applicable to the entire population. The largest threat to the external validity for this study is that Twitter doesn't make all of the Twitter data available to the developers, unless you have full firehose contract. However, since Radian6 has full firehose access, it does have access to 100% of the tweets. Another advantage of using Radian6, compared to the pretest, is that new style retweets are also considered, thereby increasing the validity.

The generalizability of influence should be interpreted with care. Since influencing another individual to pass along a piece of information is a rather narrow definition of influence, it may not necessarily imply that brands have influenced consumer opinions or purchasing behavior. Nevertheless, the influence measured in this study is believed to have significant verisimilitude to be useful for marketers. Moreover, previous studies (e.g. Bakshy et al., 2009; Cha et al., 2010) have considered influence similarly to the current study, thus this study is consistent with previous work.

3.6 RELIABILITY

Reliability is the consistency of the measurement instrument. The Twitter API has some limitations, as explained in the previous paragraph. This study overcomes these issues since Radian6 has full firehose access. Another threat to the reliability is missing data due to downtime of the Twitter servers. The downtime of the Twitter during the research period servers was minimal (see [appendix 2](#)). Twitter faced a total downtime of 111 minutes. This means the site had an uptime of 99,89%. Radian6 overcomes this issue using its smart crawling system using multiple servers. The automated sentiment analysis is controlled for manually, for a selection of 600 tweets, to control for its accuracy. The results are included in [appendix 8](#). A pattern similar to that of the results of the sentiment analysis by Tweetfeel Biz in the pretest is visible. Positively characterized tweets tend to be more accurate than tweets characterized as negative. Moreover, the accuracy results have improved over those in the pretest.

A threat to the reliability is inconsistent Twitter activity by the brands in the research sample. When brands are inactive, this disproportionally affects the analysis. Results ([appendix 5](#)) show the

Twitter activity of most brands is rather inconsistent. Most of the brands are not active during weekends; and there are also visible gaps during the week. To control for this issue, a script is executed which fixes these gaps so functions and forthcoming analysis are not disproportionately affected.

The accuracy of measuring all of the other data is guarded as it is all quantitative data, inferred using an objective standard format. There is no manual coding involved and all of the data is computed using automated processes. Concluding, the quantitative observation research is characterized by a relatively high reliability.

CHAPTER 4: RESULTS

4.1 ASSUMPTIONS

In order to execute the MANCOVA and logistic regressions, some assumptions concerning the data have to be met: independence, normality, and absence of multicollinearity and outliers. In the research model, independent variables do not have direct relations on each other. Consistent with previous studies (Cha et al, 2010; Kwat et al, 2010; Weng et al., 2010) some variables are skewed: the kurtosis threshold was exceeded for incremental followers & following, as well as for the absolute daily number of conversations and hashtags. For incremental followers some extreme outliers were removed in order to their disproportionate influence on the analysis. Next the skewed variables are log-transformed.

Figure 10: Descriptive statistics

	Following	Sum of mentions	Retweets	Followers	Sentiment	Conversations	Hashtags
N Valid	231478	231478	231478	229624	49653	231478	231478
N Missing	0	0	0	1854	181825	0	0
Mean	1,666	388,8	,17	2,754	,73	1,531	,939
Median	1,623	262	,00	2,707	1,00	1,386	,693
Std. Dev.	,137	360,97	,374	,169	,443	1,374	,957
Variance	,019	130299,16	,140	,029	,196	1,889	,916
Skewness	1,14	1,61	1,772	1,777	-1,046	,731	,618
Kurtosis	31,034	2,549	1,141	4,426	-,907	-,046	-,781

The missing values for the variable followers are the removed outliers. The variable sentiment also has a high number of missing values; this is due to the measurement instrument which cannot determine negative or positive sentiment for every tweet. The kurtosis of followers and following still exceed the kurtosis threshold. However, with such a large sample size, the effect of this violation is minimal.

Figure 11: Collinearity statistics for dependent variables

Dependent variables	Tolerance	VIF
Mentions	,780	1,282
Incremental followers	,788	1,268
Retweets	,978	1,023
Sentiment	,969	1,032

Figure 12: Collinearity statistics for independent variables

Independent variables	Tolerance	VIF
Incremental following	,942	1,062
Conversation	,931	1,075
Hashtags	,960	1,042
Sentiment	,991	1,009
External influences	,988	1,012
Day of research	,987	1,013

The collinearity statistics show little multicollinearity among both independent and dependent variables.

During the pre-analysis phase it was also found that three brands (Hertz, Canon, AMP Energy) published very little tweets, therefore they have been removed from the analysis.

4.2 ANALYSIS

In order to test the effect on the ratio variables, the MANCOVA is executed.

Figure 13: MANCOVA results

Source	Dependent Variable	F-value	Sig.	Partial Eta Squared	Observed Power
Conversation	Sum of mentions	44305,187	,000	,162	1,000
Hashtag	Sum of mentions	4077,714	,000	,017	1,000
Following	Sum of mentions	1885,020	,000	,008	1,000
Conversation	Followers	39453,319	,000	,147	1,000
Following	Followers	25627,271	,000	,100	1,000
Hashtag	Followers	4985,260	,000	,021	1,000
Covariates:					
<i>External influences</i>	<i>Followers</i>	<i>1838,749</i>	<i>,000</i>	<i>,008</i>	<i>1,000</i>
<i>Day of research</i>	<i>Followers</i>	<i>1165,812</i>	<i>,000</i>	<i>,005</i>	<i>1,000</i>
<i>External influences</i>	<i>Sum of mentions</i>	<i>9057,550</i>	<i>,000</i>	<i>,038</i>	<i>1,000</i>
<i>Day of research</i>	<i>Sum of mentions</i>	<i>2776,459</i>	<i>,000</i>	<i>,012</i>	<i>1,000</i>

It is evident that all effects are significant under the 99% confidence level. The effects are listed from strongest to weakest. As can be seen in the results of the analysis, the effect of the number of conversations is the largest. The effect of following is the smallest. The number of conversations also has the largest effect on the dependent variable followers. However, following also has a very strong effect on followers. The effect of hashtags on both dependent variables is significant, but not very large. The observed power, or P of correctly rejecting the null hypothesis, is 1,000 for all effects. Both covariates also correlate significantly with both dependent variables. Both variables are good covariates because they correlate with the dependent variables, and share little variance with the other independent variables. The explained variance of the dependent variable mentions is 27.6%, and adjusted R squared of followers is 33.8%.

Figure 14: Binary logistic regression for Retweets and Sentiment

	Retweets			Sentiment		
	B	Sig.	Exp(B)	B	Sig.	Exp(B)
Conversation	,009	,039	1,009	,033	,000	1,034
Hashtag	,031	,000	1,031	,065	,000	1,067
Following	-,763	,000	,466	-,523	,000	,593
<i>External influences</i>	<i>,068</i>	<i>,000</i>	<i>1,071</i>	<i>-,191</i>	<i>,000</i>	<i>,826</i>
<i>Day of research</i>	<i>-,005</i>	<i>,000</i>	<i>,995</i>	<i>-,009</i>	<i>,000</i>	<i>,991</i>

For the logistic regression of the dependent variable retweets, all effects but conversation are significant. The explained variance is also relatively low, 0.6%. The effect of conversation and hashtag is relatively low, conversation is even insignificant. The effect of following on both retweets and sentiment is large and, contrary to the research expectations, negative. Conversation and hashtag correlate more with sentiment. Both effects are positive and significant, although not very large. The explained variance for sentiment is 1.4%.

Independent variable following shows mixed results across the dependent variables. Its effect on mentions is a relatively small. However, its effect on followers is large. Moreover, it negatively correlates with both retweets and sentiment. The character of these results supports the notion of “the million follower fallacy” (Avnit, 2009). This theory argues that the follower indegree alone explains little about influence due to the fact some users follow back others simply because of etiquette. During the rise of twitter standards have emerged. Among some users it became polite to

follow someone back. This etiquette is leveraged by some users to elevate their follower indegree. The results show that the amount of users which a brand follows, is mainly valid to predict followers, therefore, the results support the notion of the million follower fallacy. The hashtag indegree has a moderate effect on mentions and followers, and a small effect on the amount of retweets and sentiment, however, all significant. The power of one to one communication for predicting influence been confirmed. It shows a great effect on the amount of mentions, the amount of followers, and a small effect on sentiment. Its effect on retweets, however, was insignificant. The model was able to explain a significant amount of variance of both followers and mentions. Retweets and sentiment both have little variance explained. This means that there are other important variable which explain these variables. The relatively low explained variance is nevertheless somewhat surprising, as previous work and theory expected that the strategies would show differences in their effect on both independent variables. This divergence may be due to the face that work tends to focus on observed success, i.e. research samples are biased towards observed success. When the larger number of non-successful events are also included, it may become difficult to identify the proposed relations.

CHAPTER 5: CONCLUSIONS

5.1 CONCLUSIONS

The research question to be answered in this study was:

What Twitter strategies are most effective in increasing influence on Twitter, and how should influence be measured?

The current study tried to find an answer on this question by investigating the effect of Twitter strategies suggested by professional literature on influence. Furthermore, various measures of influence were evaluated: mentions indegree, followers indegree, sentiment and retweet indegree. In order to test these effects, over 250.000 tweets produced by brands and consumers were gathered during a 10 week research period using social media monitoring tool Radian6. Brands' influence was tested by measuring the effect of Twitter strategies, inferred from the tweets of 30 brands from various industries, on consumer word of mouth.

The biggest contribution of this study to research in this field is that investigates the relationship between brands' Twitter strategies as condition, and its influence on consumer word of mouth. The study has shown that one to one communication, listening to consumers and community participation all significantly influence consumer word of mouth. One to one communication shows the largest effect across influence measures, although its effect on retweet indegree is insignificant. This highlights the important of consumer engagement in order to increase a brand's influence.

The effect of community participation is also significant as well as listening to consumers, although its effect on retweet indegree and sentiment is negative. The amount of following has the greatest effect on followers indegree. This shows that relationships on Twitter may be more reciprocal than how relationships on Twitter are generally presented. Moreover, these results support the notion of the "million follower fallacy" which assumes people make use of etiquette to elevate their followers indegree.

