The Pennsylvania State University

The Graduate School

College of Information Sciences and Technology

BRANDS ALL A TWITTER:
THE INFLUENCES OF TWITTER ON BRANDS AND CONSUMERS

A Dissertation in

Information Sciences and Technology

by

Mimi Zhang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2010

The dissertation of Mimi Zhang was reviewed and approved* by the following:

Bernard J. (Jim) Jansen
Associate Professor of Information Sciences and Technology
Dissertation Advisor
Chair of Committee

Lynette Kvasny
Associate Professor of Information Sciences and Technology

Shawn Clark
Professor of Practice of Information Sciences and Technology
Executive Director of Institute for Global Prescience

James L. Rosenberger
Professor of Statistics

Abdur Chowdhury
Chief Scientist at Twitter, Inc.
Special Member

Michael D. McNeese
Professor of Information Sciences and Technology
Professor-in-Charge of College of Information Sciences and Technology

*Signatures are on file in the Graduate School

ABSTRACT

Local, national, and global commercial businesses continue to be increasingly interested in leveraging Twitter to present the brand, manage word-of-mouth communication, and interact with consumers. However, there is a general lack of solid and comprehensive understanding of the platform. This dissertation attempts to fill that gap by providing an in-depth analysis of word-of-mouth communications among consumers and businesses on Twitter and by uncovering Twitter community dynamics from a business perspective. I examined three aspects important to current and prospective business Twitter users, namely benefit (i.e., what can a business get from Twitter?), role (i.e., how active should a business be on Twitter?), and audience (i.e., who connects to a business on Twitter?).

To address my research questions, I collected approximately 2 million tweets pertaining to nine brands from May, 2008, to May, 2009. I performed bootstrap-based nonparametric analysis of variance to address the benefit question and found that following a brand or being followed by a brand has statistically significant main effect on the number of word-of-mouth messages consumers send out as well as the number of word-of-mouth conversations that consumers participate in with other consumers and the brand. I conducted path analysis to explore questions about a business' role and discovered that as an active participant in the word-of-mouth dialogue, a business can increase the engagement level of consumers in word-of-mouth communication. I carried out TwoStep cluster analysis to analyze audience makeup and identified five types of consumers in a brand's immediate social network.

This dissertation advances the understanding of the potential business value of Twitter deployment and brand management. This work provides insights about the analytics of social networking on micro-communication platforms such as Twitter. Considering that online word-of-mouth marketing is one of the most effective brand enhancing and selling tools, this work offers guidance to brands on the ways to approach customers, communicate with them, and persuade them to talk about the brands.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, to Professor Jim Jansen, my advisor, my academic father

Because of your mentorship, your support and your enthusiasm in helping me, I can finish my PhD and my dissertation. I am and will always remain deeply indebted to you. You are truly an extraordinary individual.

To Dr Abdur Chowdhury

Without your gracious help on data collection, I could never have started my dissertation. I really appreciate your help.

To Professor Lynette Kvasny, Professor Shawn Clark and Professor James Rosenberger

Thanks for your diligent services on my dissertation committee.

To my friends

Jung-Woo Sohn, thanks for teaching me Python, helping me debug, and giving me rides home when I stayed in office too late.

Mithu Bhattacharya, thanks for loaning me your sweet apartment as my temporary office during the winter break, taking those long walks with me, and pumping me up along the way.

Last but most of all, to my beloved parents

Thanks for loving me, believing in me, giving me emotional and financial support, and always being there for me whenever I need you. I love you!

谨以此论文献给

我亲爱的爸爸(张秉优)妈妈(董玉华),

世界上最好的爸爸妈妈!

I dedicate this dissertation to my parents

Yuhua Dong and Bingyou Zhang,

the best parents in the world!

CHAPTER 1. INTRODUCTION

Twitter is entering the social media mainstream, providing a platform for word-of-mouth communication, and becoming a critical communication channel for commercial brands. Twitter has grown very fast. According to comScore, the total number of unique visitors to Twitter grew 1,200% from November 2008 to November 2009 (Rao, December 15, 2009). However, this growth rate, based only on the Web traffic directly to the Twitter Website, was underestimated: the real growth was much higher. Twitter has a very long tail of traffic coming from third-party clients such as TweetDeck, Seesmic, and Tweetie, which was not taken into consideration by comScore. Twitter had approximately 60 million users by the end of 2009, among which only 20 million were on twitter.com (Schonfeld, January 18, 2010). Twitter not only has a large and fast-growing user base, but their users are active in generating and distributing information. In early 2010, the company reported they had 50 million tweets per day, roughly 600 tweets per second (Weil, February 22, 2010).

Some commercial companies sense the business potential of Twitter. They are seizing the opportunity and employing it as a new online marketing tool. They create their brands' presence on Twitter and mingle with their customers. In 2009, among the Inc. 500 companies (a list of the fastest-growing private American companies compiled by Inc. Magazine), 80% of companies claimed they were using social networking sites, and 52% of the companies reported they were using Twitter; 87% of companies agreed the use of social networking for their business was successful, and 82% reported the use of Twitter was successful (Barnes & Mattson, 2009).

A number of early adopters have benefitted from their brand presence on Twitter in terms of gaining attention, loyalty, traffic and sales. Jame-Ane Ervin (a professional consulting company) reported a 400% increase in conversions (completing a desired action) by using Twitter (Social Media B2B, 2009). In 2009, Dell gained more than $6.5 million in revenue from its Twitter accounts (Ionescu, 2009). The American Customer Satisfaction Index rating of Comcast increased 9.3% in satisfaction after the company used Twitter (Israel, 2009). As many companies witness the effectiveness and great potential of social media services like Twitter, they tend to invest more resources in the social media space. According to Vignette's (2009) survey, 71% of companies planned to increase investments in social media by an average of 40% in 2009 due to reasons such as low cost, feeling compelled to do so, and getting traction. Two consultant companies teamed up in September 2009 to survey more than 1,100 client-side marketers, public relations professionals, and digital agencies, reporting that 86% of companies had a plan to spend more money on social media in 2010 (bigmouthmedia & Econsultancy, 2009). The use of social media technology by potential customers and by businesses to reach these customers demonstrates the importance of this area of research.

Twitter

What is Twitter? Twitter is by far the most popular micro-communication service. It was founded by Jack Dorsey, Biz Stone and Evan Williams in 2006 (Sagolla, 2009). Twitter's major function is to enable one to send out short messages to one's friends. The message is referred to as a tweet and its length can be up to 140 characters. Figure 1.1 and 1.2 show a personal account that I use to explain how Twitter functions. The other third-party clients have all these basic functions, but the interfaces may be a little different. On Twitter, each user has a Twitter profile page, as shown in Figure 1.1, which displays profile information including name, location, Website, and biography in the upper right corner. Below the profile section is social network information, which displays the number of followings (a.k.a., people who she follows to receive their tweets, which are displayed on her home page (Figure 1.2)), the number of followers (a.k.a., people who follow her to receive her tweets, which are displayed on their Twitter home pages), and the number of lists that include her. The network section shows the number of tweets sent out by her, the links to the page displaying her favorite tweets, and the pictures of her followings. Her picture and handle (a.k.a. user name on Twitter) are shown in the upper left hand. Below that is her tweet stream.

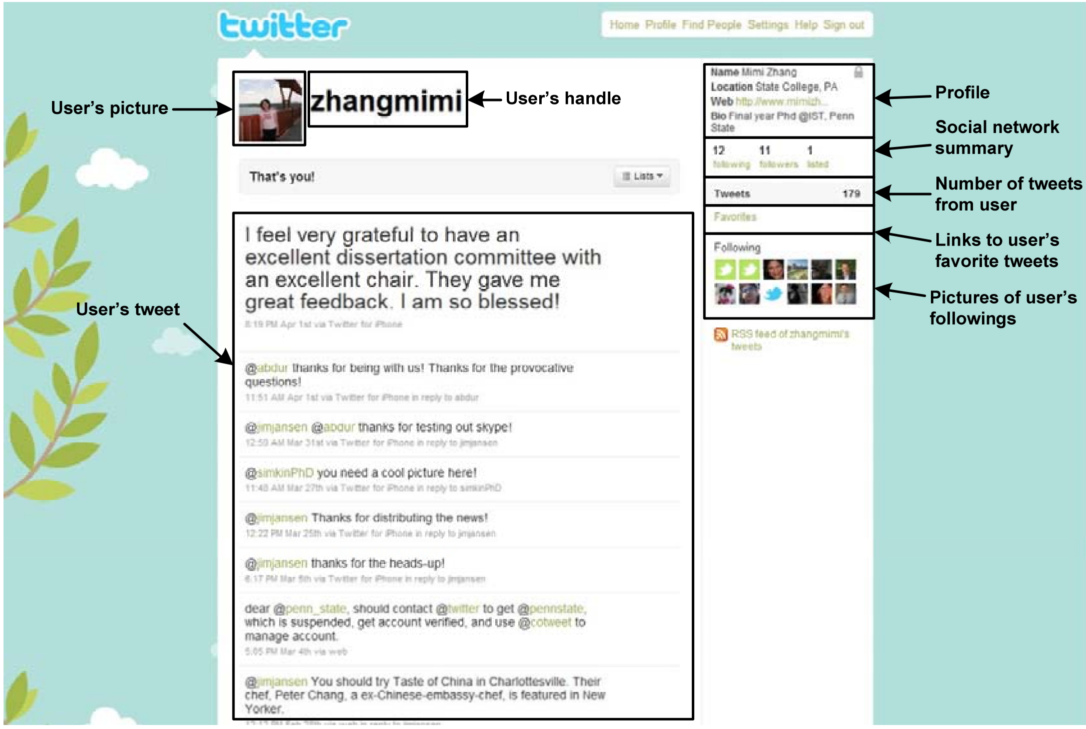


Figure 1.1 Twitter user profile page

The user also has a Twitter home page (Figure 1.2), which is more focused on tweets. On the upper left side, there is a text box where she can craft tweets to answer the question "What's happening?" which used to be "What are you doing?" Below that are the tweets from her and her followings. On the upper right hand side is a quick summary about her Twitter account including her handle, profile picture, number of tweets, number of followings, number of followers, and number of lists related to her. Below that is the Twitter dictionary showing the Twitter-related word or application and the corresponding explanation. Below the Twitter dictionary section are links to four types of tweet filters including tweets mentioning her, the direct message box where she can send and receive tweets in private, her favorite tweets, and different types of retweets relevant to her (retweets from her followings, her tweets retweeted by others, and tweets retweeted by her). Below the tweet section is the list, which can be used to classify people on Twitter and display their tweet streams under the same group label. Those people do not have to be her followers or followings, and literally can be anyone. This function can be used to classify followings and reduce the number of people the user follow and the number of tweets displayed on the user's home page. Table 1.1 is the summary of basic Twitter lingos relevant with this dissertation.

Figure 1.2 Twitter user home page

Table 1.1 Definitions of basic Twitter lingos

Twitter Lingo	**Definition**
Follower	To one Twitter user, a follower is someone who follows him/her in order to receive his/her tweets
Following	To one Twitter user, a following is someone who he/she follows in order to receive tweets
Tweet	Tweet is a message created by one on Twitter and is limited to 140 characters.
@reply[a]	@reply is to craft a tweet with "@user handle" in the beginning.
@mention[a]	@mention is to craft a tweet with "@user handle" in the tweet but not in the beginning.
Retweet (RT)	Retweet is the action of forwarding one's tweet with the acknowledgment of its sources by using RT, via, etc., in the tweet.
#hash-tag	#hash-tag sign is used before a term to create a label for the tweet and more importantly make it easy to find the tweets when using the hash-tagged terms as keywords during search by reducing term ambiguity.
List	List is to classify people on Twitter and display their tweet streams under the same group label. Those people do not have to have any connection with the link creator and literally can be anyone. This function can be used to organize Twitter users and help Twitter users to uncover interesting Twitter users.

a. In this dissertation, I do not differentiate @reply and @mention. I refer to both as "@user handle".

Twitter is widely adopted by both regular and organizational users. Based on my understanding, Twitter becomes popular for the following five reasons. First, Twitter is a real-time and flexible communication platform. Tweets are updated instantaneously in the system. The system supports one-to-one, one-to-many, and many-to-many communications.

Second, the system is simple and intuitive. When registering, people do not need to fill in lengthy forms on their backgrounds and preferences. They only need to answer a couple of basic questions to finish the registration. They can start using the system immediately since the concept of the system is very straightforward and the learning curve is short.

Third, the affordance of Twitter is lower than most of the services or systems on the Internet. Micro-messages are easy to craft and consume. It also encourages people to craft and consume the messages in real-time.

Fourth, Twitter is a very open system. It can be openly accessed in three different ways: Web interface (its own Website, MySpace, Facebook, AIM), via third-party application (desktop application and mobile phone application), and cell phone short message.

Fifth, Twitter is a network on many different levels. It is a social network that can grow many companies, forming a pseudo-corporate network with Twitter in the center. Twitter's openness and flexibility allow others to develop application programming interfaces (API) freely, which creates a number of small companies and forms a Twitter ecosystem. These companies all depend on the mother company (Twitter), so the fate of Twitter is not just its fate. It is also relevant to all those small companies. Twitter reportedly had 50,000 applications built by using Twitter (API) by December 2009 (Wauters, December 9, 2009).

Twitter is a place for people to express themselves and pass along information. Consumers can spread word-of-mouth (WOM) messages, share their brand consumption experiences, seek advice, or make recommendations on Twitter. These messages will influence their followers. Tweets are the real-time information flow recording what we are doing, planning, and thinking, or what is happening. In the context of a specific situation and occasion when WOM communication happens (Allsop, Bassett, & Hoskins, 2007, p. 402), consumers can use this real-time channel to send out the message immediately. It will have real-time influence on the message recipients. WOM messages on Twitter can be even more powerful than offline because they can reach more people than offline channels. In addition, the people one follows can be friends, experts, or celebrities. WOM from the latter two groups will be very influential (Wangenheim & Bayon, 2004) to its followers on Twitter, but they might not be all that approachable in the other non-real-time social contexts. In the meantime, commercial business users can push out brand messages to their followers, monitor tweet streams, and take appropriate actions. Interactions with the consumers can show how much the brand genuinely cares about its customers. Therefore, Twitter is a great communication channel for consumers and business, which needs in-depth exploration from both the academic and the practical standpoints.

Problem Statement

Twitter is a typical, functional, parsimonious, and popular social media service, which makes it a good starting point for businesses to enter the social media arena and one of the best social tools empowering the brand to connect with customers. It provides the basic social media functions, such as owning a profile page, connecting with people, and sharing text and multimedia information. A business can get started on Twitter easily because the registration process is simple and the concept of the system is intuitive. Moreover, it is a very popular service with considerable media spotlight, which in turn makes it even more popular. Therefore, it is appropriate and timely to study Twitter.

Due to the popularity of Twitter among regular users and commercial business users, there are hundreds of thousands of publications on guidance of using Twitter for business. Not to mention millions of posts online, there are a lot of published books like "Twitter power 2.0: How to dominate your market one tweet at a time", "Twitter marketing for dummies", "Twitter marketing: An hour a day", "Twitter marketing: Promote yourself and your business on earth's hottest social network", and so on. However, most of these publications are based on the intuitive understanding of the platform and lacking scientific support. This is what motivates my dissertation. My dissertation serves as a bridge between the academic and the practical, both of which are critical for marketing strategy development. It advances the understanding about Twitter from the commercial business' perspective and provides actionable recommendations. I tackle three fundamental questions business having in mind about Twitter: what can a business get from Twitter? How active should a business be on Twitter? Who connects to a business on Twitter?

Based on the potential impact of social media on marketing for businesses, the goals of this dissertation are to:

1. uncover Twitter community dynamics, particularly those relevant to branding and word-of-mouth communication;
2. understand the way word-of-mouth disseminates in the Twitter community; and
3. define the influence of the Twitter community on word-of-mouth marketing.

The primary research question driving the research of this dissertation is to explore *what are influences on word-of-mouth communications, brand management and customer management on Twitter?* This meta-research question is addressed by answering the following component research questions:

Research question 1: What are the branding influences of social network on word-of-mouth communication on Twitter?

Research question 2: What are the influences of brand engagement in word-of-mouth communication on consumers' level of engagement in word-of-mouth communication on Twitter?

Research question 3: What are the characteristics of consumers connecting to brands in the Twitter community?

Brand Selection and Data Collection

In order to perform comprehensive analyses on branding and WOM message, I chose to study several specific brands. Given that this dissertation focuses on the WOM communication in Twitter community, it is important to select the business valuing its brand presence on Twitter, particularly those actively participating on the platform and successfully creating a community. Another way to understand the brand selection criterion is to choose business having successfully used Twitter as an online WOM communication tool.

I took a heterogeneous approach instead of a homogenous approach on brand selection and selected brands from multiple industries. When I designed this research and tried to come up with a list of business to study in 2009, there was not that many businesses using Twitter. It was hard to sample a decent amount of businesses maintaining active presences on Twitter in one industry.

On the other hand, consumers in a certain demographic sector could be clustered by industry. My dissertation focuses on overall consumers' behavior, and I want to research a more diversified sample of people. Moreover, consumers' behavior toward one brand is generally consistent with their behavior towards another brand. Between-industry differences are transparent on the consumers' behavior. Therefore, I opted for the heterogeneous brand selection approach. I did not want to limit my dissertation research to several demographic sectors, but rather I wanted to study everyone to increase generalizability of my dissertation work.