Moreover, the study contributes to research in this field in that it found adequate measures of influence. The study shows that all dependent variables practice strong correlations among each other. It therefore partly supports results of recent influence studies by Kwak et al., Cha et al. and Weng et al., all published in 2010 during the execution of this study. Inherently the study only partially supports the notion of the "million follower fallacy", since followers indegree shows strong correlations with the other measures of influence, while the theory suggests reciprocity distorts the relationship between followers and measures of influence.

5.2 LIMITATIONS

Previous studies and theory are conflicting; using followers indegree as an indicator of influence is suggested by various professional literature and is the most widely used measure. On the other hand, the theory of "the million follower fallacy" as well as some researchers (Cha et al., 2010; Kwak et al., 2010; Weng et al., 2010) suggest followers indegree is highly susceptible to distortion due to users elevating the followers indegree using etiquette. This study cannot clear this conflict; it only partially supports the notion of "the million follower fallacy".

While the study was able to explain a large part of the variance of mentions and followers indegree, there's still a great portion of variance unexplained of the dependent variables sentiment and retweet indegree. Conclusions and generalizability of these results have to be treated with caution.

Real life observational research is generally considered to have the ultimate level of generalizability. Inherent to the type of research, is the influence of external events. In order to cope with some external influences, the study has incorporated the growth of the body tweets during the execution of the study and external events as control variables. However, there still might be other external influences which are not controlled for in the study.

Sentiment analysis still is a rather novel instrument. Although a human check of its positive and negative judgments showed rather good results, the tool flagged a large amount of tweets as neutral. Some of these were truly neutral, others were flagged neutral because the tool was unable to identify its sentiment as positive or negative. Sentiment analyses in general are more capable of identifying positive sentiment than they are at identifying negative sentiment. Aforementioned issues are likely to have caused the rather low explained variance of sentiment. As technology evolves and stronger algorithms are developed, future sentiment analyses will be more accurate. Hence future studies might be more capable of predicting sentiment.

5.3 DIRECTIONS FOR FUTURE RESEARCH

Since the research is of quantitative nature it is able to explain such a large portion of data. Inherent to quantitative research is that it is less suitable for explaining the context. It might be fruitful to combine quantitative research with a qualitative study, such as a case study, in which underlying structures for successfully influencing consumer tweets are inferred. Moreover, qualitative studies might investigate to what extent the influence measures translate into positive growth of traditional marketing performance measures such as ROI, image, satisfaction and share of wallet.

Personal characteristics or brand values as predictors of influence were outside the scope of the current study. Future researches in this field may attempt to predict influence as a function of personal characteristics or brand values. These factors may be alike those proposed by Breakenridge (2011); e.g. trust, charisma, knowledge and expertise and topic passion. Combining this with quantitative approach will be difficult however, since determining qualitative features for large populations is practically infeasible. This is stressed by Watts & Dodds (2007), who argue that there are so many kinds of influentials, that it is practically infeasible to generalize characteristics across settings.

Investigating Twitter data is still in an early phase. More professional research instruments will be more capable of investigating relations in a direct manner, and calculating more complicated ratios, e.g. calculating true reach by retrieving followers on all levels, or discriminating between levels of influence of people in the brands' network.

In discussions with professionals about this study, it was noticed that there's a great desire from managers to determine influential users. Most social media monitoring tools determine influential users using a ratio which consists of the user's followers indegree multiplied by the amount of messages in which the user mentions the brands. Brands however demand a more professional influencer analysis. There lies a great opportunity in developing a more professional index of user influence. An index which incorporates e.g. reach, engagement, amplification and linguistics (e.g. sentiment).

5.4 IMPLICATIONS

The current study shows that brands' Twitter strategies positively influence consumer word of mouth. It highlights the importance of listening to consumers, one to one communication and community participation. Moreover, it shows that a focus on following consumers is primarily used to elevate the followers indegree. This study scientifically investigated strategies suggested by professional literature. While an increasing amount of brands is participating in social media, it is evident for them to understand what strategies might be used to increase their influence on consumers. The growing amount of people interacting online only stresses the need of brands developing an online presence on social media, thereby increasing the need for knowledge on influence.

To determine the return of their social media activities, management demands measurements of the brands' influence. This study shows four indicators (follower indegree, mentions indegree, sentiment and retweet indegree) brands can use to measure their influence on consumer word of mouth. The results of this study will assist managers in quantifying influence on

Twitter, thereby helping brands with measuring their online activities and reporting back to the management. It is noted that in certain aspects Twitter is a special case; nevertheless the observations are likely to apply in other contexts as well.

REFERENCES

- Anderson, E.W. (1998). "Customer satisfaction and word of mouth", *Journal of Service Research*, vol. 1, no. 1, pg. 5–17.
- Anderson, E.W. & Salisbury, L.C. (2003). "The formation of marketlevel expectations and its Covariates", *Journal of Consumer Research*, vol. 30, no. 1, pg. 115–124.
- Aral, S., Muchnik, L. & Sundararajan, A. (2009). "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks", *Proceedings of the National Academy of Sciences*.
- Arndt, J. (1967). "Role of product-related conversations in the diffusion of a new product", *Journal of Marketing Research*, vol. 4, pg. 291 – 295.
- Avnit, A. (2009). "The Million Followers Fallacy", *Pravda Media*, Retrieved from <http://pravdam.co/...>
- Bampo, M., Ewing, M.T., Mather, D.R., Stewart, D. & Wallace, M. (2008). "The Effects of the Social Structure of Digital Networks on Viral Marketing Performance", *Information Systems Research*, vol. 19, no. 3, pg. 273–290.
- Bakshy, E., Hofman, J.M., Mason, W.A. & Watts, D.J. (2011). "Everyone's an influencer: Quantifying influence on Twitter", *Proceedings of the fourth ACM international conference on Web search and data mining*, New York: ACM.
- Bakshy, E., Karrer, B. & L.A. Adamic. (2009). "Social influence and the diffusion of user-created content", *Proceedings of the 10th ACM Conference on Electronic Commerce*, Stanford, CA: Association of Computing Machinery.
- Bansal, H.S. & Voyer, P.A. (2000). "Word of mouth processes within a services purchase decision context", *Journal of Service Research*, vol. 3, no. 2, pg. 166–177.
- Barton, B. (2006). "Ratings, reviews & ROI: How leading retailers use customer word of mouth in marketing and merchandising", *Journal of Advertising Research*, vol. 7, no. 87.
- Bickart, B. & Schindler, R.M. (2001). "Internet forums as influential sources of consumer information", *Journal of Interactive Marketing*, vol. 15, no. 3, pg. 31 – 40.
- Blom, E. (2009). "Handboek Gemeenschap: De Kracht Van Sociale Netwerken", *Utrecht, The Netherlands: AW Bruna*.
- Bollen, J., Mao, H. & Zeng, X.J. (2010). "Twitter mood predicts the stock market", *Proceedings of the Alife XII Conference*, Cambridge, MA: MIT Press.
- Breakenridge, D. (2011). "The Influencer & The Give Factor", *Deirdre Breakenridge Blog*, retrieved from <http://www.deirdrebreakenridge.com/2011/03/the-influencer-the-give-factor>.
- Cha, M., Haddadi, H., Benevenuto, F. & Gummadi, K.P. (2010). "Measuring user influence in twitter: The million follower fallacy", *Proceedings of the 4th International Conference on Weblogs and Social Media*, Menlo Park, CA: The AAAI Press.
- Chen, Y.B. & Xie, J.H. (2008). "Online consumer review: Word-of-mouth as a new element of the marketing communication mix", *Management Science*, vol. 54, no. 3, pg. 477.
- Chevalier, J. & Mayzlin, D. (2004). "The Effect of Word of Mouth on Sales: Online Book Reviews", *Journal of Marketing Research*, vol. 43, no. 3, pg. 345–354.
- Comm, J. & Burge, K. (2009). "Twitter Power: How to Dominate Your Market One Tweet at a Time", *Hoboken, New Jersey: John Wiley & Sons*.
- Costa, C., Beham, G., Reinhardt, W. & Sillaots, M. (2008). "Microblogging In Technology Enhanced Learning: A Use-Case Inspection of PPE Summer School 2008", *Workshop at the European Conference on Technology Enhanced Learning (ECTEL) 2008*, Maastricht, The Netherlands, September 16-19.
- Danaher, P.J. & Rust, R.T. (1996). "Indirect financial benefits from service quality", *Quality Management Journal*, vol. 3, no. 2, pg. 63–75.
- Davenport, T. H. and Beck, J. C. (2002). "The Attention Economy: Understanding the New Currency of Business", *Cambridge, Massachusetts: Harvard Business Press*.

- Davis, A., & Khazanchi, D. (2008). "An empirical study of online word of mouth as a predictor for multi-product category e-Commerce sales", *Electronic Markets*, vol. 18, no. 2, pg. 130–141.
- De Bruyn, A. & Lilien, G.L. (2008). "A multi-stage model of word-of-mouth influence through viral marketing", *International Journal of Research in Marketing*, vol. 25, no. 3, pg. 151-163.
- Derksen, M. (2011). "Hoe bepaal je de strategie van een succesvolle netwerkorganisatie?", *Upstream Blog*, retrieved from http://www.upstream.nl/blog/bericht/hoe_bepaal_je_de_strategie...
- De Rooy, C. (2009). "Listen to your heart or to your hard drive?", *Unpublished Thesis*, University of Nijmegen.
- De Valck, K., van Bruggen, G.H. & Wierenga, B. (2009). "Virtual communities: A marketing perspective", *Decision Support Systems*, vol. 47, no. 3, pg. 185-203.
- Domingos, P. & Richardson, M. (2001). "Mining the network value of customers", *In Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining*. New York: ACM, pg. 57–66.
- Duana, W., Gub, B. & Whinston, A.B. (2008). "Do online reviews matter?— An empirical investigation of panel data". *Decision Support Systems*, vol. 45, no. 3, pg. 1007–1016.
- Dwayne, D. (2004). "Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the Internet?", *Journal of Interactive Marketing*, vol. 18, no. 1, pg. 38 – 52.
- Ebner, M. & Schiefner, M. (2008). Microblogging - more than fun?, *Proceeding of IADIS Mobile Learning Conference 2008*, Inmaculada Arnedillo Sánchez and Pedro Isaías ed., Algarve, Portugal, 2008, p. 155-159.
- Engel, J.E., Blackwell, R.G., & Kegerreis, R.J. (1969). "How information is used to adopt an innovation". *Journal of Advertising Research*, vol. 9, pg. 3-8.
- Feldman, S.P. & Spencer, M.C. (1965). "The effect of personal influence in the selection of consumer services", In Peter D. Bennett (Ed.), *Fall Conference of the American Marketing Association Chicago: American Marketing Association*, pg. 412-435.
- Festinger, L. (1957). "A theory of cognitive dissonance". *Stanford, California: Stanford University Press*.
- Garrett, D. E. (1987). "The effectiveness of marketing policy boycotts: Environmental opposition to marketing", *Journal of Marketing*, vol. 51, no. 2, pg. 46-57.
- Gelb, B.D., & Sundaram, S. (2002). "Adapting to 'word of mouse'." *Business Horizons*, vol. 45, no. 4, pg. 21–25.
- Gladwell, M. (2002). "The tipping point: how little things can make a big difference", *New York City, New York: Little, Brown and Company Inc.*
- Godin, S. (1999). "Permission Marketing: turning strangers into friends, and friends into customers", *New York City, New York: Simon & Schuster*.
- Goldenberg, J., Libai, B., Moldovan, S. & Muller, E. (2007). "The NPV of bad news", *International Journal of Research in Marketing*, vol. 24, no. 3, pg. 186-200.
- Goyal, A., Bonchi, F. & Lakshmanan, L.V.S. (2008). "Discovering leaders from community actions. Pages", *Proceeding of the 17th ACM conference on Information and knowledge management*, pg. 499–508. New York: ACM.
- Guyot, P. (2010). "Half of messages on Twitter are not in English Japanese is the second most used language", *Semiocast Press Release*, Paris, France.
- Hagel, J. & Armstrong, A.G. (1997). "Net gain: expanding markets through virtual communities", Boston, MA' Harvard Business School Press.
- Hennig-Thurau, T., Gwinner K.P., Walsh G., & Gremler D.D. (2004). "Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the Internet?", *Journal of Interactive Marketing*, vol. 18, no. 1, pg. 38–52.
- Herr, P.M., Kardes, F.R., & Kim, J. (1991). "Effects of word-of-mouth and product-attribute information on persuasion: an accessibility-diagnostics perspective", *Journal of Consumer Research*, vol. 17, no.4, pg. 454-62.

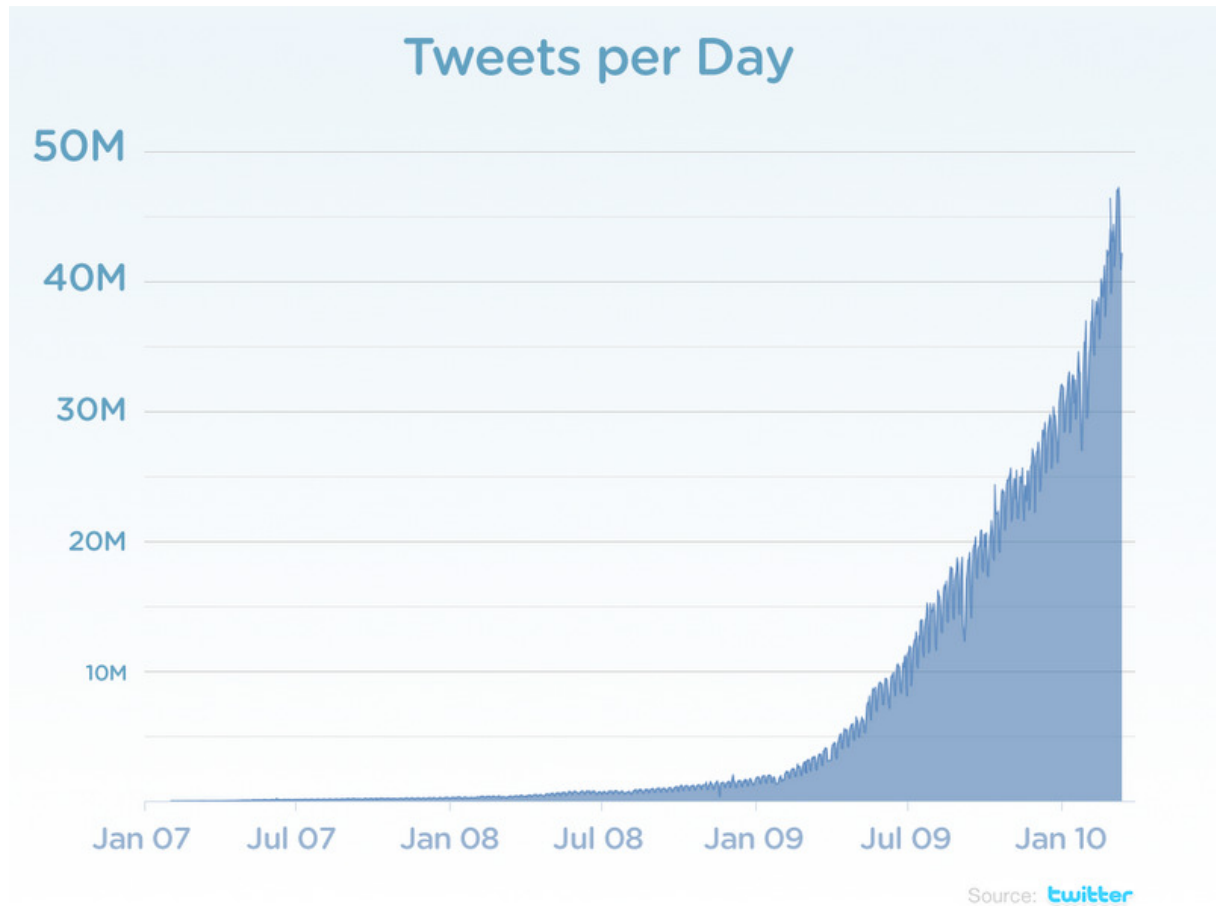
- Honeycutt, C. & Herring, S.C. (2009). "Beyond Microblogging: Conversation and Collaboration via Twitter", *Proceedings of the 42nd Hawaii International Conference on System Sciences*, IEEE Press.
- Huberman, B.A. & Asur, S. (2010). "Predicting the Future With Social Media", *Whitepaper*, Palo Alto, California: HP Labs.
- Huberman, B.A., Romero, D.M. & Wu, F. (2008). "Social Networks That Matter: Twitter Under The Microscope", *First Monday*, vol. 14, no. 1, pg. 1-5.
- Jansen, B.J. & Zhang, M. (2009). "Twitter Power: Tweets as Electronic Word of Mouth", *Journal of the American Society for Information Science and Technology*, vol. 60, no. 11, pg. 2169-2188.
- Java, A., Song, X., Finin, T. & Tseng, B. (2007). "Why We Twitter: Understanding Microblogging Usage and Communities", *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop*.
- Kaikati, A., & Kaikati, J. (2004). "Stealth marketing: How to reach consumers surreptitiously", *California Management Rev*, vol. 46, no. 4, pg. 6–22.
- Katz, E. & Lazarsfeld, P.F. (1955). "Personal influence", *Glencoe, IL*: Free Press.
- Kaplan, A.M. & Haenlein, M. (2010). "Users of the world, unite! The challenges and opportunities of social media", *Business Horizons*, vol. 53, no. 1, pg. 59-68.
- Kaplan, A.M. & Haenlein, M. (2011). "The early bird catches the news: Nine things you should know about micro-blogging", *Business Horizons*, vol. 54, no. 2, pg. 105-113.
- Keller, E. (2005). "The state of word of mouth, 2005: The consumer perspective", *Conference paper presented at the Word-of-Mouth Marketing Association Summit Chicago*, March 29-30.
- Keller, E. & Berry, J. (2003). "The Influentials", Free Press.
- Kelly, K., in Wijnberg, R., Deters, M., de Putter, J. & Hagen, H. (2011). "Backlight: The post-WikiLeaks world", *VPRO Tegenlicht*, retrieved from <http://tegenlicht.vpro.nl/afleveringen/2010-20...>
- Kerkhofs, S., Beerthuyzen, M., Planken, M. & Adriaanse, J. (2010). "Social Marketing Strategie", Tilburg, The Netherlands: Pondres BV.
- Kiecker, P., & Cowles, D.L. (2001). "Interpersonal communication and personal influence on the Internet: A framework for examining online word-of-mouth", *Journal of Euromarketing*, vol. 11, no. 2, pg. 71–88.
- Kirby, J. & Marsden, P. (2006). "Connected Marketing: The Viral, Buzz and Word of Mouth Revolution", Oxford, United Kingdom: Butterworth-Heinemann.
- Kiss, C. & Bichler, M. (2008). "Identification of influencers - Measuring influence in customer networks", *Decision Support Systems*, vol. 46, no. 1, pg. 233-253.
- Kwak, H., Lee, C., Park, H. & Moon, S. (2010). "What is Twitter, a social network or a news media?", *Proceedings of the Nineteenth International WWW Conference*, pg. 591–600. New York: ACM.
- Lacy, K. (2010). "Twitter Marketing for Dummies", *Indianapolis, Indiana: Wiley Publishing Inc*.
- Lerman, K. (2007). "Social information processing in news aggregation", *IEEE Internet Computing*, vol. 11, no. 6, pg. 16–28.
- Lerman, K. & Jones, L.A. (2007). "Social browsing on flickr", *Proceedings of the International Conference on Weblogs and Social Media*, Boulder, CO.
- Locke, C., Levine, R., Weinberger, D. & Searls, D. (2000). "The Cluetrain Manifesto: The End of Business As Usual", Cambridge, MA: Perseus.
- Maglana, M.S. (2007). "Adopting a Service-Dominant Logic in Marketing", *Technopreneurship Notes*, Retrieved from <http://technopreneurship.wordpress.com/2007/03/15/adopting-a-service-...>
- Nardi, B.A., Schiano, D.J., Gumbrecht, M., & Swartz, L. (2004). "Why we blog", *Communications of the ACM*, vol. 47, no. 12, pg. 41–46.
- Pang, B & Lee, L. (2008). "Opinion mining and sentiment analysis", *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pg. 1–135.
- Peppers, D., Rogers, M. & Dorf, R. (1999). "Is your company ready for one-to-one marketing?", *Harvard Business Review*, vol. 77, no. 1, pg. 151–160.
- Pine, B.J., Peppers, D., & Rogers, M. (1995). "Do You Want to Keep Your Customers Forever?", *Harvard Business Review*, vol. 73, no. 2, pg. 103-114.