The nine brands (Table 1.2) I researched in this dissertation represent two broad brand categories on Twitter. Coffee Goundz (@CoffeeGroundz), Kogi BBQ (@kogibbq), and Naked Pizza (@NAKEDpizza) represent small local brands. The rest are nationwide or worldwide brands. These brands are from different industries. Coffee Goundz (@CoffeeGroundz), Kogi BBQ (@kogibbq), Naked Pizza (@NAKEDpizza), and Starbucks (@Starbucks) are from the fast food industry. Coffee Goundz (@CoffeeGroundz) and Starbucks (@Starbucks) are coffee houses. Kogi BBQ (@kogibbq) is a food truck business. Naked Pizza (@NAKEDpizza) is a take-out and delivery pizza restaurant. Comcast (@comcastcares) is a cable and Internet provider. Home Depot (@HomeDepot) is a home improvement retailer. H&R Block (@HRBlock) is a tax service provider. Whole Foods (@WholeFoods) is a grocery supermarket. Zappos (@zappos) represents the online store. Detailed brand briefings are available in Appendix A.

Table 1.2 Brands' Twitter accounts on May 31, 2009

Brand	Twitter	Date of First Tweet	Location	Biography on Profile	Following (*n*)	Follower (*n*)	Tweet (*n*)
Coffee Groundz	@CoffeeGroundz	August 26, 2008	Houston, TX	I am a strong cup of coffee and by night I am a Belgium beer.	5,963	8,761	5,503
Comcast	@comcastcares	Before May 1, 2008	Philadelphia, PA	Comcast Director of Digital Care Email: We_Can_Help@cable.comcast.com	23,001	31,702	29,062
Home Depot	@HomeDepot	May 16, 2008	Atlanta, GA	I'm a spokesperson, I moonlight on Twitter to offer another way for customers to ask about their projects and our stores. information@homedepot.com	8,645	12,171	1,665
H&R Block	@HRBlock	Before May 1, 2008	Kansas City, MO	Your tax people.	2,904	4,048	1,013
Kogi BBQ	@kogibbq	November 21, 2008	Los Angeles, CA	Korean BBQ Taco Truck	1,365	56,947	1,724
Naked Pizza	@NAKEDpizza	March 6, 2009	New Orleans, LA	an all natural and good for you pizza joint in new orleans. doing it one day at a time. I care. I really do.	4,914	7,073	1,702
Starbucks	@Starbucks	August 12, 2008	Seattle, WA	Freshly brewed tweets from Brad at Starbucks in Seattle, WA.	146,562	375,695	1,820
Whole Foods	@WholeFoods	June 20, 2008	Austin, TX	Fresh organic tweets from Whole Foods Market HQ in Austin, TX.	487,444	1,396,748	2,701
Zappos	@zappos	Before May 1, 2008	Las Vegas, NV	www.zappos.com blogs.zappos.com twitter.zappos.com	417,426	1,318,834	1,349

I have four data sets (Table 1.3) about these nine brands. First, I collected the profiles of the nine brands and their tweets from May 1, 2008 to May 31, 2009.

Second, I collected the follower and following information of all the nine brands including their Twitter user name, the brand one is connecting with, and the time.

Third, for each brand, I took a random sample of 300 accounts (including the followers and followings). Sample size is determined based on confidence interval as well as the overhead of data collection. Assuming the number of consumers connecting to the business Twitter account is at the million level, I can achieve 5.66% confidence interval at the confidence level of 95% with sample size of 300. I can achieve similar confidence interval (4.74% to 5.65%) at the confidence level of 95% with sample size 300, presumably the number of consumers connecting to business is from several thousands to several hundreds of thousands. Thus, choosing sample size of 300 can maintain approximately 5% confidence interval at the confidence level of 95%.

I sampled 150 followers and 150 followings. I also cross-referenced these two sets of 150 and made sure there was no overlap between the groups. Then I downloaded the profiles of all 2,700 accounts, their tweets, and account related information from May 1, 2008, to May 31, 2009.

Fourth, I collected all the tweets mentioning these nine brands during five weeks between May 1, 2008 and May 31, 2009. The queries I used to get the tweets mentioning about the brands are presented in Table 1.4. I tried to include as many variants of brand names as possible. To make sure the five weeks were not clustered and representative, I first had a stratified sample of five months, which were May 2008, August 2008, November 2008, February 2009, and May 2009. For each of the first four months, I selected different weeks to capture the behavior in response to the potentially different marketing strategy in different periods of a month. So I chose the first week in May 2008, second week in August 2008, third week in November 2008, and fourth week in February 2009. For May 2009, I chose the last week to represent the latest branded tweet trend of the period I studied. For each tweet, I knew the sender's user name, the time, and the message.

Table 1.3 Data set description

Data Set	Description
1	Nine brands: their profiles and tweets from May 1, 2008 to May 31, 2009
2	All the handles of the followers and followings of the nine brands
3	300 followers and followings of each of these nine brands: their profiles and tweets from May 1, 2008 to May 31, 2009
4	All tweets mentioning these nine brands during these five weeks: • 05/05/2008 to 05/11/2008 • 08/11/2008 to 08/17/2008 • 11/17/2008 to 11/23/2008 • 02/23/2009 to 03/01/2009 • 05/25/2009 to 05/31/2009

Table 1.4 Queries to collect tweets mentioning about brands

Brand	Query
Coffee Groundz	coffeegroundz OR #coffeegroundz OR @coffeegroundz OR coffee groundz
Comcast	comcastcare OR #comcastcare OR @comcastcare OR Comcast OR #comcast OR @comcast
Home Depot	homedepot OR #homedepot OR @homedepot OR home depot
H&R Block	hrblock OR #hrblock OR @hrblock OR hr block
Kogi BBQ	kogibbq OR #kogibbq OR @kogibbq OR kogi bbq OR #kogi OR @kogi
Naked Pizza	naked pizza OR #nakedpizza OR @nakedpizza OR nakedpizza
Starbucks	starbucks OR #starbucks OR @starbucks OR sbux OR #sbux OR @sbux
Whole Foods	wholefoods OR #wholefoods OR @wholefoods OR wholefood OR #wholefood OR @wholefood OR whole foods
Zappos	zappos OR #zappos OR @zappos OR zappo OR #zappo OR @zappo

Dissertation Outline

The rest of this dissertation is structured as follows:

- In "Chapter 2. Word-of-mouth communication under "follow" influence on Twitter", I present my study addressing Research Question 1 with regards to the influences of connecting to a brand Twitter account on consumers' engagement in WOM communication.

- In “Chapter 3. What ‘tweets’ around comes around: Business engagement in online word-of-mouth communication on Twitter”, I describe my research addressing Research Question 2 about the influences of business engagement in online WOM communication on consumers’ WOM engagement in Twitter community.
- In “Chapter 4. Consumer Who: Profiling consumers on Twitter”, I report my work on Research Question 3 pertaining to understanding consumers based on their engagement in WOM communication and Twitter community.
- In “Chapter 5. Conclusion”, I bring together all findings, provide a comprehensive description of WOM communication on Twitter, list potential strategies for business to use Twitter, and point out the directions of future research.

CHAPTER 2. WORD-OF-MOUTH COMMUNICATION UNDER "FOLLOW" INFLUENCE ON TWITTER

Twitter has a large number of users (Schonfeld, January 18, 2010) and an enormous amount of information (Weil, February 22, 2010). Where there are crowds and information, there is also potential business opportunity. Several companies like Dell (Ionescu, 2009) and Comcast (Israel, 2009) have successfully employed Twitter during the past couple of years. Some companies plan to invest more resources on Twitter (bigmouthmedia & Econsultancy, 2009), but other companies are struggling to find value and their positioning on Twitter because they are uncertain about what benefits Twitter can provide. The research reported in this chapter can bring insight to address this dilemma and argue that social media is an essential brand and customer management.

According to Jansen, Zhang, Sobel and Chowdhury's (2009) Twitter study, roughly 19.0% of tweets mention some brand or product, which indicates that Twitter can be a brand channel and a word-of-mouth communication channel. On Twitter, brands and customers exchange tweets on a wide range of topics. Jansen and fellow researchers (Jansen, et al., 2009) analyzed the content tweets mentioning Starbucks and found those tweets were relevant to product, service, new idea, marketing, and organization. In short, Twitter has an extensive coverage of the brand. Schmitt (2009) reported that following a brand on Twitter has dramatic influences on branding. Thus, Twitter is a place to connect with customers and manage the brand.

However, exactly how should the brands connect with customers on Twitter? Should they wait for customers to come to them, or should they reach out for customers? What will be the influence of these different connections on the way customers talk about the brand? These are some of the questions motivating my research. Moreover, word-of-mouth (WOM) research is critical to commercial brands, which has a range of influences on branding. Nielson (2007) claimed that WOM messaging is the most effective selling tool. Keller (2007) argued that WOM communication is “the most important and effective communications channel” (p.448). He also reported that WOM messages are perceived as being credible, motivating information sharing, and stimulating purchase intent. WOM communication also highly correlates with company growth (Reichheld, 2001).

In the following sections of this chapter, I first present relevant research in WOM communication and point out the position of my research. I pose my research question and explain the procedure that I took to tackle my question. Afterwards, I report my findings, discuss the ways to understand the findings, and present the practical contributions.

Word-of-mouth Communication

WOM communication is information propagation in social network from the information scientist's perspective. WOM communication is the branding media and advising channel from the marketer's perspective. WOM communication was defined as "oral, person-to-person communication between a receiver and a communicator whom the receiver perceives as being non-commercial, concerning a brand, a product or a service" (Arndt, 1967a, p. 190). This definition described WOM communication in a traditional offline fashion. It captures three key characteristics including information propagation between individuals, non-commercial nature of the message creator, and message on a commercial object.

With the introduction of social networking services, WOM communication goes beyond offline. Its one-to-one communication nature becomes less apparent. Most of the time, a WOM message is delivered in a one-to-many way, such as product reviews on Amazon, branded status updates on Facebook, and branded tweets on Twitter. Also, with the assistance of technology, the brand owners can participate in the WOM communication. They can initiate the WOM process by providing incentives. For example, Starbucks started an outdoor ad campaign in six major US cities accompanied with a WOM marketing campaign on May 19, 2009. It sent out a tweet as "To win: first 5 to post a picture of one of our new outdoor ads on twitter, use the hashtag #top3percent rules: http://bit.ly/QzoI5 US only." (http://twitter.com/Starbucks/status/1848921500) The company encouraged its customers on Twitter to take pictures and post on Twitter with the predetermined hash tag. Those who posted the pictures selected as the winners would get a reward. It was a well-designed WOM marketing campaign blurring the border of offline and online marketing campaigns and maneuvering the power of social networks.

Certainly, the company can hold an account on the social media sites, such as Twitter, Facebook, YouTube, etc. The creator of WOM message is no longer completely non-commercial. The company can also encourage its employees to engage in the WOM communication. For example, Tony Hsieh, Zappos' CEO, highly encourages his employees to engage in social media and talk about Zappos. Therefore, I define WOM communication in a broad sense as online or offline communication regarding a brand, a product or a service. Specifically, it is a person-to-person communication, either in person or via technology, between a receiver and a communicator concerning a brand, a product or a service. I use word of tweet (WOT) to refer to a WOM tweet regarding a brand, a product, or a service on Twitter.

Litvin, Goldsmith, and Pan (2008) presented a conceptual model of the entire WOM process (Figure 2.1). They investigated the WOM process from motivations to sources to outcomes. I use this model to organize the literature, identify the missing pieces, and position my research in the WOM process.

Litvin, Goldsmith, and Pan (2008) presented the reasons that motivate people to spread WOM, including affects, altruism, self-interest, and reciprocation. Arndt (1967b) presented four reasons motivating consumers engaging in WOM communications including altruism, instrumental, interest and reduce cognitive dissonance. The consumers attempt to help other's making decisions; they can benefit from the communication and the benefit most likely irrelevant with the content of the message; they initiate the WOM conversation due to self-interest or ego-involvement; they attempt to change one's opinion or behavior. Mangold, Miller, and Brockway (1999) performed content analysis on a list of factors stimulating WOM communication and pointed out the top three factors including the communicator's perception on the receiver need, coincidental WOM communication relating to a broad subject, and a high level satisfaction or dissatisfaction with the product.

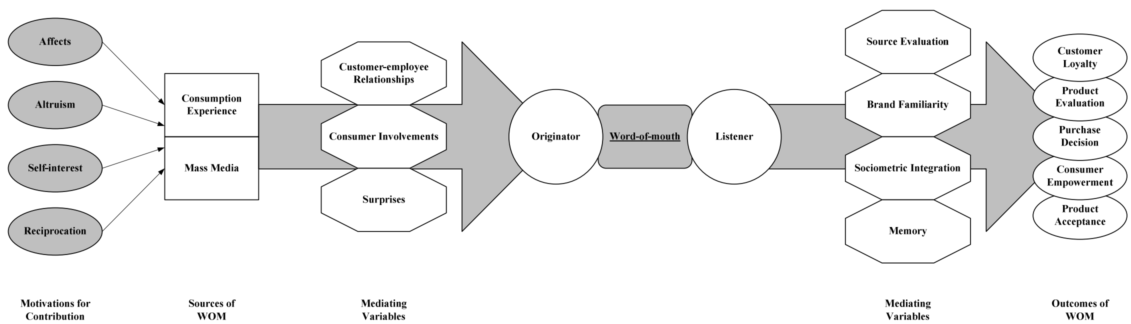


Figure 2.1 Conceptual model of the word-of-mouth process (Litvin, et al., 2008, p. 460)

Litvin, Goldsmith, and Pan (2008) presented that WOM messages usually came from one's own consumption experiences or mass media, which can influence customer loyalty, product evaluation, purchase decision, consumer empowerment, and product acceptance. The view on the source of WOM messages is limited. The WOM messages can come from marketers, businesses, or brands by initiating WOM communication among customers, encouraging employees to engage in WOM communication, or proactively generating WOM messages on social media sites themselves. Researchers also debated whether WOM message is an antecedent or a consequence of the consumer behavior (Godes & Mayzlin, 2004). Sun, Youn, Wu, and Kuntaraporn's (2006) hypothesized that WOM messaging is a driving force behind consumer behavior. It really depends on the content and context of the WOM communication. However, mentioning brand, product, or service is a productive way to increase brand awareness.

Litvin, Goldsmith, and Pan's (2008) model provides a great overview of WOM process. However, the research included in this framework has focused on the external aspects of WOM message, but fails to explore the internal aspects of WOM message. Two aspects of WOM communication process have attracted lots of attention, which are WOM communication channel and WOM message. WOM communication channel has been extended from offline to online. There has been research regarding how the online WOM communication differs from WOM in the traditional offline settings. Brown, Broderick, and Lee (2007) argued that existing offline theories on WOM communication are inappropriate for online WOM communication. Their work supported their argument and suggested that Websites are the primary actors in the online social networks and online communities provide a social proxy for individual identification. Therefore, different Websites mean different social proxies, which leads to different individual identification. It is critical to explore the dynamics of different social networks. Dellarocas (2003) differentiated online WOM networks with the traditional WOM networks by the following three aspects: (1) the large scale due to the Internet's low-cost and fast communication natures; (2) the ability of the system designers to manipulate the online WOM communications; and (3) challenges introduced by the unique characteristics of online communication, such as the anonymous nature of online identities and lacking contextual cues.

Moreover, not all WOM communication networks are equal (Allsop, et al., 2007, p. 399). It is critical to explore the dynamics of different social networks. There are different types of online channels to distribute WOM messages such as blogs (e.g. Blogger), micro-blogs (e.g. Twitter), multimedia sharing (e.g. Flickr, YouTube), reviews (e.g. Amazon, ePinions, Yelp), social bookmarking (e.g. Delicious), social networking (e.g. Facebook, LinkedIn), social news (e.g. Digg), and wikis (e.g. Wikipedia). It is interesting to explore how WOM messages are delivered on these different networks. In this dissertation, I study Twitter. Instead of studying the network at a macro level, I focus on the micro level of the Twitter network in this chapter. Twitter is a directional social network. The network is formed by the "follow" relationship, a dyadic relationship between two parties, which introduces the direction to the network and makes two parties in a communication process unequal.

Another aspect missing in this framework is the classification of WOM messages. Not all WOM messages are equal. I can classify the messages according to sentiment such as positive, negative, neutral, and no sentiment (Jansen, et al., 2009), which can advance the understanding of the consumers' attitudes toward the brand, product, and service. Haigie, Feick, and Price (1987) investigated WOM messages in the retail industry and classified WOM messages based on the evaluative dimensions of stores. They studied three groups including merchandise-related aspects (e.g., quality, pricing, and assortment), service-related aspects (e.g., quality in general and salesperson service), and pleasantness of shopping at a store. This research can advance the understanding about the content of WOM messages. Patti and Chen (2009) investigated WOM messages about credence-based services in the information-gathering process and classified WOM messages into three groups including service information gathering trigger and guidance, subjective personal experience, and personal advice. This research can advance the understanding about functions and origins of WOM messages.

On the research question in this chapter, I focus on the diffusion scope of WOT communication. Diffusion is an important characteristic and demonstrates the rippling feature of WOT. Some tweets are only read by the followers of the message creator in its concise form while some tweets are relayed from one person to another. Some tweets have the brand highlighted and have the potential to be accurately surfaced by searching for the branded term on the Twitter search engine. Some tweets in their 140-character form contain more than 140-character worth of information. I study the influence of the Twitter network on all these different types of WOT.

Research Question

Based on the review of prior work in the WOM communication area, my research question is "*What are the branding influences of social network on word-of-mouth communication?*"

Word-of-mouth communication is critical to businesses in an environment of online social networking. It is the ultimate marketing, branding and selling tool. However, previous research focuses on the motivation on the intrapersonal level (Litvin, et al., 2008; Mangold, et al., 1999; Sundaram, Mitra, & Webster, 1998), but it ignores factors on the contextual level. I agree with Dellarocas (2003) that online WOM communication can lack contextual information, which could be the reason why this topic has not been explored. But from social networks like Twitter, one context factor can be known, the relationship between account holders. The consumers' connection with the brand on Twitter can reflect the closeness of the relationship and the direction of information flow between these two parties, which can potentially influence consumers' WOT communications.

Independent variables

I operationalize social network connections on the micro level as the "follow" direction. Person A can follow person B to receive B's tweets. But, B does not necessarily have to follow back to receive A's updates. A is referred as B's follower on B's account and B is referred as A's following on A's account. Therefore, "follow" direction between brand and consumers has two levels including brand follows consumer (brand's following) and consumer (brand's follower) follows brand (Figure 2.2), which creates four different ways of consumers connecting to the brand: the consumer can follow the brand but the brand does not follow back; the brand can follow the consumer but the consumer does not follow back; the brand and the consumer follow each other; the brand and the consumer do not follow each other.

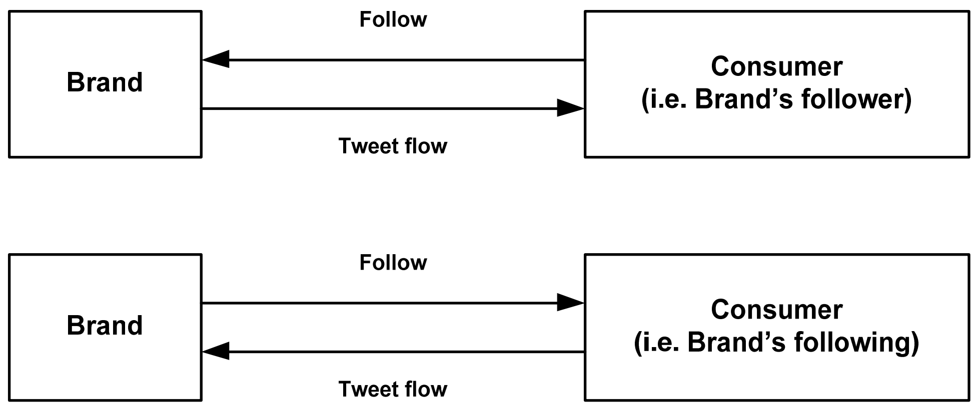


Figure 2.2 Brand and consumer relationships on Twitter

These different connection styles indicate different levels of relationship or social network ties due to asymmetrical communication between the brand and consumers, and the consumers' perceptions about being followed by a brand. It is very interesting to explore the influence of the connection style since this design ignores the common ground theory for online communication system. A brand can tweet but will not know if consumers are listening. Therefore, it is important to evaluate the "follow" influence.

Dependent variables

To address this research question, I identified 7 dependent variables (DV) with regards to the frequency of the following tweet types for WOT communication:

DV 1. WOT message: it measures WOM communication. All the messages regarding the brand are counted as WOM messages. The key to qualify as the WOT message is to mention the brand or product name in the message.

DV 2. WOT communication with other consumers (@others): it measures WOM communication in the traditional fashion, which is the interpersonal communication regarding the brand with the creator of the message perceived as being non-commercial.

DV 3. WOT communication with brand (@brand): it measures consumers' engagement with the brand account on Twitter.

DV 4. Retweet of WOT message from other consumers (RT(others)): one of the important features of WOM communication is information propagation. It used to be very hard to track the information flow in the traditional offline setting and other Web systems such as forum, review sites, etc. However, Twitter has the community norm that one should acknowledge the origins of the message when forwarding the message, which provides a way for the measurement of information propagation.

DV 5. Retweet of brand message (RT(brand)): it measures consumers' effort in propagating messages originated from the brand.

DV 6. WOT with brand hash-tagged (#brand): hash-tagging brand means the message is mainly about the brand. The hash-tag is particularly used to surface certain brand related messages in the search results by removing the ambiguity of the terms. Singhal from Google commented that hash-tags can be useful to maximize the exposure of a tweet (Talbot, January 13, 2010).

DV 7. WOT with hyperlinks (HTTP): one of the major functions of WOT is to channel traffic by including hyperlinks in the message. Consumers can be exposed to richer information or converted.

These variables capture the different branding and WOT communication aspects. I can view the variables from two angles: impact and relevance with the brand account (Table 2.1). Different types of tweets have different levels of impact. WOT communication with brand and WOT communication with other consumers are interpersonal communication between two individuals, so the tweets most likely influence these two parties. On the contrary, retweet, WOT with brand hash-tagged, and WOT with hyperlinks have much wider influence. Retweet is like a message evaluation action. Interesting and important tweets are more likely to be retweeted. This action can make WOT reach a wider audience and have higher impact. WOT with brand hash-tagged is also a high-impact type of tweet. Hash-tag sign can help remove the term ambiguity, label the tweet, and make the brand relevant tweets accurately surface in the result page on the Twitter search engine. It is also a way to show consumers' brand awareness. WOT with hyperlinks has high-impact on the content. It potentially gets people in touch with some information beyond 140 characters. A WOT message can be any type of the tweet above, so it has medium impact overall.

I can also classify WOT messages based on how relevant they are with the brand account, which shows the interaction between the consumer and the business on Twitter. WOT communication with brand shows the conversation from the consumer to the brand and the brand can check on these tweets fairly easily by clicking @brand (similar to @zhangmimi in Figure 1.2). Brand message retweet indicates the consumers' enthusiasm to spread the word and actively engage in an interaction process with the brand, which often shows the likeness to the brand, product or service. Therefore, these two types of tweets are relevant with branding and worth measuring.

Table 2.1 Dependent variables overview

	Relevant with Brand Account	Irrelevant with Brand Account
Low Impact	WOT communication with brand	WOT communication with other consumers
Medium Impact	WOT message	
High Impact	Retweet of brand message	Retweet of WOT from other consumers
	WOT with brand hash-tagged; WOT with hyperlinks	

Methodology

To address the research question in this chapter, I used Data Set 2 on the connection information of consumers linking with the brand Twitter account and Data Set 4 on all tweets mentioning the brand name in Twitter community. I conducted tweet analysis (i.e. content analysis of Twitter tweets). I labeled each tweet to enrich the tweet log based on the creator's relationship with the brand and the characteristics of the message. I employed statistical inference techniques to model the WOT communication. Below is the detailed description of the problem solving process.

Data processing

I collected 144,064 tweets in total from the five-week period. First, I labeled each tweet by the brand based on the occurrence of brand name in the tweet. I dropped 197 (0.14%) tweets since they were associated with two brands in the same tweet. Thus, I had 143,867 tweets from 96,716 individuals for analysis. Table 2.2 is an overview of the 5-week data I used in this research. It obviously shows that Twitter had growth on the user base, but the average number of branded tweet per user was about the same throughout the data collection period. The overall average number of tweets per user is 1.37 per week.

Table 2.2 Data overview

Period	**Tweet (*n*)**	**User (*n*)**	**Tweet per User**
05/05/2008 to 05/11/2008	8,301	5,636	1.47
08/11/2008 to 08/17/2008	10,287	7,605	1.35
11/17/2008 to 11/23/2008	17,332	12,439	1.39
02/23/2009 to 03/01/2009	34,910	26,411	1.32
05/25/2009 to 05/31/2009	73,037	54,970	1.33
Total	143,867	96,716	1.37 (average)

Second, I labeled the tweets by the following groups: WOT message, WOT communication with other consumers, WOT communication with brand, retweet of WOT from other consumers, retweet of brand message, WOT with brand hash-tagged, and WOT with hyperlinks (Table 2.3). 21.63% of these branded tweets occurred during a conversation with someone. 10.88% branded tweets could be viewed by the brand by going to the mention brand section on the interface. Only 1.13% of these branded tweets had brand names hash-tagged. The consumers generated 2.29% tweets about the brand retweeted by others, which is 2.75 times of the brand message retweet. 15.02% branded tweets contained hyperlinks.

Table 2.3 Breakdown of different types of tweet

Dependent Variable	n	%
WOT	143,867	
@others	31,112	21.63%
@brand	15,657	10.88%
RT (others)	3,292	2.29%
RT (brand)	1,192	0.83%
#brand	1,619	1.13%
HTTP	21,602	15.02%

Third, I labeled the sender of the tweet by his/her connection direction with the brand: follower of the brand or not; following of the brand or not (Tables 2.4, 2.5, and 2.6). Among all the individuals in my data set, about 80% of consumers talking about the brands did not connect with the brands and 20% of consumers only connected with one and only one brand (Table 2.4). Only a small proportion of consumers connecting with the brand actually engaged with the brand and the proportion ranged from 0.40% to 14.01% (Table 2.5). In most cases, consumers talking about the brand and connecting with the brand had the identities as the followers and followings of the brand (Table 2.6).

Table 2.4 Breakdown of consumers by number of brand they engage with

Brand (*n*)	Follower (*n*)	Follower (%)	Follower (% within Group)	Both (*n*)	Both (%)	Both (% within Group)	Following (*n*)	Following (%)	Following (% within Group)
0	75,131	77.68%		79,297	81.99%		77,715	80.35%	
1	20,593	21.29%	95.40%	16,565	17.13%	95.10%	18,058	18.67%	95.04%
2	895	0.93%	4.15%	770	0.80%	4.42%	853	0.88%	4.49%
3	87	0.09%	0.40%	77	0.08%	0.44%	83	0.09%	0.44%
4	10	0.01%	0.05%	7	0.01%	0.04%	7	0.01%	0.04%
Total	96,716	100.00%	100.00%	96,716	100.00%	100.00%	96,716	100.00%	100.00%

Table 2.5 Breakdown of active followers and followings within brands

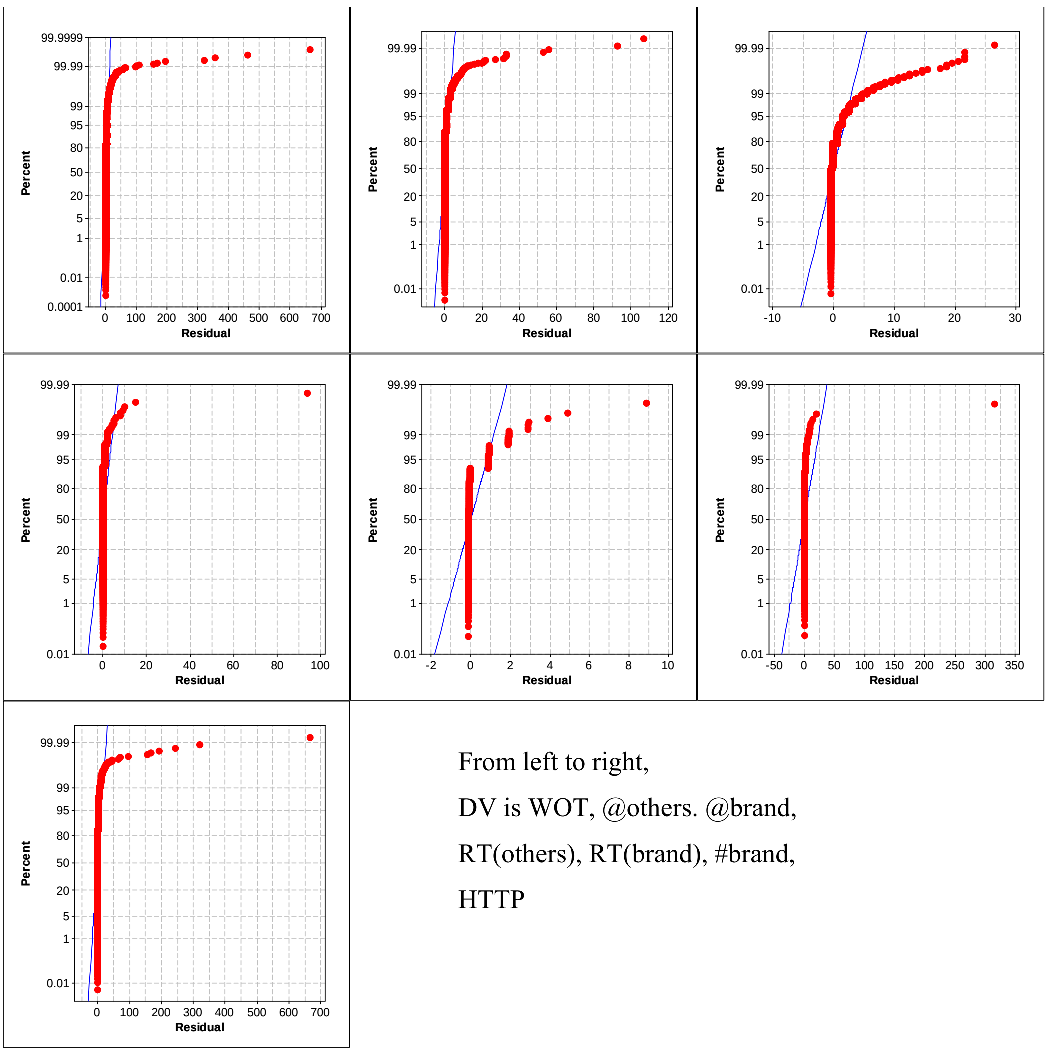
Brand	Total Follower (*n*)	Active Follower (*n*)	Active Follower (%)	Total Both (*n*)	Active Both (*n*)	Active Both (%)	Total Following (*n*)	Active Following (*n*)	Active Following (%)
Coffee Groundz	6,124	327	5.34%	5,647	325	5.76%	5,741	326	5.68%
Comcast	20,867	1,014	4.86%	20,288	1,011	4.98%	20,659	1,022	4.95%
Home Depot	8,241	198	2.40%	6,683	156	2.33%	7,436	183	2.46%
H&R Block	3,021	23	0.76%	2,753	23	0.84%	2,788	23	0.82%
Kogi BBQ	26,644	959	3.60%	1,171	164	14.01%	1,346	179	13.30%
NakedPizza	3,475	237	6.82%	1,322	133	10.06%	3,325	151	4.54%
Starbucks	198,152	12,570	6.34%	136,533	10,082	7.38%	141,886	10,675	7.52%
Whole Foods	724,284	2,892	0.40%	423,703	2,391	0.56%	474,495	2,726	0.57%
Zappos	691,458	4,464	0.65%	372,721	4,079	1.09%	417,235	4,756	1.14%

Table 2.6 Breakdown of engaging followers and followings within brands

Brand	Total Follower (*n*)	Unique Follower (*n*)	Unique Follower (%)	Both (*n*)	Both (% in Follower)	Both (% in Following)	Total Following (*n*)	Unique Following (*n*)	Unique Following (%)
Coffee Groundz	327	2	0.61%	325	99.39%	99.69%	326	1	0.31%
Comcast	1,014	3	0.30%	1,011	99.70%	98.92%	1,022	11	1.08%
Home Depot	198	42	21.21%	156	78.79%	85.25%	183	27	14.75%
H&R Block	23	0	0.00%	23	100.00%	100.00%	23	0	0.00%
Kogi BBQ	959	795	82.90%	164	17.10%	91.62%	179	15	8.38%
NakedPizza	237	104	43.88%	133	56.12%	88.08%	151	18	11.92%
Starbucks	12,570	2,488	19.79%	10,082	80.21%	94.44%	10,675	593	5.56%
Whole Foods	2,892	501	17.32%	2,391	82.68%	87.71%	2,726	335	12.29%
Zappos	4,464	385	8.62%	4,079	91.38%	85.77%	4,756	677	14.23%

Statistical analysis method: Bootstrap-based nonparametric analysis of variance

The research question in this chapter could be addressed by using two-way analysis of variance (ANOVA) to test hypotheses on the 7 dependent variables, assuming the error was normally distributed. However, the usage data is not normally distributed; it follows a power law distribution like most of the Web usage data (Raban & Rabin, 2009), meaning the error is unlikely to be normally distributed, which is true for my data based on the testing (Figure 2.3). Therefore, I cannot simply calculate *F* statistics and use *F* table to determine the *P* values. I cannot transform data to make it normal by using Box-Cox transformation method like Raban and Rabin did (2009), since the data is very far from being normal and severely skewed.



From left to right,
DV is WOT, @others. @brand,
RT(others), RT(brand), #brand,
HTTP

Figure 2.3 Normal plots of residual achieved by performing ANOVA test

Instead, I adopted an approach called bootstrap-based nonparametric analysis of variance (BNANOVA) (Zhou & Wong, Forthcoming). Through bootstrap, I can determine the distribution of *F* statistic that is achieved from my data. Then, based on this distribution, I can determine the *P* value.

According to Zhou and Wong (Forthcoming), the possible models for a two-way factorial experiment include:

$$\text{Model 1: } y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

$$\text{Model 2: } y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

$$\text{Model 3: } y_{ijk} = \mu + \alpha_i + \varepsilon_{ijk}$$

$$\text{Model 4: } y_{ijk} = \mu + \beta_j + \varepsilon_{ijk}$$

$$\text{Model 5: } y_{ijk} = \mu + \varepsilon_{ijk}$$

Note: y_{ijk} (k=1, …, n_{ij}) represents the response and k represents the subscript for replicate. μ is the overall mean effect. α_i (i=1, …, I) is the effect of the ith level of row factor A, β_j (i=1, …, J) is the effect of jth level of column factor B, $(\alpha\beta)_{ij}$ is the interaction effect between α_i and β_j, ε_{ijk} is a random error component. $\sum_i \alpha_i = 0$; $\sum_j \beta_j = 0$; $\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$

So in this chapter y_{ijk} is the frequency of different WOT messages. Factors A and B are the brand's follower and following, respectively.