- Rabe-Hesketh, S., Skondral, A. & Pickles, A. (2004). "GLLAMM Manual", *U.C. Berkeley Division of Biostatistics Working Paper Series, Berkeley, CA: University of California*.
- Raman, D. (2010). "Twitteren Met Resultaat", *Netherlands, Haarlem: Easy Computing*.
- Reichheld, F.F. (2003). "The one number you need to grow", *Harvard Business Review*, vol. 81, no. 12, pg. 46–54.
- Richins, M.L. (1983). "Negative word-of-mouth by dissatisfied consumers: a pilot study", *Journal of Marketing*, vol. 47, pg. 68–78.
- Richins, M.L., & Root-Shaffer, T. (1988). "The role of involvement and opinion leadership in consumer word-of-mouth: An implicit model made explicit", *Advances in Consumer Research*, vol. 15, pg. 32–36.
- Rosenbloom, S. (2007). "On Facebook, Scholars Link Up With Data", *New York Times*, Retrieved from <http://www.nytimes.com/2007/12/17/style/17facebook.html>.
- Shannon, C. E., & Weaver, W. (1949). "The mathematical theory of communication", *Urbana, Illinois: University of Illinois Press*.
- Simon, H. A. (1971). "Designing Organizations for an Information-Rich World", *Martin Greenberger, Computers, Communication, and the Public Interest*, The Johns Hopkins Press.
- Sun, E.S., Rosenn, I., Marlow, C.A. & Lento, T.M. (2009). "Gesundheit! Modelling contagion through Facebook news feed", *Proceedings of the 3rd International Conference on Weblogs and Social Media*, San Jose, CA: The AAAI Press.
- Thompson, S.A. & Sinha, R.K. (2008). "Brand Communities and New Product Adoption: The Influence and Limits of Oppositional Loyalty", *Journal of Marketing*, vol. 72, no. 6, pg. 65–80.
- Trusov, M., Bucklin, R.E. & Pauwels, K. (2009). "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site", *Journal of Marketing*, vol. 73, no. 5, pg. 90-102.
- Vargo, S.L. & Lusch, R.F. (2004). "Evolving To A New Dominant Logic For Marketing", *Journal of Marketing*, vol. 68, no. 1, pg. 1-17.
- Van Veelen, I., Denters, M., Kieft, M., de Bruijn, W., van den Berg, J., de Putter, J. & Romeijn. (2008). "Backlight: What happened to the truth", *VPRO Tegenlicht*, retrieved from <http://tegenlicht...>
- Ward, J.C. & Ostrom, A.L. (2006). "Complaining to the masses: The role of protest framing in customer-created complaint web sites", *Journal of Consumer Research*, vol. 33, no. 2, pg. 220-230.
- Watts, D.J. (2002). "A simple model of global cascades on random networks", *Proceedings of the National Academy Of Sciences*, vol. 99, pg. 5766–5771.
- Watts, D.J. (2007). "The accidental influential", *Harvard Business Review*, vol. 85, no. 2, pg. 22-23.
- Watts, D.J. & Dodds, P. (2007). "Influential, networks, and public opinion formation", *Journal of Consumer Research*, vol. 34, no. 4, pg. 441-458.
- Watts, D.J., Frumin, M. & Peretti, J. (2007). "Viral Marketing for the Real World", *Harvard Business Review*, vol. 85, no. 5, pg. 22–23.
- Weng, J., Lim, E.P., Jiang, J. & He, Q. (2010). "TwitterRank: Finding Topic-sensitive Influential Twitterers", *Proceedings of the third ACM international conference on Web search and data mining*, February 4–6, 2010, New York: ACM.
- Werry, C. C. (1996). "Linguistic and interactional features of Internet Relay Chat. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and crosscultural perspectives*, pg. 47-63, Amsterdam, Netherlands: John Benjamins.
- Whyte, W.H. (1954). "The Web of word of mouth", *Fortune*, vol. 50, pg. 140-143.
- Zeithaml, V.A., & Bitner, M.J. (1996). *Services marketing*. New York City, New York: McGraw-Hill.

APPENDIX

APPENDIX 1: AMOUNT OF TWEETS PER DAY, BY TWITTER INC.

Obtained February the 22th, 2010 from <http://blog.twitter.com/2010/02/measuring-tweets.html>



APPENDIX 2: TWITTER DOWNTIME PER MONTH, BY PINGDOM

Obtained February the 1st, 2010 from http://www.pingdom.com/reports/wx4vra365911/check_ov...

Month	Uptime (%)	Downtime (d, h, m)	Avg response time
January 2011	99.92%	21m	926.56 ms
December 2010	99.83%	1 h, 15 m	922.1 ms
November 2010	100%	1 m	983.86 ms
October 2010	99.95%	22 m	946.69 ms
September 2010	99.97%	10 m	845.89 ms
August 2010	99.82%	1 h, 17 m	1107.86 ms
July 2010	99.66%	2 h, 32 m	1238.71 ms
June 2010	98.52%	10 h, 32 m	2523.91 ms
May 2010	99.77%	1 h, 40 m	2281.34 ms
April 2010	99.83%	1 h, 13 m	2179.76 ms
March 2010	99.82%	1 h, 21 m	1314.09 ms
February 2010	99.87%	50 m	887.38 ms
January 2010	99.78%	1 h, 40 m	1000.56 ms
December 2009	99.67%	2 h, 27 m	810.48 ms
November 2009	99.95%	22 m	796.16 ms
October 2009	99.84%	1 h, 10 m	1065.25 ms
September 2009	99.87%	57 m	1767.68 ms
August 2009	99.06%	6 h, 57 m	1540.93 ms
July 2009	99.92%	35 m	1083.68 ms
June 2009	99.86%	59 m	907.46 ms
May 2009	99.51%	3 h, 37 m	991.08 ms
April 2009	99.87%	58 m	995.68 ms
March 2009	99.79%	1 h, 34 m	907.78 ms
February 2009	99.91%	37 m	807.35 ms
January 2009	99.92%	37 m	794.72 ms
December 2008	99.97%	14 m	871.13 ms
November 2008	99.29%	5 h, 8 m	811.79 ms
October 2008	99.77%	1 h, 41 m	976.53 ms
September 2008	99.87%	54 m	911.05 ms
August 2008	99.86%	1 h, 3 m	798.14 ms
July 2008	99.43%	4 h, 12 m	1188.36 ms
June 2008	98.39%	11 h, 36 m	2151.92 ms
May 2008	97.13%	21 h, 22 m	1556.6 ms
April 2008	99%	7 h, 10 m	970.6 ms
March 2008	99.57%	3 h, 12 m	889.4 ms
February 2008	98.09%	13 h, 17 m	1098.2 ms
January 2008	98.17%	13 h, 37 m	1196.67 ms
December 2007	97.87%	15 h, 48 m	1169.74 ms
November 2007	98.74%	9 h, 4 m	1244.23 ms
October 2007	99.02%	7 h, 19 m	1984.19 ms
September 2007	99.06%	6 h, 44 m	2281.15 ms
August 2007	98.47%	11 h, 22 m	2682.29 ms
July 2007	99.59%	3 h, 2 m	1742.24 ms
June 2007	99.08%	6 h, 37 m	1806.91 ms
May 2007	97.67%	17 h, 10 m	2327.11 ms
April 2007	99.11%	6 h, 15 m	1824.76 ms

March 2007	92.09%	2 d, 10 h, 4 m	3591.5 ms
February 2007	98.21%	6 h, 22 m	2070.89 ms
Total average	99.11%	11 d, 9 h, 21 m	1413.22 ms

APPENDIX 3: DATA ALTERATION WITHIN SQLYOG

Set values in the column stats_conversation. When a tweet matches the text string (in this example, HTC), it sets value=1. If false, the value is set to 0.

```
UPDATE tweets tw
SET stats_conversation = 1, stats_conversation_user = 'HTC'
WHERE tw.tweettype = 2 AND
      message RLIKE CONCAT('^@htc([A-Za-z0-9_].*)?$');
```

Set values in the column stats_mention. When a tweet matches the text string (in this example, HTC), it sets value=1. If false, the value is set to 0.

```
UPDATE tweets tw
SET stats_mention = 1, stats_mention_user = 'HTC'
WHERE tw.tweettype = 2 AND tw.message RLIKE CONCAT('^.*@htc([A-Za-z0-9_].*)?$');
```

Set values in the column stats_retweet. When a tweet matches the text string (in this example, HTC), it sets value=1. If false, the value is set to 0.

```
UPDATE tweets tw
SET stats_retweet = 1, stats_retweet_user = 'HTC'
WHERE tw.tweettype = 2 AND
      (
        tw.message LIKE CONCAT('%RT @htc:%')
        OR
        tw.message LIKE CONCAT('%RT @htc %')
        OR
        tw.message LIKE CONCAT('%via @htc')
        OR
        tw.message LIKE CONCAT('%(via @htc)')
      );
```

Set values in the column stats_hashtag. When a tweet matches the text string (in this example, HTC), it sets value=1. If false, the value is set to 0.

```
UPDATE tweets tw
SET stats_hashtag = 1
WHERE tw.message RLIKE '(^.*|^)[A-Za-z0-9_]+.*$'
```