Mean squares for treatments are estimated in the following ways for balanced design (same number of observations under each combination of the experimental conditions) and unbalanced design (different numbers of observations under each combination of the experimental conditions) (Montgomery, 2008; Searle, Casella, & McCulloch, 1992; Zhou & Wong, Forthcoming):

$$MSE = \frac{1}{N - IJ}\sum_i\sum_j\sum_k (y_{ijk} - \bar{y}_{ij.})^2 ;\; N = \sum_i\sum_j n_{ij}$$

For balanced design:

$$MSAB = \frac{k}{(I-1)(J-1)}\sum_i\sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$MSA = \frac{kJ}{I-1}\sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$$

$$MSB = \frac{kI}{J-1}\sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2$$

For unbalanced design:

$$x_{ij} = \bar{y}_{ij.} = \frac{\sum_{k=1}^{n_{ij}} y_{ijk}}{n_{ij}} ;\; \bar{x}_{i.} = \frac{\sum_j x_{ij}}{I} ;\; \bar{x}_{.j} = \frac{\sum_i x_{ij}}{J} ;\; \bar{x}_{..} = \frac{\sum_i\sum_j x_{ij}}{IJ}$$

$$MSAB_u = \frac{1}{(I-1)(J-1)}\sum_i\sum_j (x_{ij} - \bar{x}_{i.} - \bar{x}_{.i} + \bar{x}_{..})^2$$

$$MSA_u = \frac{J}{I-1}\sum_i (\bar{x}_{i.} - \bar{x}_{..})^2$$

$$MSB_u = \frac{I}{J-1}\sum_j (\bar{x}_{.j} - \bar{x}_{..})^2$$

Note: *u* in the footnote denotes unbalanced design.

The algorithm for testing interaction effect by using BNANOVA follows these steps (Montgomery, 2008; Searle, et al., 1992; Zhou & Wong, Forthcoming):

Step 1: Calculate the observed $F_{Interaction} = MSInteraction / MSE$ statistic . $MSInteraction$ can be $MSAB$ for the balanced design and $MSAB_u$ for the unbalanced deign.

Step 2: Calculate observed errors $e_{ijk} = y_{ijk} - \bar{y}_{ij.}$

Step 3: Sample n_{ij} errors e^*_{ijk} from e_{ijk} for each *i*th level of row and *j*th level of column with replacement.

Step 4: $y^*_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + e^*_{ijk}$ where $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$ are the least square estimates from Model 2, calculate $F^*_{Interaction}$

Step 5: Repeat Step 3 and Step 4 a large number of times b (for example, b=10,000 in my study) to get $F^{(1)*}_{Interaction}$, $F^{(2)*}_{Interaction}$, …, $F^{(b)*}_{Interaction}$

Step 6: Find out the percentage of $F^*_{Interaction}$ which are greater than or equal to the observed $F_{Interaction}$ statistic. This percentage is *P* value.

The algorithm for testing on main effect by using BNANOVA follows these steps (Montgomery, 2008; Searle, et al., 1992; Zhou & Wong, Forthcoming):

Step 1: Calculate the observed $F_{Main} = MSMain / MSE$ statistic . $MSMain$ can be MSA or MSB for the balanced design and MSA_u or MSB_u for the unbalanced design

Step 2: Calculate observed errors $e_{ijk} = y_{ijk} - \bar{y}_{ij.}$

Step 3: Sample n_{ij} errors e^*_{ijk} from e_{ijk} for each *i*th level of row and *j*th level of column with replacement

Step 4: $y^*_{ijk} = \hat{\mu} + \hat{\alpha}_i / \hat{\beta}_j + e^*_{ijk}$ where $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$ are the least square estimates from Model 3 or 4, calculate F^*_{Main}

Step 5: Repeat Step 3 and Step 4 a large number of times b (for example, b=10,000 in my study) to get $F^{(1)*}_{Main}$, $F^{(2)*}_{Main}$, …, $F^{(b)*}_{Main}$

Step 6: Find out the percentage of F^*_{Main} which are greater than or equal to the observed F_{Main} statistic. This percentage is *P* value.

Results

The linear regression is run to analyze the relationship between the amounts of tweets the individual user sent before and after connecting to the business on Twitter. This analysis is performed to understand the potential differences of before- and after-connection behaviors. Table 2.7 presents the descriptive statistics for the four variables in the analysis included: Branded Tweet Volume from Follower, Tweet Volume after Becoming Follower, Branded Tweet Volume from Following, and Tweet Volume after Becoming Following. For the first two variables used to test the follower influence, almost all the consumers (17,727/19,224=92.21%) only sent out no more than 3 branded tweets. For the latter two variables used to test the following influence, almost all the consumers (14,769/16,197=91.18%) also sent out no more than 3 branded tweets. Therefore, it is reasonable to treat consumers sending out more than 3 branded tweets as the outliers, which can be the high-leverage and distort the relationship of before- and after-connection behavior.

The slopes between tweet volume and tweet volume after connecting with the brand for each individual are almost 1 and the intercepts are small enough to be ignored, which indicate if the consumers connect with the brand, almost all of their branded tweets are created after the connection (Table 2.8 and Figure 2.4). Therefore, there is no need to differentiate before- and after-connection behavior for consumers with a brand connection. There is almost no before-connection behavior for consumers with brand connection. I rule out the case that the consumers connecting with the brand are originally active in generating branded tweets before they connect with the brand. Almost all of this active information generating behavior happens after they connect with the brand. It shows the potential influences of connecting with brand on the within-subject level.

Table 2.7 Five-number summary of variables in regression, mean, standard deviation, and sample size

Variable	Min	Q1	Median	Q3	Max (Adjusted)	Max	Mean	SD[a]	Sample Size
Branded Tweet Volume from Follower	1	1	1	2	3	356	1.77	3.58	19224
Tweet Volume after Becoming Follower	1	1	1	2	3	356	1.65	3.48	19224
Branded Tweet Volume from Following	1	1	1	2	3	356	1.87	4.17	16197
Tweet Volume after Becoming Following	1	1	1	2	3	356	1.71	3.87	16197

a: SD=Standard Deviation

Table 2.8 Regression

Predictor	Response	Slope	*P* Value[a]	Intercept	*P* Value[a]	R^2
Tweet Volume after Becoming Follower	Tweet Volume	0.98	<0.01*	0.08	<0.01*	0.81
Tweet Volume after Becoming Following	Tweet Volume	0.98	<0.01*	0.10	<0.01*	0.78

*a: * denotes that P value is less than 0.05*

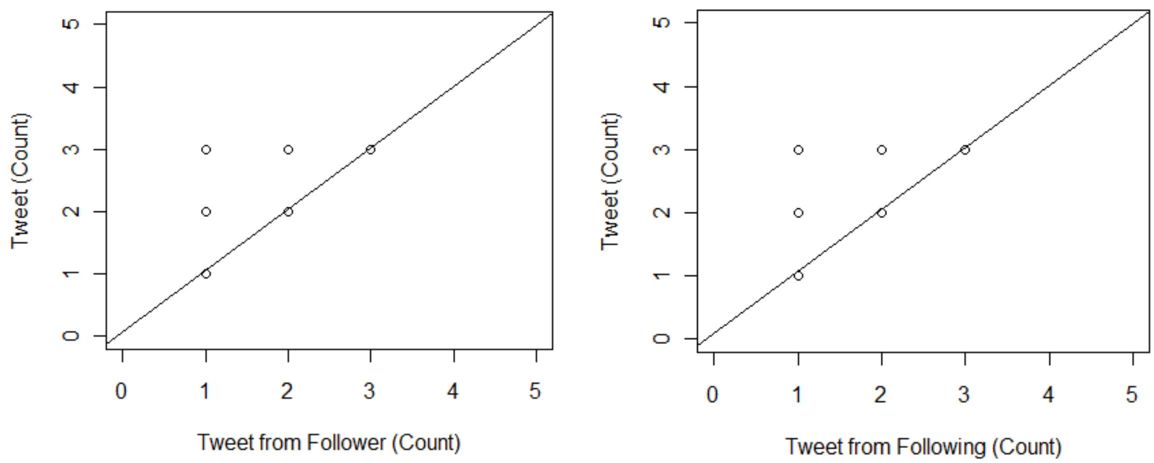


Figure 2.4 Regression plot

The results of BNANOVA test are presented in Table 2.9. Appendix B shows the distributions of *F* values. There is no significant interaction effect of follower and following on WOT message frequency (*F* value<0.001, *P* value=0.66). But follower (*F* value=0.001, *P* value=0.05) and following (*F* value=0.010, *P* value<0.01) have statistically significant influence on the number of WOT messages sent out by the consumers. In addition, consumers as the brand's follower or following will send out more messages about the brand (Table 2.10 and Figure 2.5).

For the two types of interpersonal WOT communication, I have similar results (Table 2.9). There is no significant interaction effect of follower and following on WOT communication frequency with other consumers (*F* value<0.001, *P* value=0.31) and WOT communication frequency with brand (*F* value=0.001, *P* value=0.12). But follower (*F* value=0.008, *P* value<0.01) and following (*F* value=0.009, *P* value<0.01) have the main effect with statistical significance on WOT communication frequency with other consumers. Follower (*F* value=0.005, *P* value<0.01) and following (*F* value=0.023, *P* value<0.01) also have the main effect with statistical significance on WOT communication frequency with the brand. Moreover, consumers who are the brand's follower or following tend to communicate more about the brand, product, or service with other consumers or the brand (Table 2.10 and Figure 2.5).

However, follower and following have no influence with statistical influence on the number of retweets of WOT messages from other consumers or the brand, WOT messages with brand hash-tagged, and WOT messages with hyperlinks (Table 2.9). All of the *P* values are much larger than 0.05.

Table 2.9 BNANOVA results

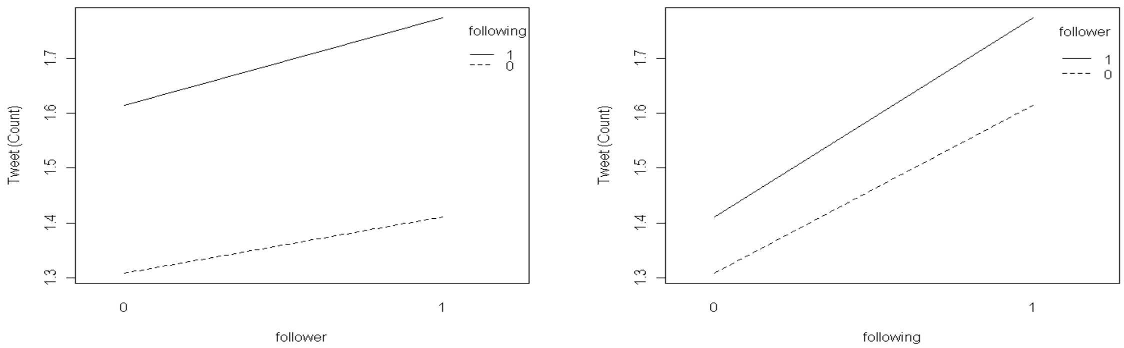
Dependent Variable	Follower		Following		Follower×Following	
	F Value	*P* Value[a]	F Value	*P* Value[a]	F Value	*P* Value[a]
WOT	**0.001**	**0.05***	**0.010**	**<0.01***	<0.001	0.66
@others	**0.008**	**<0.01***	**0.009**	**<0.01***	<0.001	0.31
@brand	**0.005**	**<0.01***	**0.023**	**<0.01***	0.001	0.12
RT (others)	0.005	0.32	<0.001	0.79	0.003	0.38
RT (brand)	0.001	0.86	0.008	0.52	0.001	0.86
#brand	0.001	0.50	<0.001	0.81	<0.001	0.75
HTTP	0.003	0.20	0.002	0.37	<0.001	0.72

*a: * denotes that P value is less than 0.05*

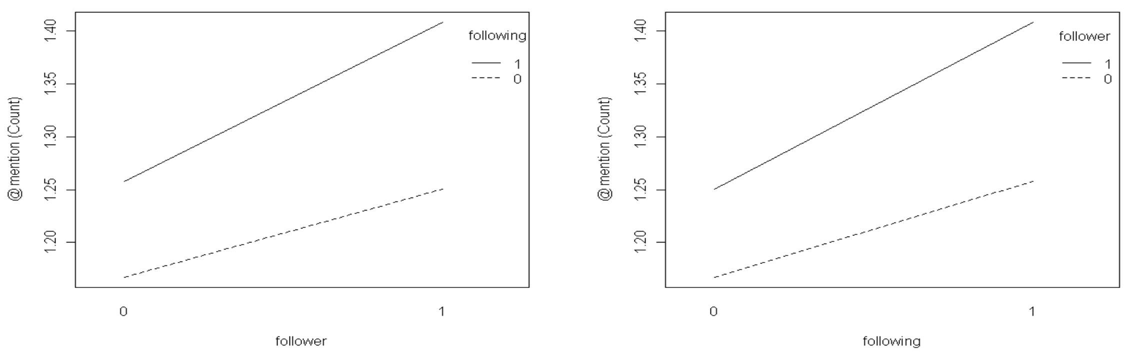
Table 2.10 Average tweet number by sender's relationship with brand

Group	WOT	@others	@brand	RT (others)	RT (brand)	#brand	HTTP
Non-follower, non-following	1.31	1.17	1.22	1.12	1.07	1.70	1.51
Non-follower, following	1.61	1.26	1.36	1.04	1.12	2.00	1.87
Follower, non-following	1.41	1.25	1.27	1.14	1.09	1.50	1.16
Follower, following	1.77	1.41	1.50	1.26	1.13	1.46	1.38

WOT message (WOT)



WOT communication with other consumers (@others)



WOT communication with brand (@brand)

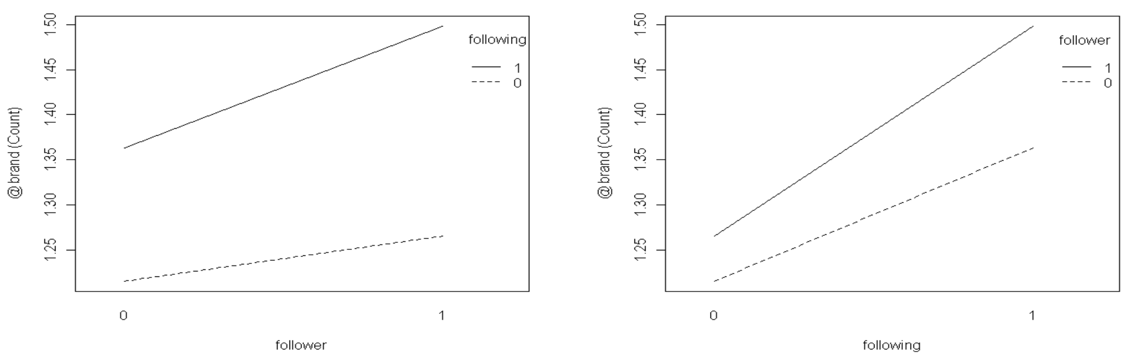


Figure 2.5 Interaction plots

Discussion

In this chapter, I investigated the potential differences on seven types of word-of-tweet (WOT) — WOT message, WOT communication with other consumers, WOT communication with brand, retweet of WOT from other consumers, retweet of brand message, WOT with brand hash-tagged, and WOT with hyperlinks — among consumers having different connections with the brand on Twitter, which include brand's follower or not and/or brand's following or not. To address the research question in this chapter, I performed tweet analysis by using linear regression and bootstrap-based nonparametric analysis of variance correspondingly.

I found that on the within-subject level, WOT messages are sent after the consumers connect with the brand, becoming the brand's follower and/or the brand's following, if they have connection with the brand. I had similar findings on the between-subject level, being the brand's follower and brand's following have main effect on the number of WOT messages with statistical significance, but their interaction has no statistical significant influence on WOT messages. This result can simply be understood as having connection with the brand, either being a follower or following, means more WOT messages. Therefore, it is worthwhile for a brand to recruit as many interested consumers to follow the brand's account, and it is also worthwhile for the brand to follow those potential consumers who express interest in the brand.

In addition, I also found the same influence of being the brand's follower and brand's following, having main effect but not interaction effect, on the frequency of WOT communication with other consumers and the brand. However, being the brand's follower and brand's following has no statistically significant influence, no main effect, and no interaction effect on the number of retweeting of WOT messages from other consumers and the brand, the number of WOT with brand hash-tagged, and the number of WOT with hyperlinks.

WOT message, WOT communication with other consumers, and WOT communication with the brand are three types of WOT messages that I list in Table 2.1 as having low and medium impacts because they reach fewer people compared with retweets and WOT messages with brand hash-tagged or they are less likely to expose people with richer information compared with WOT with hyperlinks. One way to understand the association is that Twitter is a communication platform, meaning its major function is to facilitate interaction. It is understandable that people interact more or try to interact more with people connecting with them. It is similar to our behavior in the physical space. We typically interact more with our next door neighbor than people living down the street or a friend in another city. Thus, consumers connecting with the brand interact more with the brand. Companies should connect with consumers on Twitter.

But what is more interesting here is that consumers connecting with the brand also interact with other consumers more on the branded topic. More importantly, almost all of the communication happens after the connection with the brand. This connection can be understood as creating some sort of cues in consumers' mind through the connection. The consumers, who are brand's followers, can receive tweets from the brand once in a while. This can be viewed as a mind-sharing process, during which a set of abstract associations are created (Holt, 2004). It increases consumers' brand awareness. This can also be viewed as a relationship building process, during which the consumers feel the emotional side of the brand (Holt, 2004) and the brand becomes anthropomorphized. It will increase consumers' brand attachment. Therefore, companies should utilize this channel and send out polished branded messages, which will benefit brand and customer management.

On the other hand, being the brand's following, the consumers' tweets can be received by the brand. The brand can know the consumers' day-to-day activity and thinking. Therefore, to the consumers, their tweeting about the brand can become a potential rewarding action, making profit from the high status individual like business (Thibaut & Kelley, 1975). Imagining from the consumers' perspective, it could be very rewarding to the consumers if they perceive that their tweets can be received by the brand about being brand's loyal customer, compliments towards the brand, and recommendations on improving brand. Thus, companies should follow back to the consumers if the consumers express the interest in following them, and in the meantime, they can reach out to follow consumers who have potential interest in the brand.

Even though I did not show that connecting with the brand is relevant with retweeting of WOT from other consumers, retweeting of brand message, WOT with brand hash-tagged, and WOT with hyperlinks, these WOT communication types are still popular and prevalent on Twitter. These actions may be more relevant with the Twitter network rather than the dyadic relationship between the consumer and the brand.

Researchers have debated whether WOM message is a reason or an outcome of the consumption related behavior (Godes & Mayzlin, 2004). My research has shown that connection with the brands happens first and then WOT communications start, and not the other way. Thus, connections definitely motivate the WOT communication to a certain extent. It would be interesting to study why there are people talking about the brand on Twitter but who do not connect with the brand. Except some intuitive answers like not wanting to receive spam from the brand, what would be some other interesting reasons?

My takeaways from this research question on the commercial businesses using Twitter include that the brand should connect with consumers on Twitter and allow them to choose their most comfortable connection fashion. The connections on Twitter enhance both the management of consumers having direct connections to the business through possibly increasing the brand awareness, strengthening brand attachment, and rewarding consumers on the emotional and material levels.

I highly recommend the brand follow the consumers on Twitter given that following (P value<0.01) has much more influence than follower (P value=0.05) on the number of WOT messages (Table 2.9). It creates an emotional bond in the consumers' mind between consumers and brand. It also brings the brand closer to the consumers from the consumers' perspective.

I do not recommend that businesses buy followers or purchase an existing Twitter account with a large number of followers and followings. In such cases, the business seemingly has a large community but it will not be an engaging community for the brand. It is a useless community for the business. On TV, the consumers do not have the option to choose to watch the commercial or not and which commercial to watch. But in the Internet age, marketers or businesses have less control over information (Weber, 2009). De Chernatony (2000) recommends adopting a looser form of brand control and a more relaxed brand management approach. On social networks, the consumers can select the brands that are interesting for them to follow. This mechanism helps to form a brand community with people having interest in the brand. Consumers' interest is very important. It is the reason for which receiving messages from the brand does not cause annoyance among consumers, and the communication between consumers and brand can be more effective. Thus, the business should have a clear goal on its Twitter account and polish their tweet crafting skills. Tweet content is the ultimate driving force of the community formation.

Considering consumers connecting with the brand are more involved in WOT communication, they are very likely to belong to the loyal customer group. Moreover, the branded experience is about "actively engaging consumers" (Schmitt, 2009, p. 4), since advertising does not just mean grabbing a "30-second spot" (Schmitt, 2009, p. 5) and it can be employing the digital branded experience to deliver the message, which converts the consumer "from passive reactors to advocates" (Schmitt, 2009, p. 11). The company can utilize the connection to promote their business and brand by engaging these consumers or distributing ads. They can increase their influence and form deeper impressions compared with mass communication channels, such as email and television.

Conclusion

I concluded from this research question that being the brand's follower and/or following can motivate consumers on the frequency of talking about the brand, discussing the brand with other consumers, and communicating with the brand on Twitter. I showed that connection with consumers on Twitter means more than just a simple connection. It actually means these groups of people are more likely to be involved in the WOM communication than people who are not connecting with the brand will be. My work provides insight on the analytics of social network.

Summary

Word of mouth marketing as a branding media is undergoing considerable change due to the increasing use of micro-communication services. Of all these services, Twitter is the most popular one, with a lot of press attention and a rapidly growing user base. In this chapter, I investigated the potential differences of seven types of word-of-tweet (WOT) communication among consumers having different connections with the brand on Twitter, which include consumers that are the brand's followers or not and/or consumers that are the brand's followings or not. The seven types of WOT communication investigated are WOT message, WOT communication with other consumers, WOT communication with the brand, retweet of WOT from other consumers, retweet of brand message, WOT with the brand hash-tagged, and WOT with hyperlinks. To address the research question in this chapter, I performed tweet analysis by using linear regression and bootstrap-based nonparametric analysis of variance correspondingly. I found that being the brand's follower and brand's following have main effect but not interaction effect on the volume of WOT messages, WOT communication with other consumers and the brand. However, they have no statistically significant influence, no main effect, and no interaction effect on the number of retweet of WOT messages from other consumers and the brand, the number of WOT with brand hash-tagged, and the number of WOT with hyperlinks. The implications for brands include that they should connect with the consumers on Twitter, particularly follow the consumers; they should be aware that the consumers connecting with them are the hardcore group, and thus offer great potential in using the Twitter channel in terms of engaging consumers, promoting business, and managing brands.

CHAPTER 3. WHAT "TWEETS" AROUND COMES AROUND: BUSINESS ENGAGEMENT IN ONLINE WORD-OF-MOUTH COMMUNICATION ON TWITTER

Online word-of-mouth (oWOM) communication can have huge influences on commercial businesses. Research has shown that exposure to oWOM messages can generate more interest in product category than can exposure to information produced by marketers (Bickart & Schindler, 2001). OWOM messages have strong influences over online brand trust (Ha, 2004). Moreover, oWOM messages highly correlate with companies' sales (Bharati & Tarasewich, 2002, May; Lleti, Ortiz, Sarabia, & Sanchez, 2004). Therefore, commercial businesses need to consider oWOM messages when developing and managing their advertising, branding, and marketing strategies.

Moreover, successful commercial businesses must go beyond simply being aware of or taking into consideration oWOM messages and instead must engage in the oWOM communication process as both communication initiators and active participants. While noting differences between online and offline environments, de Chernatony (2000) highlights the importance of interacting with consumers, engaging consumers in the branding process, and participating in consumers' conversations. The Internet empowers consumers to gain access to information sources, enabling them to be the "active co-producers of value" (de Chernatony, 2000, p. 191). Building on this idea, my research in this chapter focuses on how a business can engage in oWOM communication, how engagement creates consumers' oWOM communication, and how consumers respond to the business' engagement.

Jansen, Zhang, Sobel, and Chowdhury (2009) researched word-of-mouth communication on Twitter and found that about one fifth of all tweets contain the name of a brand, product, or service. Among these WOM tweets, about one fifth express some sentiments. More than half of the branded tweets with sentiments are positive tweets, and only one third of them are negative tweets. Their study showed that the linguistic structure of tweets is similar to the linguistic patterns of natural language expressions. They concluded that Twitter is a potentially rich WOM venue for companies to explore as part of their overall branding strategy. It is a key application in the attention economy and a competitive intelligence source.

My research extends the work of Jansen, Zhang, Sobel, and Chowdhury (2009) by focusing on the interaction aspect of business and consumer engagement in the oWOM communication platform. I am particularly interested in studying oWOM communication from the business' perspective and investigating overall oWOM communication based on the engagement of business, the exchange of oWOM between businesses and consumers, and the consumers' reaction to a business' engagement. It brings insight on how active businesses should be on Twitter. Should they engage as often as they can, at least once a day, or whenever they have news to release? In addition, can a business' engagement cause consumers' to engage in oWOM communication on Twitter? If so, in what These are some of the questions which motivate my research reported in this chapter.

In the remainder of this chapter, I review WOM communication with the focus on how companies can manage WOM communication. I present the model tested in this project and explain the approach to tackle the problem. Then I report the results by using path analysis and conclude by discussing research and managerial implications.

Managing Word-of-mouth Communication

With the introduction of the Internet, forum, online review, and social media services, traditional WOM communication, defined as "oral, person-to-person communication between a perceived non-commercial communicator and a receiver concerning a brand, a product, or a service offered for sale" (Arndt, 1967a, p. 190), has been evolving into oWOM communication, defined as "any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the internet" (Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004, p. 39). OWOM communication is perceived as spreading faster, reaching out to a larger audience, and having deeper and wider influences when compared with traditional WOM communication. Companies, then, must find ways to harness the potential of oWOM.

There are several ways that companies can possibly develop WOM communication management strategies. The company can manage consumers' WOM communications by leveraging the motivations for consumers to become involved in

WOM communication. Sundaram, Mitra, and Webster (1998) identified eight motives for consumers to engage in WOM communication. They further differentiated factors motivating consumers in positive WOM communication, including altruistic, product involvement, and self-enhancement reasons from factors motivating consumers in negative WOM communication including altruistic, anxiety reduction, vengeance, and advice seeking reasons. Overall, motives to engage in WOM communication are significantly related to consumers' consumption experiences. Building on the work of Balasubramanian and Mahajan (2001), Hennig-Thurau, Gwinner, Walsh, and Gremler (2004) developed one of the most comprehensive frameworks of factors motivating people to express themselves and engage in WOM communication on the Web. Their framework basically identifies five utilities motivating consumers to engage in oWOM communication. The first is the focus-related utility, which consumers receive when making contributions to the community, especially those with added value. The second, consumption utility, refers to consumers obtaining value through direct consumption of other consumers' contributions. In contrast, approval utility comes when one's contributions are consumed or approved by other consumers. Moderator-related utility is achieved when a third party makes the complaint act easier because consumers hope the platform operator will serve as an intermediary between them and the company. The last one is homostate utility, which is a balanced state that individuals strive to restore after they lose the original equilibrium according to the balance theory (Hennig-Thurau, et al., 2004, p. 44). After a satisfying or dissatisfying consumption experience, consumers can restore their balance by expressing either positive or negative sentiment toward the brand.

The company can influence oWOM via its own marketing channel. Keller (2007) argued that traditional media and marketing channels still drive oWOM. Roughly 50% of branded conversations include a reference to some kind of media or marketing that is consumed by at least one of oWOM communication participant. These media and marketing references include advertising, editorial and programming from various types of media, company websites, and marketing materials at the point of purchase, coupons, and other promotions.

The company can manage WOM communication by playing different roles in the communication process. Godes and fellow researchers (Godes, et al., 2005) described four WOM management strategies for business: the company can be an observer, moderator, mediator, or participant in the WOM communication. As an observer of WOM communication, the company only collects information and learns the ecosystem. It can know how its consumers think about it and what its competitors are doing. As a moderator, the company goes beyond listening to actually foster the conversation. It usually realizes the moderator role by establishing a platform to allow consumers to exchange information or adopting a customer recommendation program. As a mediator, the company takes control of the oWOM message and disseminates it by itself. It tries to manipulate the communication content and channel. The company can be more active and serve as a participant in the WOM communication directly. Social media sites like Twitter can enable the company to play this role. In this research, I argue that business participation in WOM communication can also be a driving factor for WOM communication, which can be viewed as an approach for business to manage WOM communication.

My model. Given that commercial businesses own brand presences on Twitter, I postulate that businesses' tweets can be viewed as media and marketing materials, the driving factor for oWOM according to Keller (2007). Due to proximity to the source, business tweets are likely to be the major driving factor for oWOM on Twitter. The business plays a role as active oWOM communication initiator or participant in the process as shown in (Godes, et al., 2005).

Because businesses establish their Twitter accounts, they are able to mange oWOM communication. Businesses try to form a brand community by getting consumers to follow it and, in turn, by following the consumers. Broadcasting tweets is an important part of this process. The more tweets a company sends out and the more consumers it follows, the bigger impact it has in the Twitter community. The impact is reflected by greater brand awareness and increased number of followers; more consumers connecting to it learn more about the brand from the tweets and have more things to share about the brand. These consumers, in turn, influence other consumers who are connected with them but who are not connected to the brand. These relationships are modeled in Figure 3.1.

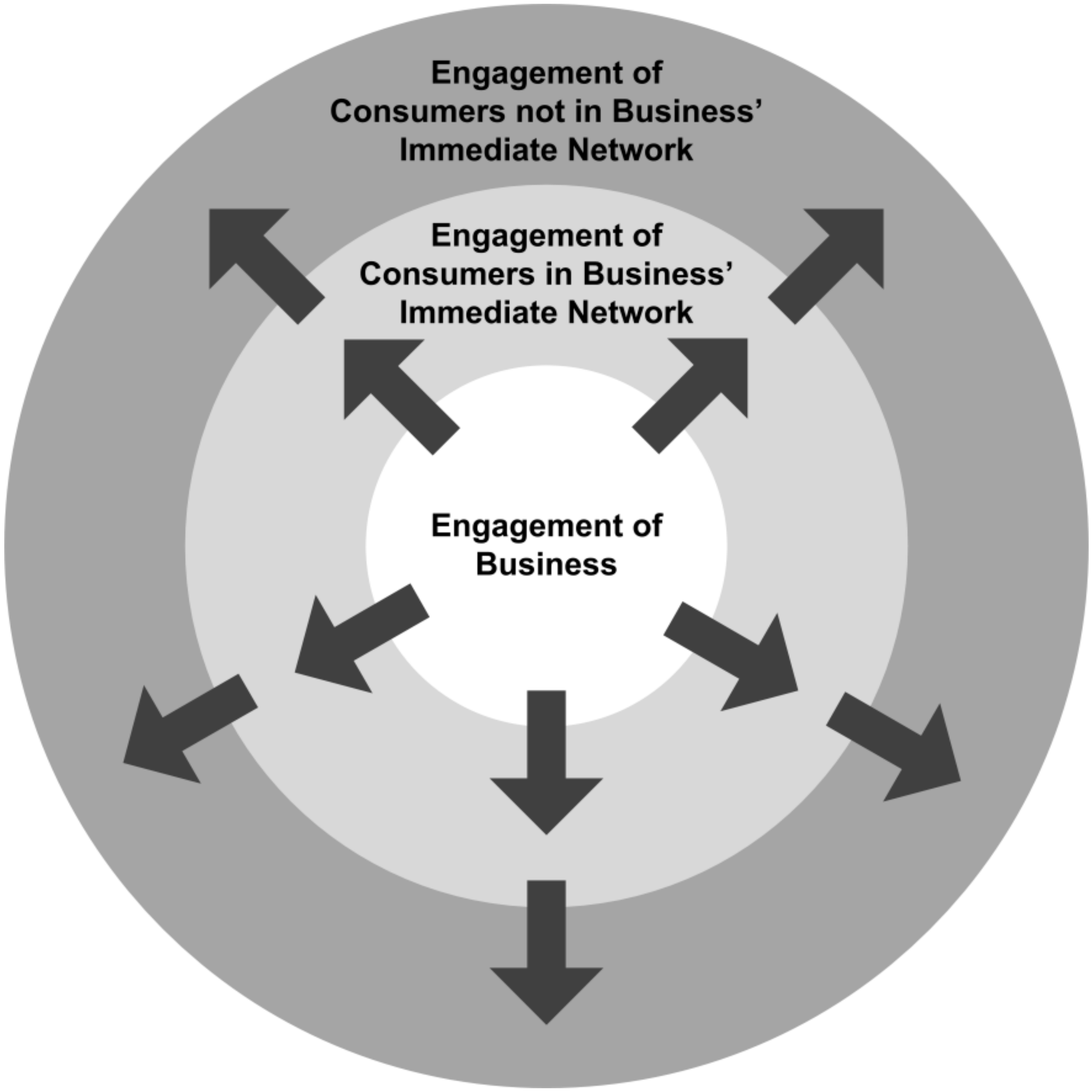


Figure 3.1 Business engagement in oWOM communication

Research Question

The study in this chapter is motivated by previous research on different strategies companies use to manage oWOM communication, and thus, my research question is “*what are the influences of brand engagement in online word-of-mouth communication on consumers’ level of engagement in word-of-mouth communication on Twitter?*”

Internet and social network sites are changing the way oWOM messages diffuse and enable businesses to participate actively in oWOM communication. However, there is general lack of empirical research on businesses as participants in managing oWOM communication. Understanding the business role as oWOM participant has strategic meaning to management. It brings insight on how influential the business can be on Twitter, how proactive the business should be in oWOM communication process, and how many resources the business should allocate on oWOM advertising.

There are two ways to address my research question in this chapter and operationalize my model (Figure 3.1). One way is to assume that business engagement can create cues and associations in consumers’ minds, that it is a mind-sharing process (Holt, 2004). From this perspective, consumers tend to engage more in the oWOM communication as they become more aware of the brand. A second way to address my research question is to stick strictly to the explicit influence of business engagement by investigating consumers’ behavior of retweeting messages from business Twitter accounts. Retweeting is the action of forwarding one’s tweet with the acknowledgment of its source in the message

Using the first approach to my research question, I can measure business engagement on Twitter by how active the business is in sending tweets by estimating tweet frequency (*WOM(Brand)*). I can also evaluate how active the business is in listening to consumers and in trying to understand them by measuring the number of consumers it follows (*Brand's Following Number*), which may be the precursor for delivering high quality oWOM messages.

The consumers' engagement is evaluated on two levels based on the distance of the relationships consumers have with the brand. The immediate influence the brand can have is on consumers with direct connection to it, either the brand's followings or followers. So one way is to measure how actively these consumers participate in the oWOM communication process by branded tweet frequency (*WOM(Consumers Connecting to Brand)*). Another way to operationalize consumers' engagement is to measure the number of a brand's followers (*Brand's Follower Number*), which shows the intention of the consumers to receive up-to-date brand information. Since Twitter is a social network, brand influence can permeate through the layer of consumers connecting to it and then penetrate through the network and in turn potentially affect all the consumers in Twitter. Therefore, another level to operationalize consumer engagement is to measure the branded tweet frequency among consumers having no connection with the brand (*WOM(Consumers Not Connecting to Brand)*).