The following script extracts the amount of mentions per tweet. In order to extract this information, first a new function is added to the SQL database which makes the extraction possible. Basically, it searches the tweets for the at character (@), checks whether it is followed by valid characters (so it's actually a mention, not a random sign, which sometimes occurs when people type they're frustrated (e.g. !&@)#\$^)), and counts the amount of mentions.

```
DELIMITER //
DROP FUNCTION IF EXISTS AMOUNT_MENTION//
```

```

CREATE FUNCTION AMOUNT_MENTION (tweetmessage VARCHAR(320))
RETURNS INT
DETERMINISTIC
BEGIN
    DECLARE returnValue INT;
    DECLARE amountAtCharacter INT;
    DECLARE amountAtProcessed INT;
    DECLARE lastPositionAt INT;

    SET returnValue = 0;
    SET amountAtCharacter = LENGTH(tweetmessage) -
LENGTH(REPLACE(tweetmessage, '@', ''));
    SET amountAtProcessed = 0;
    SET lastPositionAt = 1;

    WHILE amountAtCharacter != amountAtProcessed DO
        IF SUBSTRING(tweetmessage, LOCATE('@', tweetmessage,
lastPositionAt)) RLIKE '^@[A-Za-z0-9_].*$' THEN
            SET returnValue = returnValue + 1;
        END IF;
        SET lastPositionAt = LOCATE('@', tweetmessage,
lastPositionAt) + 2;
        SET amountAtProcessed = amountAtProcessed + 1;
    END WHILE;
    RETURN returnValue;
END//
DELIMITER ;

```

Subsequently, the following script executes the previously explained SQL function and inserts the amount of mentions per tweet in a column (stats_mention_amount).

```

UPDATE tweets tw
SET stats_mention_amount = AMOUNT_MENTION(tw.message);

```

Extract the amount of hashtags per tweet. In order to extract this information, first a new function is added to the SQL database which makes the extraction possible. Basically, it search the tweets the hash character (#), checks whether it is followed by valid characters (so it's actually a hashtag, not a random sign, which sometimes occurs when people type they're frustrated (e.g. !&@^#\$^)), and counts the amount of hashtags.

```

DELIMITER //
DROP FUNCTION IF EXISTS AMOUNT_HASHTAG//
CREATE FUNCTION AMOUNT_HASHTAG (tweetmessage VARCHAR(320))
RETURNS INT
DETERMINISTIC
BEGIN
    DECLARE returnValue INT;
    DECLARE amountHashCharacter INT;
    DECLARE amountHashProcessed INT;
    DECLARE lastPositionHash INT;

    SET returnValue = 0;
    SET amountHashCharacter = LENGTH(tweetmessage) -
LENGTH(REPLACE(tweetmessage, '#', ''));
    SET amountHashProcessed = 0;

```

```

SET lastPositionHash = 1;

WHILE amountHashCharacter != amountHashProcessed DO
    IF SUBSTRING(tweetmessage, LOCATE('#', tweetmessage,
lastPositionHash)) RLIKE '^#[A-Za-z0-9_].*$' THEN
        SET returnValue = returnValue + 1;
    END IF;
    SET lastPositionHash = LOCATE('#', tweetmessage,
lastPositionHash) + 2;
    SET amountHashProcessed = amountHashProcessed + 1;
END WHILE;
RETURN returnValue;
END//
DELIMITER ;

```

Subsequently, the following script executes the previously explained SQL function and inserts the amount of mentions per tweet in a column (stats_hashtag_amount).

```

UPDATE tweets tw
SET stats_hashtag_amount = AMOUNT_HASHTAG(tw.message);

```

The next function declares the sum of mentions for a brand during a 24 hour period.

```

DELIMITER //
DROP FUNCTION IF EXISTS SUM_MENTION//
CREATE FUNCTION SUM_MENTION (func_brand VARCHAR(32), func_dateoftweet DECIMAL(20))
RETURNS INT
DETERMINISTIC
BEGIN
    DECLARE returnValue INT;
    SET returnValue = 0;
    IF (func_dateoftweet % 86400) >= 7200 THEN
        SET returnValue = (SELECT COUNT(*) FROM tweets_copy WHERE brand = func_brand AND
dateoftweet
            BETWEEN (func_dateoftweet - (func_dateoftweet % 86400)) + 7200
            AND (func_dateoftweet - (func_dateoftweet % 86400)) + 7200 + 86399);
    ELSE
        SET returnValue = (SELECT COUNT(*) FROM tweets_copy WHERE brand = func_brand AND
dateoftweet
            BETWEEN (func_dateoftweet - (func_dateoftweet % 86400)) + 7200 - 86400
            AND (func_dateoftweet - (func_dateoftweet % 86400)) + 7200 + 86399 - 86400);
    END IF;
    RETURN returnValue;
END//
DELIMITER ;

```

The function is then called by the following script, which sets the actual values for the sum_mentions variable.

```

UPDATE tweets AS tw
SET tw.sum_mentions = SUM_MENTION(tw.brand, tw.dateoftweet / 1000)
WHERE tw.tweettype = 2;

```

This function calculates the incremental amount of following for a brand during the 24 hour timespan.

```
DELIMITER //
DROP PROCEDURE IF EXISTS INCR_FOLLOWING//
CREATE PROCEDURE INCR_FOLLOWING()
BEGIN
    DECLARE done BOOLEAN DEFAULT 0;
    DECLARE starttime DECIMAL;
    DECLARE endtime DECIMAL;
    DECLARE minfollowing INT;
    DECLARE maxfollowing INT;
    DECLARE currentbrand VARCHAR(32);
    DECLARE cursor_brand CURSOR FOR SELECT username FROM tweets WHERE tweettype = 1
GROUP BY username;
    -- Declare continue handler
    DECLARE CONTINUE HANDLER FOR SQLSTATE '02000' SET done=1;
    UPDATE tweets SET incr_following = 0;
    SET starttime = (SELECT MIN(dateoftweet) / 1000 FROM tweets);
    IF starttime % 86400 >= 7200 THEN
        SET starttime = starttime - (starttime % 86400);
    ELSE
        SET starttime = starttime - (starttime % 86400) - 86400;
    END IF;
    SET endtime = (SELECT MAX(dateoftweet) / 1000 FROM tweets);
    IF endtime % 86400 >= 7200 THEN
        SET endtime = endtime - (endtime % 86400);
    ELSE
        SET endtime = endtime - (endtime % 86400) - 86400;
    END IF;
    SELECT starttime, endtime;
    WHILE starttime != endtime DO
        SET done = 0;
        OPEN cursor_brand;
        -- Loop through all rows
        REPEAT
            -- Get order number
            FETCH cursor_brand INTO currentbrand;
            SET minfollowing = (SELECT following FROM tweets WHERE username = currentbrand AND
(((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) = starttime ORDER BY
dateoftweet LIMIT 1);
            IF minfollowing IS NOT NULL THEN
                SET maxfollowing = (SELECT following FROM tweets WHERE username = currentbrand AND
(((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) = starttime ORDER BY
dateoftweet DESC LIMIT 1);
                SELECT CONCAT("[", starttime, "] - Found ", minfollowing, "-", maxfollowing, " for brand: ",
currentbrand) AS STATUS;
                UPDATE tweets SET incr_following = maxfollowing - minfollowing WHERE tweettype = 2
AND brand = currentbrand AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) %
86400)) = starttime;
            ELSE
                SELECT CONCAT("[", starttime, "] - No tweets for brand: ", currentbrand) AS STATUS;
            END IF;
        UNTIL done
        END REPEAT;
    END WHILE;
END
```



```

        END IF;
    -- End of loop
    UNTIL done END REPEAT;
    CLOSE cursor_brand;
    -- go to the next day
    SET starttime = starttime + 86400;
    END WHILE;
END//
DELIMITER ;

```

This function calculates the incremental amount of followers for a brand during the 24 hour timespan.

```

DELIMITER //
DROP PROCEDURE IF EXISTS INCR_FOLLOWERS//
CREATE PROCEDURE INCR_FOLLOWERS()
BEGIN
    DECLARE done BOOLEAN DEFAULT 0;
    DECLARE starttime DECIMAL;
    DECLARE endtime DECIMAL;
    DECLARE minfollowers INT;
    DECLARE maxfollowers INT;
    DECLARE currentbrand VARCHAR(32);
    DECLARE cursor_brand CURSOR FOR SELECT username FROM tweets WHERE tweettype = 1
    GROUP BY username;
    -- Declare continue handler
    DECLARE CONTINUE HANDLER FOR SQLSTATE '02000' SET done=1;
    UPDATE tweets SET incr_followers = 0;
    SET starttime = (SELECT MIN(dateoftweet) / 1000 FROM tweets);
    IF starttime % 86400 >= 7200 THEN
        SET starttime = starttime - (starttime % 86400);
    ELSE
        SET starttime = starttime - (starttime % 86400) - 86400;
    END IF;
    SET endtime = (SELECT MAX(dateoftweet) / 1000 FROM tweets);
    IF endtime % 86400 >= 7200 THEN
        SET endtime = endtime - (endtime % 86400);
    ELSE
        SET endtime = endtime - (endtime % 86400) - 86400;
    END IF;
    SELECT starttime, endtime;
    WHILE starttime != endtime DO
        SET done = 0;
        OPEN cursor_brand;
        -- Loop through all rows
        REPEAT
            -- Get order number
            FETCH cursor_brand INTO currentbrand;
            SET minfollowers = (SELECT followers FROM tweets WHERE username = currentbrand AND
            (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) = starttime ORDER BY
            dateoftweet LIMIT 1);
            IF minfollowers IS NOT NULL THEN

```

```

        SET maxfollowers = (SELECT followers FROM tweets WHERE username = currentbrand AND
        (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) = starttime ORDER BY
        dateoftweet DESC LIMIT 1);
        SELECT CONCAT("[", starttime, "] - Found ", minfollowers, "-", maxfollowers, " for brand: ",
        currentbrand) AS STATUS;
        UPDATE tweets SET incr_followers = maxfollowers - minfollowers WHERE tweettype = 2
        AND brand = currentbrand AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) %
        86400)) = starttime;
        ELSE
        SELECT CONCAT("[", starttime, "] - No tweets for brand: ", currentbrand) AS STATUS;
        END IF;
    -- End of loop
    UNTIL done END REPEAT;
    CLOSE cursor_brand;
    -- go to the next day
    SET starttime = starttime + 86400;
END WHILE;
END//
DELIMITER ;