The descriptions above and the model (Figure 3.1) tested in this research can be specified by the following path equations and demonstrated by path model (Figure 3.2):

$$\begin{aligned} &WOM(ConsumersNotConnectingToBusiness) \\ &= a_1 + b_{11}WOM(Business) + b_{12}BusinessFollowingNumber \\ &+ b_{13}WOM(ConsumersConnectingToBusiness) \\ &+ b_{14}BusinessFollowerNumber + e_1 \end{aligned} \tag{1}$$

$$\begin{aligned} &WOM(ConsumersConnectingToBusiness) \\ &= a_2 + b_{21}WOM(Business) + b_{22}BusinessFollowingNumber + e_2 \end{aligned} \tag{2}$$

$$BusinessFollowerNumber \\ = a_3 + b_{31}WOM\,(Business) + b_{32}BusinessFollowingNumber + e_3 \tag{3}$$

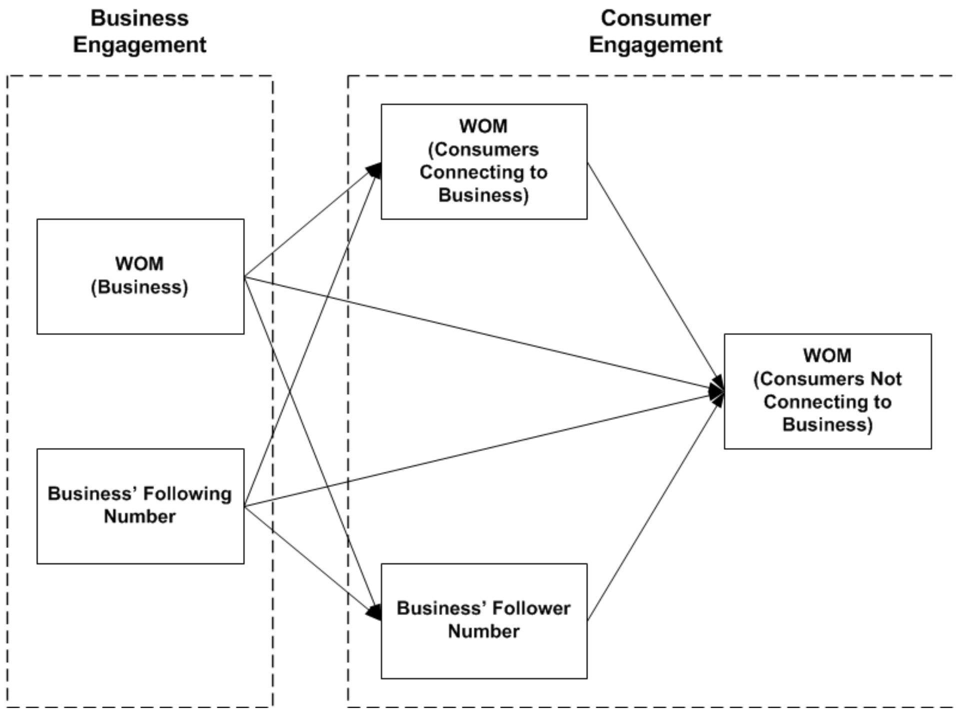


Figure 3.2 Business engagement in oWOM communication

In the second approach to address my research question, I measure how tweets from the brand diffuse throughout the Twitter network. The business sends out tweets from its Twitter account. Some of the messages are retweeted by consumers who received the tweets directly from the company. Those retweeted messages are subsequently read and retweeted by consumers connecting to consumers having direct connection with the brand. In such a fashion, consumers pass along the messages, potentially spreading them throughout the whole Twitter community. I model this process to evaluate the explicit influence of business engagement in the oWOM communication process.

Methodology

In order to answer my research question in this chapter, I used Data Set 1 of all tweets from the brand, Data Set 2 with regards to the follower and following information of the brand, and Data Set 4 on all tweets mentioning the brand name on Twitter. I performed path analyses to investigate the relationship between all variables and assessed the contribution of each predictor on the overall oWOM communication in the Twitter community. I also investigated trajectory of retweeting the message from the brand.

Data processing

All the data were processed at the week level since previous research found a strong weekly pattern of Twitter usage. Users tweet more in the middle of the week and less during weekend and the beginning of the week (Jansen, et al., 2009). Therefore, it is comparatively stable and more comparable to measure by week. Thus, I measured, by week, the number of tweets from the brand from Data Set 1 (*WOM(Brand)*) and the number of the brand's new followers and followings from Data Set 2 (*Brand's Follower Number*, *Brand's Following Number*).

For all the tweets mentioning these nine brands during 5-week study period in Data Set 4, I differentiated the tweets based on whether the sender of the message connects to the brand, which can be following the brand and/or letting the brand follow him/her. I then categorized the tweets as those sent by consumers connecting to the brand (*WOM(Consumers Connecting to Brand)*) and those sent by consumers having no connection to the brand (*WOM(Consumers Not Connecting to Brand)*). These tweet volumes are summarized at the week level.

My data, however, is not multivariate normal (Figure 3.3), but rather it has a power law distribution. I transformed data to the normal distribution (Figure 3.4) via the Box-Cox power transformation, using lg(variable+1).

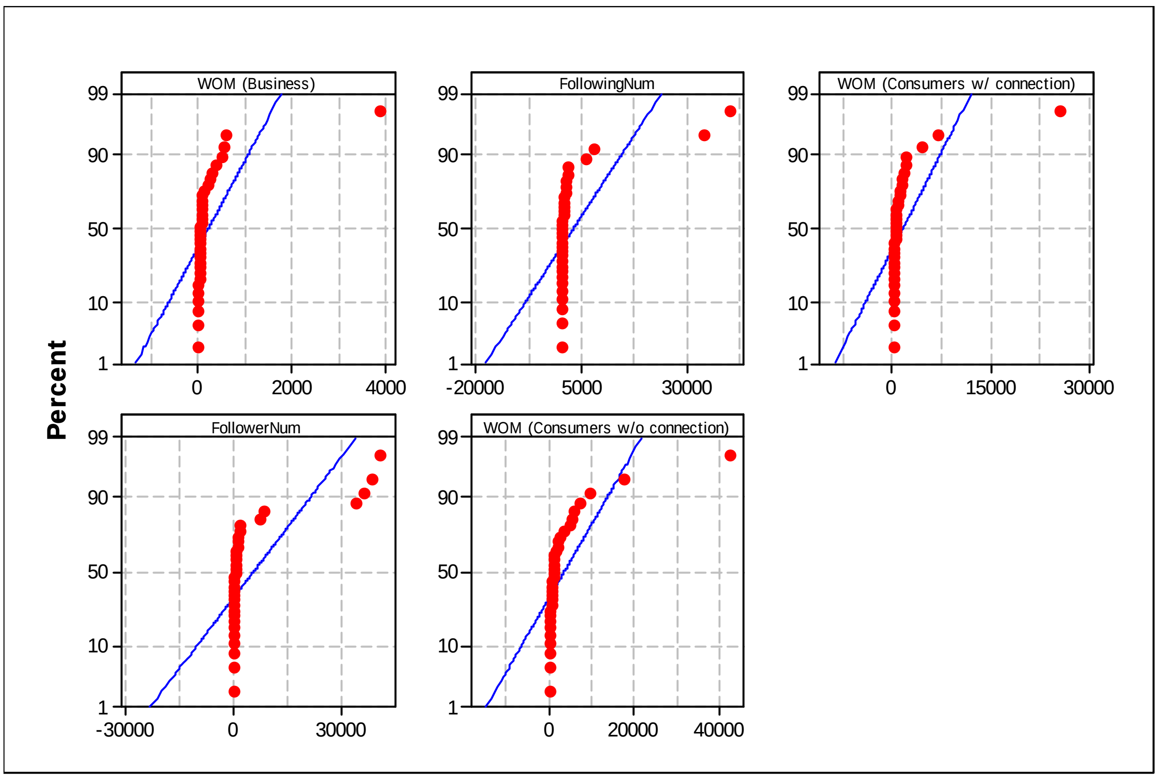


Figure 3.3 Normal probability plots before data transformation

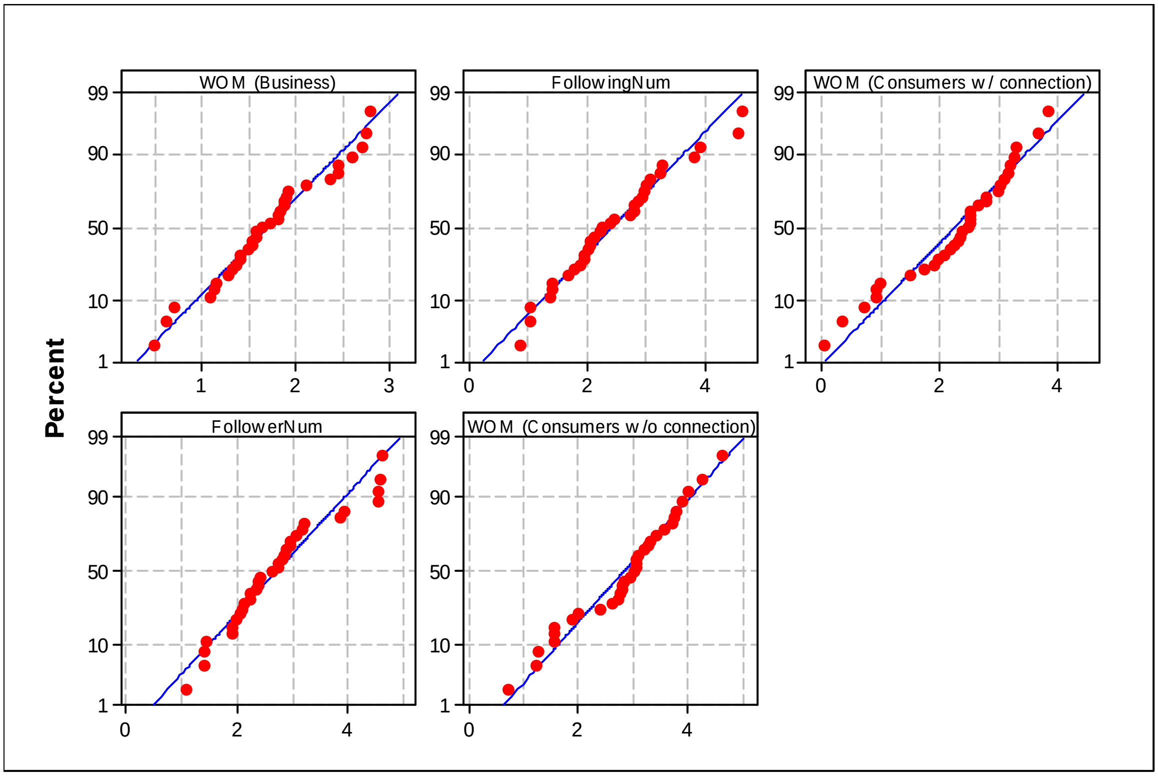


Figure 3.4 Normal probability plots after data transformation

Retweet formats have many different variations. In order to guarantee the rigor of my research methods and validity of my research results, I manually evaluated all the tweets sent by consumers and picked out the messages originating from the brand.

Path analysis

The major statistical approach I used to test my model (Figure 3.2) is path analysis, "a statistical technique that uses both bivariate and multiple linear regression techniques to test the causal relations among the variables specified in the model" (Olobatuyi, 2006, p. 32). I want to emphasize that path analysis itself can only show the existence of correlation. Causal relationship can be indicated on top of correlation based on other data and/or theoretical supports. One specialty of path analysis is that it reveals the direct and indirect effects that predictor variables have on responding variables (Olobatuyi, 2006).

Results

Descriptive analysis of brand's followers and followings

Tables 3.1, 3.2, and 3.3 summarize the follower and following information by May 31, 2009, for all of the nine brands I studied. Given that all of the brands I studied are popular and well-known in the Twitter community, the odds of consumers connecting with multiple brands is very high. However, Table 3.1 tells us that about 84% of these consumers with connections to the brands on Twitter connected to one and only one brand as the brand's follower and/or following, and approximately 14% connected to 2 brands. Therefore, consumers generally connect with a small handful of brands rather than a large number of brands, which indicates the special preference consumers have for the brands they connect to.

Table 3.1 Breakdown of followers and followings by number of brands they connect to

Number of Brand	Follower (*n*)	Follower (%)	Both (*n*)	Both (%)	Both (% within Follower)	Both (% within Following)	Following (*n*)	Following (%)
1	**1,173,504**	**83.25%**	**685,244**	**84.01%**	**58.39%**	**90.75%**	**755,099**	**83.78%**
2	208,893	14.82%	112,445	13.78%	53.83%	89.24%	125,996	13.98%
3	20,540	1.46%	13,263	1.63%	64.57%	89.18%	14,873	1.65%
4	4,757	0.34%	3,378	0.41%	71.01%	89.58%	3,771	0.42%
5	1,355	0.10%	984	0.12%	72.62%	91.36%	1,077	0.12%
6	463	0.03%	328	0.04%	70.84%	91.88%	357	0.04%
7	97	0.01%	70	0.01%	72.16%	85.37%	82	0.01%
8	12	0.00%	1	0.00%	8.33%	50.00%	2	0.00%
Total	1,409,621	100.00%	815,713	100.00%	57.87%	90.51%	901,257	100.00%

Table 3.2 presents the detailed breakdown of followers and followings within the brand. The follower/following ratios for eight out of nine brands were around 1, but the ratio for Kogi BBQ was about 20. Interestingly, 60% to 98% of consumers with connections to Coffee Groundz, Comcast, Home Depot, H&R Block, Starbucks, Whole Foods, and Zappos were both their followers and followings. For Kogi BBQ, roughly 95% of followers were just their followers. Naked Pizza had about 40% of consumers with connections to it as both the brand's followers and followings. Thus, for most brands, the consumers connecting to them are both their followers and followings, with follower/following ratios very close to 1.

Table 3.2 Breakdown of followers and followings within brand

Brand	Coffee Groundz	Comcast	Home Depot	H&R Block	Kogi BBQ	NakedPizza	Starbucks	Whole Foods	Zappos
Total Follower (n)	6,124	20,867	8,241	3,021	26,644	3,475	198,152	724,284	691,458
Unique Follower (n)	477	579	1,558	268	25,473	2,153	61,619	300,581	318,737
Unique Follower (%)	7.79%	2.77%	18.91%	8.87%	**95.61%**	**61.96%**	31.10%	41.50%	46.10%
Overlap (n)	5,647	20,288	6,683	2,753	1,171	1,322	136,533	423,703	372,721
Overlap Follower (%)	**92.21%**	**97.23%**	**81.09%**	**91.13%**	4.39%	38.04%	**68.90%**	**58.50%**	**53.90%**
Overlap Following (%)	**98.36%**	**98.20%**	**89.87%**	**98.74%**	**87.00%**	39.76%	**96.23%**	**89.30%**	**89.33%**
Total Following (n)	5,741	20,659	7,436	2,788	1,346	3,325	141,886	474,495	417,235
Unique Following (n)	94	371	753	35	175	2,003	5,353	50,792	44,514
Unique Following (%)	1.64%	1.80%	10.13%	1.26%	13.00%	**60.24%**	3.77%	10.70%	10.67%
Follower Following Ratio	1.07	1.01	1.11	1.08	19.79	1.05	1.4	1.53	1.66

Table 3.3 shows the breakdown of the consumers connecting to the brand based on whether they were a brand's follower first or following first. If the consumer connects with the brand before the brand reaches out to the consumer, it means the consumer initiates the connection and is engaging with the brand. If it is the other way, it means the brand inclines to connect with the consumer and is active in the oWOM communication.

Among the brands' followings, about 90% of the consumers requested to follow the brands first. Only a very small portion (1.58%) was followed by the brands first. About 10% of the consumers among followings were just the brands' followings, and they did not follow the brand. Among the brands' followers, about 60% of the consumers requested to follow the brands first. Only 1% of consumers who were followed by the brands first then followed back to the brand. About 40% of the consumers among followers were just the followers and did not have the brands follow them. Therefore, most of the time, consumers initiate the connection with the brand, and the brand follows back to more than half of the consumers requesting to follow it.

Table 3.3 Breakdown of consumers by the order of becoming brand's follower and/or following

Order of Becoming Brand's Follower/Following	*n*	%
Follower First	953,786	88.73%
Following First	17,035	1.58%
Just Following	104,090	9.68%
Total Following	**1,074,911**	**100.00%**
Follower First	953,786	56.70%
Following First	17,035	1.01%
Just Follower	711,445	42.29%
Total Follower	**1,682,266**	**100.00%**

To sum up, the majority of consumers connecting to the brand connect to one and only one brand. Most of the time, consumers who connect to the brand do so as both the brand's followers and followings. Most of the brands have a balanced number of followers and followings. In addition, connecting to the brand is predominately the consumers initiating activity. This tells that the brand community on Twitter is a tight community with loyal brand advocates.

Path analysis

Table 3.4 shows the oWOM communication volumes outside of the brand's immediate social network increased dramatically immediately after the businesses launched their Twitter accounts. The increases of weekly oWOM message count range from 40% to 25,300%.

Table 3.4 OWOM volumes before and after

the businesses launched branded Twitter accounts

Brand	OWOM Volume 1-Week Before Business Twitter Account Launched[a]	OWOM Volume 1-Week After Business Twitter Account Launched[a]	%
Coffee Groundz	8	32	400.00%
Home Depot	356	577	162.08%
Kogi BBQ	16	553	3,456.25%
Naked Pizza	2	508	25,400.00%
Starbucks	4,148	5,949	143.42%
Whole Foods	451	830	184.04%

a: The oWOM messages are from consumers without connection to the branded Twitter account.

Table 3.5 presents the descriptive statistics of all variables for path analysis including means and standard deviations before and after transformation. On average by week, the businesses sent out 117.21 tweets and followed 2,988.48 consumers. The consumers connecting to the brand sent out 775.79 tweets mentioning the name of the brand they connected to. There were 5,389.30 consumers who started following brand. Among those consumers not linking with the brand in the Twitter community, there were 3,432.61 tweets mentioning one of the nine brands researched in this study. After transformation, the variables follow multivariate normal distribution.

Table 3.5 Descriptive statistics of variables

Variable	Mean	Standard Deviation	Minimum	Maximum
Before transformation				
WOM (Brand)	117.21	166.59	2	604
Brand's Following Number	2,988.48	8,931.20	6	39,977
WOM (Consumers Connecting to Brand)	775.79	1,409.67	0	6,912
Brand's Follower Number	5,389.30	12,304.21	11	40,822
WOM (Consumers Not Connecting to Brand)	3,432.61	7,820.87	4	42,189
After transformation				
WOM (Brand)	1.70	0.60	0.48	2.78
Brand's Following Number	2.41	0.95	0.85	4.60
WOM (Consumers Connecting to Brand)	2.25	0.95	0	3.84
Brand's Follower Number	2.71	0.95	1.08	4.61
WOM (Consumers Not Connecting to Brand)	2.81	0.94	0.70	4.63

Table 3.6 presents the Pearson correlations between all transformed variables. Most of the correlations are significant except two pairs as expected, which are the correlations between the WOM messages from the brand and the brand's follower number (γ=0.08) and the brand's following number (γ=0.17). The largest correlation is between brand's follower number and the amount of WOM from consumers connecting to the brand with statistical significance at the level of 0.01 (γ=0.80). The second largest correlation is between brand's following number and brand's follower number with statistical significance at the level of 0.01 (γ=0.78). Brand's following number also has statistically significant correlation with the amount of WOM from consumers connecting to the brand at the level of 0.01 (γ=0.77), which is the third largest correlation. The volume of WOM messages from consumers having no connection with the brand correlates with statistical significance with all four predictors: the amount of WOM

messages from the brand (γ=0.35, *p* value<0.05), brand's following number (γ=0.51, *p* value<0.01), the amount of WOM messages from consumers connecting to the brand (γ=0.66, *p* value<0.01), and brand's follower number (γ=0.62, *p* value<0.01). The correlation between the amounts of WOM messages from the brand and the consumers connecting to the brand is significant at the level of 0.05 (γ=0.42). In addition, the scatterplot matrix (Figure 3.6) shows the obvious linear relationships between all the variables except for those two pairs with insignificant correlations. However, the path analysis will provide more insight on the relationships between variables in the model.

Table 3.6 Correlations among variables

Variable	**1**	**2**	**3**	**4**	**5**
1. WOM (Brand)	1.00	0.17	0.42*	0.08	0.35*
2. Brand's Following Number	0.17	1.00	0.77**	0.78**	0.51**
3. WOM (Consumers Connecting to Brand)	0.42*	0.77**	1.00	0.80**	0.66**
4. Brand's Follower Number	0.08	0.78**	0.80**	1.00	0.62**
5. WOM (Consumers Not Connecting to Brand)	0.35*	0.51**	0.66**	0.62**	1.00

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

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

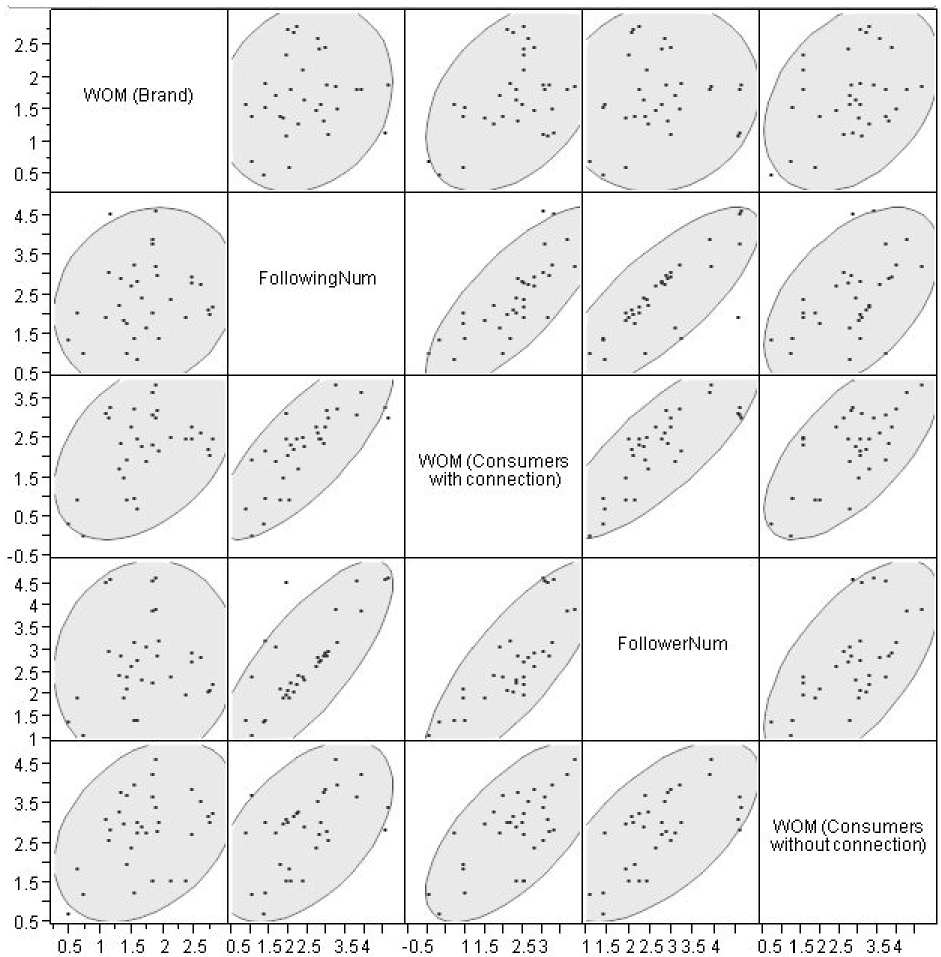
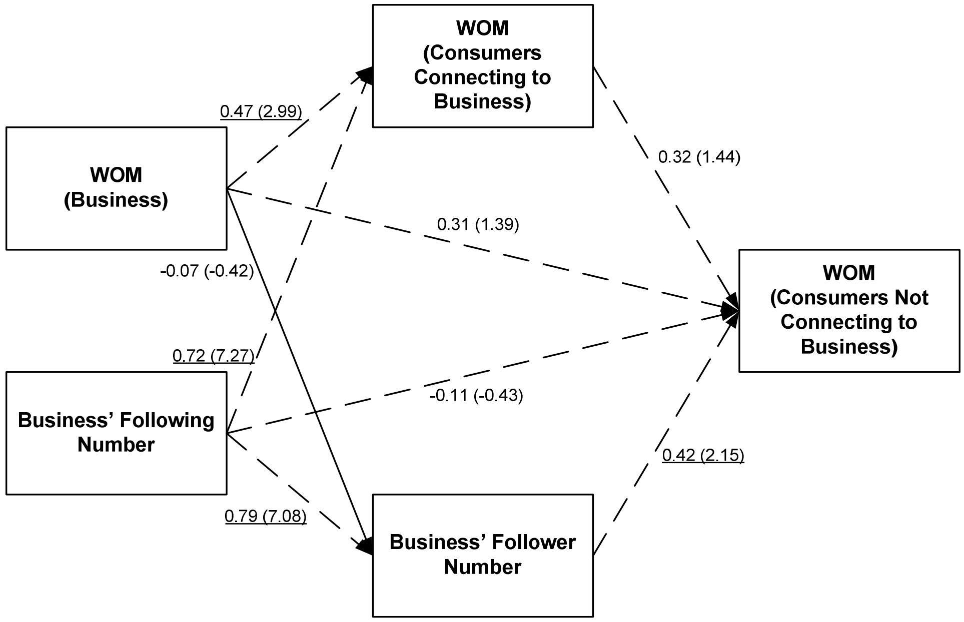


Figure 3.5 Scatterplot matrix

I conducted the path analysis employing maximum likelihood in Amos 18 developed by SPSS Inc. The full path model with estimated path coefficients is presented in Figure 3.6. According to Byrne (2010), the test statistic used in Amos is the critical ratio (C.R.), which is the parameter estimate divided by its standard error. Therefore, it operates as a z-statistic in testing. At the level of 0.05, the test statistic needs to be larger than 1.96 or smaller than -1.96 in order to reject the null hypothesis. Thus, there are four statistically significant relationships.

The follower number has a significant direct effect on the amount of WOM from consumers having no connection with the brand (*b*=0.42, C.R.=2.15). The following number has a significant direct effect on brand's follower number (*b*=0.79, C.R.=7.08) and the amount of WOM messages from consumers connecting to the brand (*b*=0.72, C.R.=7.27). The amount of WOM messages from the brand has a significant effect on the amount of WOM messages from consumers connecting to the brand (*b*=0.47, C.R.=2.99), but it does not have a significant effect on brand's follower number (*b*=-0.07, C.R.=-0.42). The brand's following number does not have a significant effect on the amount from consumers having no connection with the brand either (*b*=-0.11, C.R.=-0.43). Interestingly, both the amounts of WOM messages from the brand (*b*=0.31, C.R.=1.39) and the consumers connecting to the brand (*b*=0.32, C.R.=1.44) have no significant effect on the amount of WOM messages from the consumers not connecting to the brand at the level of 0.05, but the *p* values are around 0.15.



Note: Dashed line denotes the Pearson correlation between the two variables is statistically significant. Solid line denotes the Pearson correlation between the two variables is not statistically significant. For the results of path analysis, the value not in the parentheses is the coefficient and the value in the parentheses is the critical ratio. The value underlined denotes the statistical significance of the coefficient at the level of 0.05.

Figure 3.6 Full path model tested with path coefficients

Table 3.7 presents the direct, indirect, and total effects of all 4 predictor variables on the volume of WOM message from consumers not connecting to the brand. The amount of WOM messages from the brand has about 2.5 times stronger direct effect (0.31) than the indirect effect (0.12). The amount of WOM messages from consumers connecting to the brand has a much bigger direct effect (0.32) and indirect effect (<0.01).

The brand's following number has a larger indirect effect (0.57) than direct effect (-0.11), whereas the brand's follower number has a much larger direct effect (0.42) than indirect effect (<0.01). This explains why, in the path model, the coefficient for the following number is not statistically significant while the follower number is statistically significant. The brand's following number influences the amount of WOM message not connecting to the brand via the brand's follower number.

Table 3.7 Effects of predictors variables

on WOM (Consumers Not Connecting to Brand)

Variable	Direct Effect[a]	Indirect Effect[a]	Total Effect
WOM (Brand)	**0.31**	0.12	0.43
Brand's Following Number	-0.11	**0.57**	0.46
WOM (Consumers Connecting to Brand)	**0.32**	<0.01	0.32
Brand's Follower Number	**0.42**	<0.01	0.42

a: Value in bold denotes the stronger effect between direct effect and indirect effect

Retweet

I had 1,142 tweets retweeted by consumers, which were originally from 243 tweets sent by the brand, 0.95% (243/25,601=0.95%) of all tweets sent by the brand. I had 5 retweet styles in my data, but RT and via were the predominate styles accounting for 96.93% of all the retweet (Table 3.8).

Table 3.8 Retweet style

Style	*n*	%
RT	1,013	88.70%
Via	94	8.23%
Retweeting	19	1.66%
Retweet	14	1.23%
R/T	2	0.18%
Total	1,142	100.00%

For the retweet frequency (Table 3.9 and Figure 3.7), 50% were retweeted no more than twice, 75% were retweeted no more than 4 times, and the majority were retweeted no more than 8 times. But the retweet frequency distribution has a long tail. The maximum retweet frequency is 77 times in my data. The average retweet frequency is 4.70 times. Thus, it is very rare that the tweets from a brand become viral; selecting messages to retweet is a highly personalized behavior. Table 3.10 presents the top retweeted tweets. I manually evaluated the theme of these tweets and classified them into 6 groups: humorous (n=9), anecdotal (n=5), philanthropic (n=4), news (n=3), philosophical (n=2), and promotional (n=2). Zappos is the major contributor of these top retweeted messages (n=10), most of which are humorous messages. Starbucks and Whole Foods also have the second largest amount of top retweeted messages (n=5), most of which are anecdotal for Starbucks and philanthropic for Whole Foods.

Table 3.9 Five-number summary of retweet frequency, mean, and standard deviation

Minimum	**Q1**	**Median**	**Q3**	**Maximum (Adjusted)**	**Maximum**	**Mean**	**Standard Deviation**
1	1	2	4	8	77	4.70	10.37

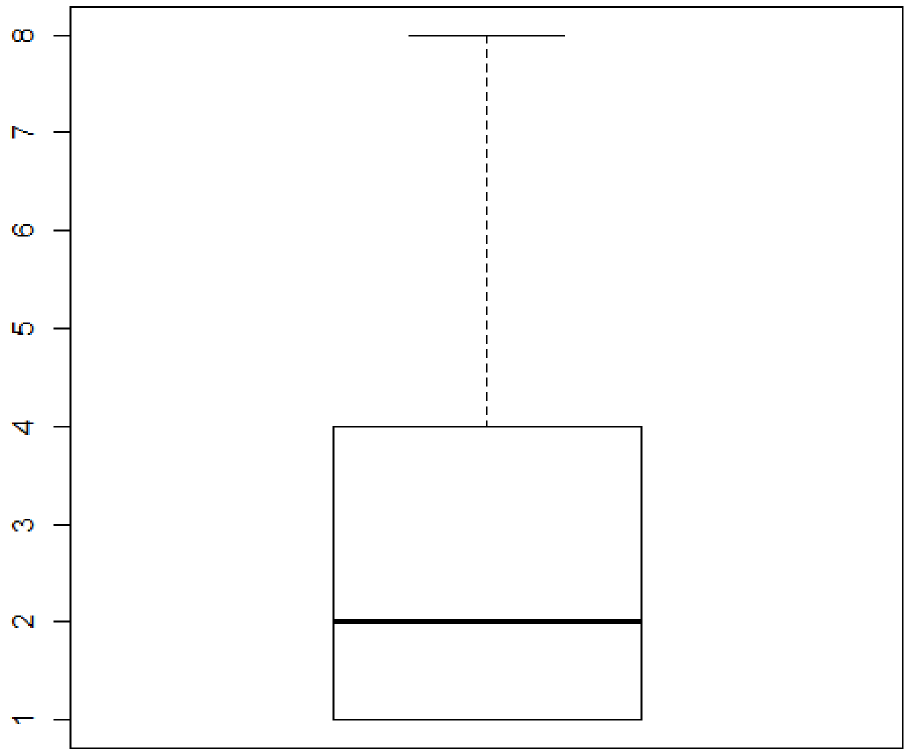


Figure 3.7 Boxplot of retweet frequency (without outliers)

Table 3.10 Top retweeted tweets

Rank	*n*	Sender	Tweet	Type
1	77	@WholeFoods	#twitterforfood Skip a meal June 1st and donate the savings to world hunger. http://tr.im/m1Pq	Philanthropic
2	77	@zappos	Trying to reduce my email inbox is like trying to lose weight. The number always seems to creep back up to where it was before.	Humorous
3	74	@zappos	Anonymous donor giving @lancearmstrong foundation $25k when @LIVESTRONGCEO "hits" 25k followers. Hope he doesn't actually hit me	Humorous
4	55	@zappos	Those who can laugh without cause have either found the true meaning of happiness or have gone stark raving mad. -N Papernick	Philosophical
5	47	@zappos	I've been wondering about this for awhile... Now I finally know who moved my cheese (thanks @missrogue): http://bit.ly/zcheese	Humorous
6	36	@zappos	Proper etiquette when you see clothes on a stranger w/ tag hanging out? Somehow I don't think "Tag! You're it!" is appropriate.	Humorous
7	34	@zappos	Dropped my laptop on floor this morning. I usually drop my phone, so good to know I'm moving on to bigger and better things.	Humorous
8	26	@Starbucks	RT @RGreenberg: http://twitpic.com/64z5h - Take a look at this @starbucks in Paris. Can you believe it? --> looks like a great store	Anecdotal
9	25	@kogibbq	oh, dear lord: http://tinyurl.com/qp2yew	Anecdotal

10	24	@Starbucks	Here they are in action ... I stayed out of the way. http://twitpic.com/623yo	Anecdotal
11	22	@CoffeeGroundz	If you are a local musician or in a band; let us know if you would like to be apart of a full day concert series. Pls RT	Promotional
12	18	@zappos	RT @ChrisKnight Strong company culture exists when your team responds favorably b/c of personal alignment to organization values	Philosophical
13	16	@WholeFoods	Our flagship store is pedaling the good stuff - just launched bicycle delivery for downtown Austin. More info: http://bit.ly/sBb49	News
14	15	@zappos	In between phone calls, Zappos employees are forced to eat marshmallows to keep speaking skills up - http://bit.ly/chubbybunny	Humorous
15	15	@Starbucks	RT @craftyasparagus: Reading: "12 Clever Ways to Reuse Coffee Grounds- The Green Gathering" http://twitthis.com/ba89tp	Anecdotal
16	13	@WholeFoods	Have you entered for your chance to win 2 tickets to Bonnaroo 2009? We pick a winner tomorrow, so enter today! http://tr.im/bonnaroo	Philanthropic, promotional
17	13	@WholeFoods	Learn about entrepreneurs who lift themselves & their communities out of poverty w/ loans from Whole Planet Foundation. http://tr.im/wpfe	Philanthropic
18	13	@Starbucks	We're having listening parties all over the country for the new DMB album @davejmatthews @larasweetworld more here: http://bit.ly/BNxvw	News
19	13	@kogibbq	Lakernation Kogi Bryant!	Humorous
20	12	@zappos	http://twitpic.com/1rjnv - My cousin's invention: 2 waffles, maple syrup, 2 eggs, 2 slices Taylor Ham, string cheese, 2 sausages, 3 bacon ...	Humorous
21	12	@NAKEDpizza	fyi: @nakedpizza sets record. 68% of sales May 29 from twitter. set store record for all sales. 41% all tickets twtr	News
22	11	@zappos	Obama landed in Las Vegas today. I wanted to board Air Force One, but apparently I didn't have enough frequent flyer miles.	Humorous
23	11	@Starbucks	We taste 250,000 cups of coffee a year to ensure quality: @jphayw some of that is done in the 'cupping room'. I'll grab a photo next time.	Anecdotal
24	10	@WholeFoods	Empower 25,000 people to lift themselves out of poverty. Donate to the Whole Planet Foundation Prosperity Campaign. http://is.gd/l7cJ	Philanthropic

In terms of time, 50% of retweeting happened within 21.26 minutes after the original message was sent, 75% happened within 99.50 minutes, and all took place within 238.77 minutes (Table 3.11 and Figure 3.8). Assuming the users consume the tweets and then immediately retweet those interesting ones, these findings indicate the tweet consumption happens within 1.5 hours or 4 hours at most.