```

The following three functions calculate the brand ratio's for conversation, hashtags and mentions related to the total brand tweets within the 24 hour timespan.

```

DELIMITER //
DROP PROCEDURE IF EXISTS RATIO_CONVERSATIONS//
CREATE PROCEDURE RATIO_CONVERSATIONS()
BEGIN
    DECLARE done BOOLEAN DEFAULT 0;
    DECLARE starttime DECIMAL;
    DECLARE endtime DECIMAL;
    DECLARE totalconversations INT;
    DECLARE totaltweets INT;
    DECLARE currentbrand VARCHAR(32);
    DECLARE cursor_brand CURSOR FOR SELECT username FROM tweets WHERE tweettype = 1
    GROUP BY username;
    -- Declare continue handler
    DECLARE CONTINUE HANDLER FOR SQLSTATE '02000' SET done=1;
    UPDATE tweets SET ratio_conversation = 0;
    SET starttime = (SELECT MIN(dateoftweet) / 1000 FROM tweets);
    IF starttime % 86400 >= 7200 THEN
        SET starttime = starttime - (starttime % 86400);
    ELSE
        SET starttime = starttime - (starttime % 86400) - 86400;
    END IF;
    SET endtime = (SELECT MAX(dateoftweet) / 1000 FROM tweets);
    IF endtime % 86400 >= 7200 THEN
        SET endtime = endtime - (endtime % 86400);
    ELSE
        SET endtime = endtime - (endtime % 86400) - 86400;
    END IF;
    SELECT starttime, endtime;
    WHILE starttime != endtime DO

```

```

SET done = 0;
OPEN cursor_brand;
-- Loop through all rows
REPEAT
  -- Get order number
  FETCH cursor_brand INTO currentbrand;
  SET totaltweets = (SELECT COUNT(*) FROM tweets WHERE username = currentbrand AND
faketweet = 0 AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) =
starttime);
  IF totaltweets IS NOT NULL THEN
    SET totalconversations = (SELECT COUNT(*) FROM tweets WHERE username = currentbrand
AND faketweet = 0 AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) =
starttime AND stats_conversation = 1);
    SELECT CONCAT("[", starttime, "] - Found ", totaltweets, "/", totalconversations, " for brand:
", currentbrand) AS STATUS;
    UPDATE tweets SET ratio_conversation = totalconversations / totaltweets WHERE tweettype
= 2 AND brand = currentbrand AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) %
86400)) = starttime;
  ELSE
    SELECT CONCAT("[", starttime, "] - No tweets for brand: ", currentbrand) AS STATUS;
  END IF;
-- End of loop
UNTIL done END REPEAT;
CLOSE cursor_brand;
-- go to the next day
SET starttime = starttime + 86400;
END WHILE;
END//
DELIMITER ;

DELIMITER //
DROP PROCEDURE IF EXISTS RATIO_HASHTAGS//
CREATE PROCEDURE RATIO_HASHTAGS()
BEGIN
  DECLARE done BOOLEAN DEFAULT 0;
  DECLARE starttime DECIMAL;
  DECLARE endtime DECIMAL;
  DECLARE totalhashtags INT;
  DECLARE totaltweets INT;
  DECLARE currentbrand VARCHAR(32);
  DECLARE cursor_brand CURSOR FOR SELECT username FROM tweets WHERE tweettype = 1
GROUP BY username;
  -- Declare continue handler
  DECLARE CONTINUE HANDLER FOR SQLSTATE '02000' SET done=1;
  UPDATE tweets SET ratio_hashtag = 0;
  SET starttime = (SELECT MIN(dateoftweet) / 1000 FROM tweets);
  IF starttime % 86400 >= 7200 THEN
    SET starttime = starttime - (starttime % 86400);
  ELSE
    SET starttime = starttime - (starttime % 86400) - 86400;
  END IF;
  SET endtime = (SELECT MAX(dateoftweet) / 1000 FROM tweets);

```

```

IF endtime % 86400 >= 7200 THEN
    SET endtime = endtime - (endtime % 86400);
ELSE
    SET endtime = endtime - (endtime % 86400) - 86400;
END IF;
SELECT starttime, endtime;
WHILE starttime != endtime DO
    SET done = 0;
    OPEN cursor_brand;
    -- Loop through all rows
    REPEAT
        -- Get order number
        FETCH cursor_brand INTO currentbrand;
        SET totaltweets = (SELECT COUNT(*) FROM tweets WHERE username = currentbrand AND
faketweet = 0 AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) =
starttime);
        IF totaltweets IS NOT NULL THEN
            SET totalhashtags = (SELECT COUNT(*) FROM tweets WHERE username = currentbrand AND
faketweet = 0 AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) =
starttime AND stats_hashtag = 1);
            SELECT CONCAT("[", starttime, "] - Found ", totalhashtags, "/", totaltweets, " for brand: ",
currentbrand) AS STATUS;
            UPDATE tweets SET ratio_hashtag = totalhashtags / totaltweets WHERE tweettype = 2 AND
brand = currentbrand AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400))
= starttime;
        ELSE
            SELECT CONCAT("[", starttime, "] - No tweets for brand: ", currentbrand) AS STATUS;
        END IF;
        -- End of loop
    UNTIL done END REPEAT;
    CLOSE cursor_brand;
    -- go to the next day
    SET starttime = starttime + 86400;
END WHILE;
END//
DELIMITER ;

DELIMITER //
DROP PROCEDURE IF EXISTS RATIO_MENTIONS//
CREATE PROCEDURE RATIO_MENTIONS()
BEGIN
    DECLARE done BOOLEAN DEFAULT 0;
    DECLARE starttime DECIMAL;
    DECLARE endtime DECIMAL;
    DECLARE totalmentions INT;
    DECLARE totaltweets INT;
    DECLARE currentbrand VARCHAR(32);
    DECLARE cursor_brand CURSOR FOR SELECT username FROM tweets WHERE tweettype = 1
GROUP BY username;
    -- Declare continue handler
    DECLARE CONTINUE HANDLER FOR SQLSTATE '02000' SET done=1;
    UPDATE tweets SET ratio_mention = 0;

```

```

SET starttime = (SELECT MIN(dateoftweet) / 1000 FROM tweets);
IF starttime % 86400 >= 7200 THEN
    SET starttime = starttime - (starttime % 86400);
ELSE
    SET starttime = starttime - (starttime % 86400) - 86400;
END IF;
SET endtime = (SELECT MAX(dateoftweet) / 1000 FROM tweets);
IF endtime % 86400 >= 7200 THEN
    SET endtime = endtime - (endtime % 86400);
ELSE
    SET endtime = endtime - (endtime % 86400) - 86400;
END IF;
SELECT starttime, endtime;
WHILE starttime != endtime DO
    SET done = 0;
    OPEN cursor_brand;
    -- Loop through all rows
    REPEAT
        -- Get order number
        FETCH cursor_brand INTO currentbrand;
        SET totaltweets = (SELECT COUNT(*) FROM tweets WHERE username = currentbrand AND
faketweet = 0 AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) =
starttime);
        IF totaltweets IS NOT NULL THEN
            SET totalmentions = (SELECT COUNT(*) FROM tweets WHERE username = currentbrand AND
faketweet = 0 AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) =
starttime AND stats_mention = 1);
            SELECT CONCAT("[", starttime, "] - Found ", totalmentions, "/", totaltweets, " for brand: ",
currentbrand) AS STATUS;
            UPDATE tweets SET ratio_mention = totalmentions / totaltweets WHERE tweettype = 2 AND
brand = currentbrand AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400))
= starttime;
        ELSE
            SELECT CONCAT("[", starttime, "] - No tweets for brand: ", currentbrand) AS STATUS;
        END IF;
    -- End of loop
    UNTIL done END REPEAT;
    CLOSE cursor_brand;
    -- go to the next day
    SET starttime = starttime + 86400;
END WHILE;
END//
DELIMITER ;

```

The next functions calculate the absolute number of conversations, hashtags and mentions related to the amount of tweets for a brand within the 24 hour time period.

```

DELIMITER //
DROP PROCEDURE IF EXISTS ABSOLUTE_CONVERSATIONS//
CREATE PROCEDURE ABSOLUTE_CONVERSATIONS()
BEGIN
    DECLARE done BOOLEAN DEFAULT 0;

```

```

DECLARE starttime DECIMAL;
DECLARE endtime DECIMAL;
DECLARE totalconversations INT;
DECLARE currentbrand VARCHAR(32);
DECLARE cursor_brand CURSOR FOR SELECT username FROM tweets WHERE tweettype = 1
GROUP BY username;
-- Declare continue handler
DECLARE CONTINUE HANDLER FOR SQLSTATE '02000' SET done=1;
UPDATE tweets SET absolute_conversation = 0;
SET starttime = (SELECT MIN(dateoftweet) / 1000 FROM tweets);
IF starttime % 86400 >= 7200 THEN
    SET starttime = starttime - (starttime % 86400);
ELSE
    SET starttime = starttime - (starttime % 86400) - 86400;
END IF;
SET endtime = (SELECT MAX(dateoftweet) / 1000 FROM tweets);
IF endtime % 86400 >= 7200 THEN
    SET endtime = endtime - (endtime % 86400);
ELSE
    SET endtime = endtime - (endtime % 86400) - 86400;
END IF;
SELECT starttime, endtime;
WHILE starttime != endtime DO
    SET done = 0;
    OPEN cursor_brand;
    -- Loop through all rows
    REPEAT
        -- Get order number
        FETCH cursor_brand INTO currentbrand;
        SET totalconversations = (SELECT COUNT(*) FROM tweets WHERE username = currentbrand
AND faketweet = 0 AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) =
starttime AND stats_conversation = 1);
        IF totalconversations IS NOT NULL THEN
            SELECT CONCAT("[", starttime, "] - Found ", totalconversations, " for brand: ", currentbrand)
AS STATUS;
            UPDATE tweets SET absolute_conversation = totalconversations WHERE tweettype = 2 AND
brand = currentbrand AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400))
= starttime;
        ELSE
            SELECT CONCAT("[", starttime, "] - No tweets for brand: ", currentbrand) AS STATUS;
        END IF;
    -- End of loop
    UNTIL done END REPEAT;
    CLOSE cursor_brand;
    -- go to the next day
    SET starttime = starttime + 86400;
END WHILE;
END//
DELIMITER ;

DELIMITER //
DROP PROCEDURE IF EXISTS ABSOLUTE_MENTIONS//

```



```

CREATE PROCEDURE ABSOLUTE_MENTIONS()
BEGIN
  DECLARE done BOOLEAN DEFAULT 0;
  DECLARE starttime DECIMAL;
  DECLARE endtime DECIMAL;
  DECLARE totalmentions INT;
  DECLARE currentbrand VARCHAR(32);
  DECLARE cursor_brand CURSOR FOR SELECT username FROM tweets WHERE tweettype = 1
GROUP BY username;
  -- Declare continue handler
  DECLARE CONTINUE HANDLER FOR SQLSTATE '02000' SET done=1;
  UPDATE tweets SET absolute_mention = 0;
  SET starttime = (SELECT MIN(dateoftweet) / 1000 FROM tweets);
  IF starttime % 86400 >= 7200 THEN
    SET starttime = starttime - (starttime % 86400);
  ELSE
    SET starttime = starttime - (starttime % 86400) - 86400;
  END IF;
  SET endtime = (SELECT MAX(dateoftweet) / 1000 FROM tweets);
  IF endtime % 86400 >= 7200 THEN
    SET endtime = endtime - (endtime % 86400);
  ELSE
    SET endtime = endtime - (endtime % 86400) - 86400;
  END IF;
  SELECT starttime, endtime;
  WHILE starttime != endtime DO
    SET done = 0;
    OPEN cursor_brand;
    -- Loop through all rows
    REPEAT
      -- Get order number
      FETCH cursor_brand INTO currentbrand;
      SET totalmentions = (SELECT COUNT(*) FROM tweets WHERE username = currentbrand AND
faketweet = 0 AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) =
starttime AND stats_mention = 1);
      IF totalmentions IS NOT NULL THEN
        SELECT CONCAT("[", starttime, "] - Found ", totalmentions, " for brand: ", currentbrand) AS
STATUS;
        UPDATE tweets SET absolute_mention = totalmentions WHERE tweettype = 2 AND brand =
currentbrand AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) =
starttime;
      ELSE
        SELECT CONCAT("[", starttime, "] - No tweets for brand: ", currentbrand) AS STATUS;
      END IF;
    -- End of loop
    UNTIL done END REPEAT;
    CLOSE cursor_brand;
    -- go to the next day
    SET starttime = starttime + 86400;
  END WHILE;
END//
DELIMITER ;

```

```

DELIMITER //
DROP PROCEDURE IF EXISTS ABSOLUTE_HASHTAGS//
CREATE PROCEDURE ABSOLUTE_HASHTAGS()
BEGIN
    DECLARE done BOOLEAN DEFAULT 0;
    DECLARE starttime DECIMAL;
    DECLARE endtime DECIMAL;
    DECLARE totalhashtags INT;
    DECLARE currentbrand VARCHAR(32);
    DECLARE cursor_brand CURSOR FOR SELECT username FROM tweets WHERE tweettype = 1
GROUP BY username;
    -- Declare continue handler
    DECLARE CONTINUE HANDLER FOR SQLSTATE '02000' SET done=1;
    UPDATE tweets SET absolute_hashtag = 0;
    SET starttime = (SELECT MIN(dateoftweet) / 1000 FROM tweets);
    IF starttime % 86400 >= 7200 THEN
        SET starttime = starttime - (starttime % 86400);
    ELSE
        SET starttime = starttime - (starttime % 86400) - 86400;
    END IF;
    SET endtime = (SELECT MAX(dateoftweet) / 1000 FROM tweets);
    IF endtime % 86400 >= 7200 THEN
        SET endtime = endtime - (endtime % 86400);
    ELSE
        SET endtime = endtime - (endtime % 86400) - 86400;
    END IF;
    SELECT starttime, endtime;
    WHILE starttime != endtime DO
        SET done = 0;
        OPEN cursor_brand;
        -- Loop through all rows
        REPEAT
            -- Get order number
            FETCH cursor_brand INTO currentbrand;
            SET totalhashtags = (SELECT COUNT(*) FROM tweets WHERE username = currentbrand AND
faketweet = 0 AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) =
starttime AND stats_hashtag = 1);
            IF totalhashtags IS NOT NULL THEN
                SELECT CONCAT("[", starttime, "] - Found ", totalhashtags, " for brand: ", currentbrand) AS
STATUS;
                UPDATE tweets SET absolute_hashtag = totalhashtags WHERE tweettype = 2 AND brand =
currentbrand AND (((dateoftweet / 1000) - 7200) - (((dateoftweet / 1000) - 7200) % 86400)) =
starttime;
            ELSE
                SELECT CONCAT("[", starttime, "] - No tweets for brand: ", currentbrand) AS STATUS;
            END IF;
        -- End of loop
        UNTIL done END REPEAT;
        CLOSE cursor_brand;
        -- go to the next day
        SET starttime = starttime + 86400;
    END WHILE;
END //

```

```
END WHILE;  
END//  
DELIMITER ;
```

The previous scripts are all executed by a simple CALL.

This script scans whether or not a tweets contains an URL. The text string is to match only valid URLs.

```
UPDATE tweets tw  
SET stats_url = 1  
WHERE tw.message RLIKE '(http://(www\.)?|www\.)[A-Za-z0-9][A-Za-z0-9\.\-]*\.[A-Za-z]{2,6}(\$|\)');
```

APPENDIX 4: TWEET DISTRIBUTION

Figure 15: Amount of tweets per day

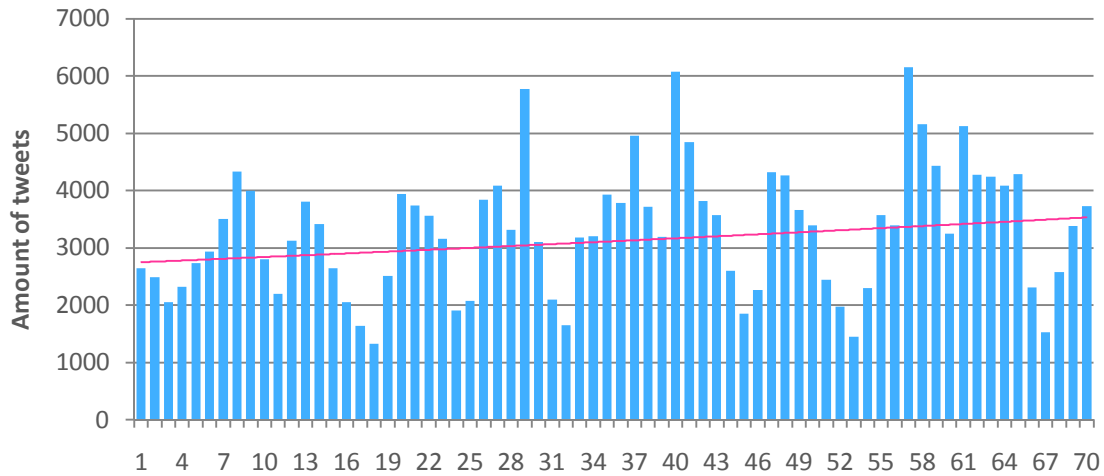


Figure 17: Amount of tweets per day of the week

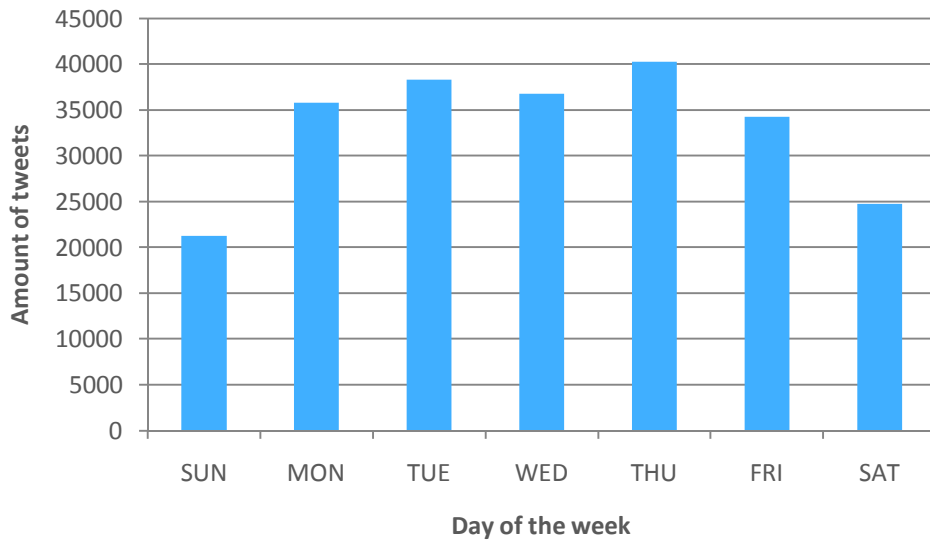
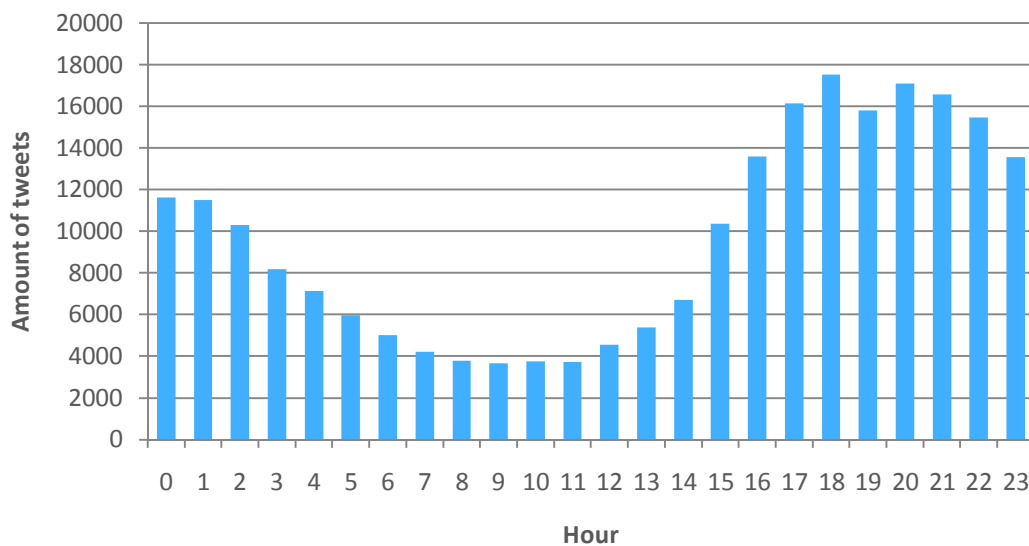


Figure 16: Amount of tweets per hour (GMT +01)



APPENDIX 5: CONTROLLING ACTIVITY GAPS

The Java application (not able to publish code here) checks the brands' Twitter activity and shows the blocks of time the brands have not tweeted. Since the activity of brands is related to consumer tweets within a 24 hour timespan, brands are to tweet at least once a day. As such, the following script controls for gaps in tweeting only to correctly execute the functions.

26-02-2011 16:07:52 -- AMPEnergy had 12 empty blocks of time. **Filtered.**
26-02-2011 16:09:32 -- Blackberry had 11 empty blocks of time.
26-02-2011 16:11:09 -- British_Airways had 10 empty blocks of time.
26-02-2011 16:12:39 -- canon_camera had 8 empty blocks of time. **Filtered.**
26-02-2011 16:14:06 -- CarnivalCruise had 8 empty blocks of time.
26-02-2011 16:15:06 -- CocaCola had 2 empty blocks of time.
26-02-2011 16:16:36 -- CokeZero had 9 empty blocks of time.
26-02-2011 16:18:27 -- ConnectByHertz had 13 empty blocks of time. **Filtered.**
26-02-2011 16:19:44 -- Delta had 6 empty blocks of time.
26-02-2011 16:20:48 -- DisneyParks had 3 empty blocks of time.
26-02-2011 16:21:43 -- drpepper had 1 empty blocks of time.
26-02-2011 16:23:28 -- Expedia had 12 empty blocks of time.
26-02-2011 16:25:04 -- Gatorade had 10 empty blocks of time.
26-02-2011 16:26:40 -- HALcruises had 10 empty blocks of time.
26-02-2011 16:28:16 -- hpnews had 10 empty blocks of time.
26-02-2011 16:30:22 -- htc had 16 empty blocks of time.
26-02-2011 16:31:35 -- JetBlue had 5 empty blocks of time.
26-02-2011 16:33:16 -- Logitech had 11 empty blocks of time.
26-02-2011 16:35:06 -- MarriottIntl had 13 empty blocks of time.
26-02-2011 16:36:51 -- Microsoft had 12 empty blocks of time.
26-02-2011 16:38:18 -- MonsterEnergy had 8 empty blocks of time.
26-02-2011 16:39:45 -- MotoMobile had 8 empty blocks of time with.
26-02-2011 16:41:48 -- mtn_dew had 16 empty blocks of time with.
26-02-2011 16:43:36 -- nokia had 12 empty blocks of time.
26-02-2011 16:45:32 -- pepsi had 14 empty blocks of time.
26-02-2011 16:46:27 -- redbull had 1 empty blocks of time.
26-02-2011 16:47:37 -- SonyElectronics had 4 empty blocks of time.
26-02-2011 16:49:28 -- sonericsson had 13 empty blocks of time.
26-02-2011 16:50:42 -- SouthwestAir had 5 empty blocks of time.
26-02-2011 16:52:49 -- TropicanaOJ had 16 empty blocks of time.

APPENDIX 6: DATA GATHERING PROCESS IN THE PRETEST

The application consists of three processes. First the application checks for the amount of followers of the brands' accounts every four hour. This process is not very resource intensive. Yet it measures the amount of followers five times a day to control for Twitter downtime. The second part of the application checks for new tweets every three hour. The second and third part of the study are both resource-intensive, and therefore are executed two minutes passed (half passed) x hour. This because it is expected that many other application check for updates on the Twitter API every round (half) hour. By running the processes at not such a general time, the chance on downtime for the application decreases. The process retries seven times to acquire the data related to a particular brand. The third process, in which the new tweets are actually downloaded and saved into the database is executed an half an hour after the second process. The process retries to acquire a tweet three times. If it fails, the tweet is scheduled for download at the next execution of the process. The according time schedule:

Process 01: Check for amount of followers

ANNE: 00:00 04:00 08:00 16:00 20:00

ELISA: 02:00 06:00 10:00 18:00 22:00

Process 02: Check for new tweets

ANNE: 00:02 03:02 06:02 09:02 12:02 15:02 18:02 21:02

ELISA: 01:32 04:32 07:32 10:32 13:32 16:32 19:32 22:32

Process 03: Actually downloads the new tweets and saves them into the database

ANNE: 00:32 03:32 06:32 09:32 12:32 15:32 18:32 21:32

ELISA: 02:02 05:02 08:02 11:02 14:02 17:02 20:02 23:02

APPENDIX 7: CELL DATA CONTAINED BY A SINGLE RECORD

Cells	Cell data	
Original data from Radian6		
ARTICLE_ID	The unique ID of the tweet in the Radian6 database	Nominal, Unique, Numeric
AUTHOR	The accountname of the user who posted the tweet	Nominal, Text
CONTENT	The actual tweet message	Nominal, Text
ARTICLE_URL	The URL to the individual tweet on Twitter.com	Nominal, Text
PUBLISH_DATE	The date-time of the tweet when posted. Follows the format: MMM DD, YYYY HH:MM TT	Nominal, Text
FOLLOWING	The amount of users the tweet posters follows	Ratio, Numeric
FOLLOWERS	The amount of users who follow the tweet poster	Ratio, Numeric
UPDATES	The cumulative number of the poster's tweets	Ratio, Numeric
SENTIMENT	The classified sentiment	Nominal, Text
Data added for analysis		
TWEET_ID	The unique ID of the tweet at Twitter	Ratio, Unique, Numeric
BRAND	Which brand is referred to	Nominal, Numeric
INDUSTRY	The industry of the referred brand	Nominal, Numeric
DATE_UNIXTIME	The date extracted from the tweet, in unixtime	Ratio, Numeric
DATE_FORMATTED	The unixtime converted to DD-MMM-YYYY HH:MM	Date, Numeric
DATE_DAYOFWEEK	The day of the week [MON-SUN]	Date, Text
DATE_DAYOFYEAR	The day of the year [001-366]	Ratio, Numeric
DATE_DAYOFRESEARCH	The day of the research [01-70]	Ratio, Numeric
DATE_WEEKOFRESEARCH	The week of the research [01-10]	Ratio, Numeric
DATE_HOUR	The hour of the day [00-23]	Ratio, Numeric
STATS_CONVERSATION	Is the tweet a conversation	Binary, Numeric
STATS_MENTION	Does the tweet contain a mention	Binary, Numeric
STATS_MENTION_AMOUNT	How many mentions does the tweet contain	Ratio, Numeric
STATS_HASHTAG	Does the tweet contain a hashtag	Binary, Numeric
STATS_HASHTAG_AMOUNT	How many hashtags does a tweet contain	Ratio, Numeric
STATS_URL	Does the tweet contain an URL	Binary, Numeric
STATS_CHARACTERCOUNT	The amount of characters in the tweet	Ratio, Numeric
RATIO_CONVERSATION	STATS_CONVERSATION / Total number of tweets within timespan	Ratio, Numeric
RATIO_HASHTAG	STATS_HASHTAG / Total number of tweets within timespan	Ratio, Numeric
RATIO_MENTION	STATS_MENTION / Total number of tweets within timespan	Ratio, Numeric
SUM_MENTIONS	The sum of mentions for the referred brand in the timespan	Ratio, Numeric
INCREMENTAL_FOLLOWING	The change in FOLLOWING for referred brand within	Ratio, Numeric

FOLLOWING	timespan	
INCREMENTAL_FOLLOWERS	The change in FOLLOWERS for referred brand within timespan	Ratio, Numeric
TRANS_INCR_FOLLOWERS	Removed some extreme outliers from INCREMENTAL_FOLLOWERS	Ratio, Numeric
LOG_INCR_FOLLOWING	INCREMENTAL_FOLLOWING log-transformed	Ratio, Numeric
LOG_INCR_FOLLOWERS	TRANS_INCR_FOLLOWERS log-transformed	Ratio, Numeric
ABSOLUTE_CONVERSATION	The sum of consumers' STATS_CONVERSATION within timespan	Ratio, Numeric
LOGGED_CONVERSATION	ABSOLUTE_CONVERSATION log-transformed	Ratio, Numeric
ABSOLUTE_HASH_TAG	The sum of consumers' STATS_HASHTAG within timespan	Ratio, Numeric
LOG_HASHTAG	ABSOLUTE_HASHTAG log-transformed	Ratio, Numeric
SENTIMENT_RECODED	SENTIMENT recoded to numeric	Ordinal, Numeric
SENTIMENT_TRANS	Removed neutral for SENTIMENT_RECODED, thereby transformed to binary	Binary, Numeric
EXTERNAL_INFLUENCES	Control variable to control for extraneous influences	Binary, Numeric
FILTER_BRANDS	Filter which filters out the all the tweets referring to the brands which didn't meet the activity requirements	Binary, Numeric

APPENDIX 8: RELIABILITY ANALYSIS OF THE SENTIMENT ANALYSIS

A summary of the manual check of the accuracy of the sentiment analysis as produced by Radian6.

	<i>Positive</i>		<i>Negative</i>	
	Correct	Incorrect	Correct	Incorrect
Beverage	92	8	66	34
Travel	92	8	94	6
Electronics	95	5	86	14
Sumtotal	93%	7%	82%	18%

A summary of the manual check of the accuracy of the sentiment analysis of the pretest as produced by Tweetfeel Biz.

	<i>Positive</i>		<i>Negative</i>	
	Correct	Incorrect	Correct	Incorrect
Beverage	91	9	51	49
Electronics	90	10	86	14
Sumtotal	90,5%	9,5%	70%	30%