Table 3.11 Five-number summary of time difference (in minutes)

between retweet time and original tweet time, mean, and standard deviation

Minimum	Q1	Median	Q3	Maximum (Adjusted)	Maximum	Mean	Standard Deviation
0.07	5.75	21.26	99.50	238.77	287,031.85	1,416.41	15,858.88

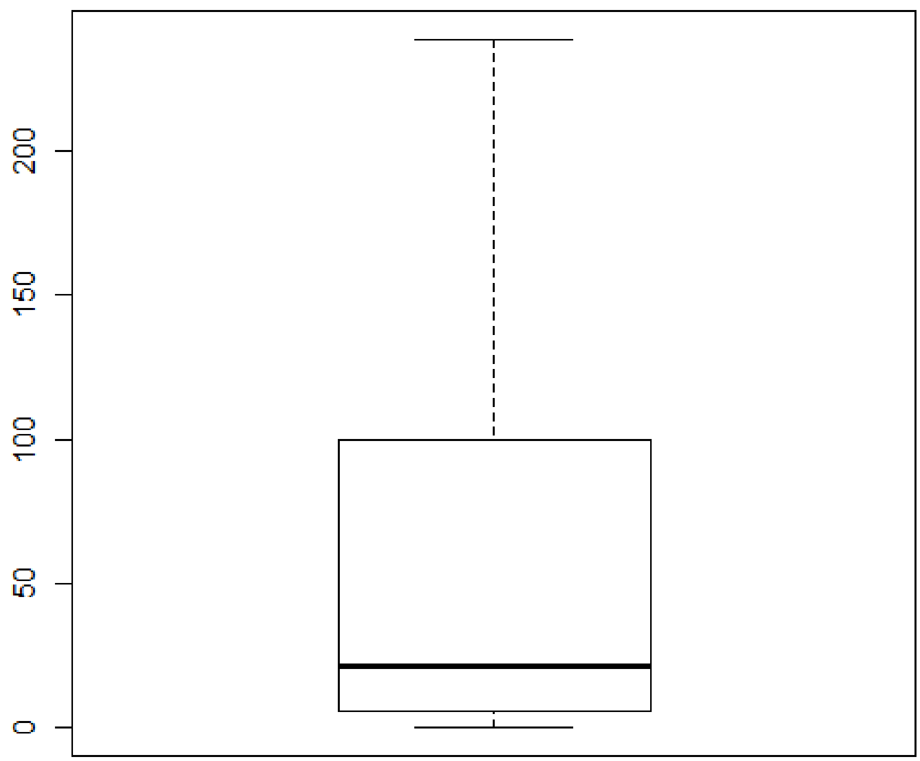


Figure 3.8 Boxplot of time difference (in minutes)

between retweet time and original tweet time (without outliers)

As for the participants of retweeting (Table 3.12), 63.13% were both the brand's followers and followings; 28.81% of the consumers were just the brand's followers. Therefore, 91.94% of consumers retweeting the brand's tweets were the direct receivers of these messages. Only 8.06% consumers were the indirect receivers, among which 6.83% had no connection with the brand at all. Thus, retweeting is mainly performed by consumers having direct connection with the brand.

Table 3.12 Retweeting breakdown by sender's relationship with the brand

Brand's Follower	**Brand's Following**	***n***	**%**
Yes	Yes	721	63.13%
Yes	No	329	28.81%
No	No	78	6.83%
No	Yes	14	1.23%
Total		1,142	100.00%

Discussion

In the project reported in this chapter, I studied the role of brand as an active participant in the WOM communication by correlating the brand engagement in WOM communication with the level of consumers' engagement and by investigating the trajectories of a brand's WOM message diffusion in the Twitter community. I used path analysis to examine 164,478 tweets from 96,725 individual Twitter users with regards to nine brands during a 5-week study period. I operationalized brand engagement as the amount of WOM messages from a brand and the number of consumers the brand follows (*Brand's Following Number*). I operationalized consumers' engagement as the amounts of WOM messages from consumers both connecting to the brand and having no connection with the brand as well as the brand's follower number.

I found major jumps in the WOM messaging volumes after the business launched branded Twitter accounts (Table 3.4), which indicates the dramatic influences of business' engagement in WOM communication on the consumers' engagement in messaging that matters to the brand. I found the statistically linear correlations between all these five variables except two pairs: the brand's following number with the amounts of WOM messages from both the brand and the consumers connecting to the brand (Table 3.6 and Figure 3.5). Path analysis shows that both the amount of WOM messages from the brand and the brand's following number are statistically significant predictors for the amount of WOM messages from the consumers connecting to the brand (Figure 3.6). The brand's following number is a statistically significant predictor for the brand's follower number, but the amount of WOM messages from the brand is not (Figure 3.6). Only a brand's follower number is the statistically significant direct predictor for the amount of WOM messages from consumers without connection to the brand, and the brand's following number is absolutely not a statistically significant direct predictor, but it has huge indirect effect seemingly through the brand's follower number (Table 3.7 and Figure 3.6). The amounts of WOM messages from the brand and the consumers connecting to the brand are not statistically significant direct predictor, but their C.R. values are close to the significance standard (Table 3.7 and Figure 3.6), on the other hand they both have statistically significant linear correlation with the amount of WOM messages from consumers without connection to the brand (Table 3.6 and Figure 3.5). This may be due to the sample size (n=33).

Path analysis is recommended for the sample size to be generally 10 times or ideally 20 times the number of the parameters, and at least 5 times for significance testing of model effects (Kling, 1998). I have 5 parameters here so the sample size is recommended to be 50 to 100 cases ideally but definitely more than 25 cases. I have 33 cases which is more than required but not the ideal situation either. This impacts the significance testing leading to the insignificance of the amounts of WOM messages from the brand and the consumers connecting to the brand, which are obviously close to being statistically significant. In other words, with a larger sample, the relationship could very possibly become significant.

I have several insights from path analysis results and other related analysis. First, business' engagement as an active participant in the WOM process motivates consumers' engagement. I demonstrated the correlations exist between businesses' engagement and consumers' engagement based on Pearson correlation and path analysis. In addition, my data shows the weekly WOM volumes jumped at least 40% with the introduction of branded Twitter accounts. On the theoretical side, marketing and advertising materials from business are major sources for WOM communication (Keller, 2007). Together, these results indicate that the brand's engagement in the WOM communication is the major cause of the consumers' engagement in the WOM communication.

Second, proximity to the communication channel plays a central role in WOM message diffusion. Before businesses launched Twitter accounts, they definitely owned accounts on some other social media platforms. However, with their presences and participations on Twitter, their WOM message volumes got a major boost. This indicates that getting close to the communication channel can have major influence over the communication that exists in the channel. With conversations happening in multiple places, brands, whether they want to do or not, have to engage in these medium to influence the dialogue. Therefore, the business should have the brand presence on many different social media sites to potentially an influence larger audience. Moreover, it should be active on the social media site with high density of its target audience.

Third, the business should be as active as possible on Twitter. One very common question that businesses always ask is how often it should tweet. The answer is at least once every 1.5 to 4 hours. Given that my research shows the causation of business' engagement in WOM communication to the consumers' engagement, the business should actively participate in WOM communication. What "tweets" around comes around. Moreover, most retweeting actions happen within 1.5 hours or 4 hours at most, which indicates the life cycle for the tweet.

On the analysis of retweet, I found about 1% of tweets from the brand are retweeted by the consumers, who are mainly the brand's followers (91.94% of all consumers involved in the retweet). 50% of retweeting actions take place within 20 minutes after the messages are sent out and 75% happen within 1.5 hours. The quickest retweeting happens in 42 seconds, which is close to real time. The majority of retweeted messages are forwarded between 1 to 8 times, which shows that retweeting behavior is a very personal decision. However, the distribution of the retweeting frequency has a long tail. The top retweeted messages are mostly humorous and anecdotal, which is in accordance with previous research from Phelps, Lewis, Mobilio, Perry and Raman (2004). Their research shows that messages sparking strong emotion — humor, fear, sadness, or inspiration — are likely to be forwarded. Therefore, they recommended that businesses craft messages sparking the appropriate emotion for the appropriate causes. The same recommendation also applies for businesses owning Twitter account. Only a very small portion of tweets gets forwarded. Thus, businesses should think of crafting tweets with their unique brand styles and embedding the appropriate emotion.

On the other hand, the retweeted messages do not seem spread out through the whole Twitter community. Almost all of them only reach consumers following the brand and are forwarded by these consumers. There is no doubt that consumers following the brand belong to the core consumers group based on this phenomenon. It means this group is the appropriate audience for the brand. But the more important question the business may want to ask itself is what they can do to break the wall, get the messages forwarded by consumers not having a one-degree separation relationship with it on Twitter since these messages can have strategic or marketing meanings. The brand Twitter account associates may consider reposting the important messages after 1.5 hours in order to increase the WOM message exposure.

Conclusion

I conclude that the brand's engagement in the WOM communication on Twitter enhance the consumers' engagement in the WOM communication. In addition, retweeting as a way to show consumers' response to brand engagement indicates that the influence only reaches consumers with a second-degree relationship to the brand. My research advances the understanding of businesses' roles as an active participant in the WOM communication process. The commercial business Twitter users can leverage the findings in this study to develop effective advertising, marketing and brand management strategies.

Summary

Social media services, such as Twitter, enable commercial businesses to participate actively in online word-of-mouth communication. In this chapter, I examined the potential influences of brand engagement in online word-of-mouth communication on the level of consumers' engagement and investigated the trajectories of a brand's online word-of-mouth message diffusion in the Twitter community. I used path analysis to examine 164,478 tweets from 96,725 individual Twitter users with regards to nine brands during a 5-week study period. I operationalized brand engagement as the amount of online word-of-mouth messages from brand and the number of consumers the brand follows. I operationalized consumers' engagement as the number of online word-of-mouth messages from consumers both connecting to the brand and having no connection with the brand as well as the number of consumers following the brand. I concluded that the brand's engagement on Twitter relates directly to consumers' engagement with online word-of-mouth communication. In addition, retweeting, as an explicit way to show consumers' response to brand engagement, indicates that the influence only reaches consumers with a second-degree relationship to the brand and that the life cycle of a tweet is generally 1.5 hours or 4 hours at most. My research advances understanding of the business' role in the online word-of-mouth communication and brings insight to the analytics of social networks and online word-of-mouth message diffusion patterns.

CHAPTER 4. CONSUMER WHO: PROFILING CONSUMERS ON TWITTER

For commercial business in Web 2.0, developing an online community by maintaining a brand presence on social network sites can enhance awareness and perhaps increase conversions (i.e., completing a desired action), but more importantly, it can create or enlarge the customer base (Schmitt, 2009). Two major Twitter users' characteristics make Twitter a good place for businesses to form a community and present the brand online. First, Twitter has a large number of users actively generating information. Businesses need such community members because they may be more likely to generate word-of-mouth messages, which are critical for awareness and conversions. Twitter users created approximately 100,000 books worth of content from August 2006 to August 2008, although the service allows users to post a maximum of 140 characters at a time (Milstein, Chowdhury, Hochmuth, Lorica, & Magoulas, 2008). Twitter reported their users posted 50 million tweets per day, roughly 600 tweets per second (Weil, February 22, 2010).

Twitter users are also very active in engaging the community in a real-time fashion. For businesses, that means their word-of-mouth messages can potentially travel much faster throughout the network on Twitter than on other platforms, such as online

review sites. Speed is a critical factor for awareness and conversions. Two examples show the power of real-time engagement of Twitter users. Bill Gates signed up on Twitter on January 19, 2010, and he had 100,000 followers in 8 hours, which was 12,500 followers per hour and 208 followers per minute (Parr, January 20, 2010). Conan O'Brien talked about his marketing experience of promoting his tour "Legally Prohibited from Being Funny on Television" on Twitter. He commented, "What was fascinating is, by the time we launched the tour, or announced the tour, I didn't spend one penny on advertising. I sent out one tweet that directed people to a Website where you could buy your ticket. That was it. And the show sold out in a couple of hours across the country. And that got everybody, a lot of people, rethinking how things are marketed. And there is not one billboard. I didn't have to go to one radio station and sit with morning DJs, like, to hawk my show. I didn't have to do any of that. It was one tweet. And I think people are starting to understand that the whole world has completely changed".
2010)

Twitter is widely adopted by organizational users. According to a report from bigmouthmedia & Econsultancy (2009), Twitter was used by 78% of company respondents and 74% of agency respondents among more than 1,100 organizations. Even though a large number of companies start using Twitter, there is still a general lack of systematic Twitter brand and consumer management strategies. The major reason is due to our minimum understanding about Twitter users, particularly those close to businesses (i.e., consumers connecting to the brand Twitter account). Our understanding about this group of people is generally inferred from our understanding about Twitter users in the entire Twitter community. Businesses cannot develop detailed meaningful strategies based on this. Certainly businesses can know consumers connecting with it on the individual level. However, the brand and consumer management strategies cannot be based on individual level due to cost and benefit reasons. A reasonable level of consumers' profiles for Twitter is needed to develop for business users. This need motivates my research on consumer classification.

Jansen and fellow researchers (Jansen, et al., 2009) conducted a case study of Starbucks and explored communication patterns between the brand and consumers. They covered four aspects including range, frequency, time, and content (i.e., how varied were the topics of communication between Starbucks and its customers? How often did Starbucks and its customers twitter each other? When did they twitter? What did they twitter about?). I will follow this direction to advance understanding of consumers' communication behaviors. Who are they? What do they do for living? What do they want to get from Twitter? How active are they on Twitter? How actively are they involved in word-of-mouth communication on Twitter?

In the following sections of this chapter, I first review the relevant research on the consumer classification. I then explain the research question and the statistical approach I took to tackle the question. Next, I present the findings and follow with discussion, and practical implications.

Consumer Classification

Online community is an online space where a group of people interact with each other. A social network is the purest form of an online community, focusing more on the relationship, interaction, and socialization of members compared to the other online communities. For the other online communities, users' interactions are usually centered on a theme such as outdoor activities, photography, etc. But the theme for social network is each individual user. Therefore, on the social network, every user has a prominent profile page to represent one's real identity. The connections between users are presented explicitly. (boyd & Ellison, 2007; Strauss & Buss, 2009) On top of these two basic elements — the profile page and connections — users share information and communicate with each other.

If a brand is on Twitter, it will have its own profile page and a group of friends, who are usually its customers. It interacts with its customers under the theme, which is the brand itself. This is very similar to the online community with the brand as the theme. Given that there is not much research on brand-related social networks, if we want to understand the brand on social network sites, we need to first understand online communities, especially those relevant to brands and marketing. I focus on the literature with regards to community members. These are two streams of research on this topic: theory driven and data driven.

On theory driven research, researchers classify community members based on a set of theoretical standards. The group number and label are known beforehand. Kozinets (1999) studied marketing-related online communities by looking at their members, interaction modes and community types. He used two interrelated factors — relations with the consumption activity and relations with the virtual community — to divide members of virtual communities into four groups, including tourist, mingler, devotee, and insider (Figure 4.1). The relations with the consumption activity refer to perceived centrality or importance of the consumption activity. The relations with the virtual community refer to the intensity of one's social relationship with some others in the community. He believed that the devotees and the insiders are the heavy or loyal consumers and thus are important marketing targets. He also believed the proportion between the devotees and the insiders to the tourists and the minglers follows Pareto's Law (i.e., devotees and insiders are about 20% of the virtual community while tourists and minglers are about 80%). He argued that the community itself may "propagate the development of loyalty and heavy usage by culturally and socially reinforcing consumption" (p.255) so that tourists and minglers can be converted to devotees and insiders.

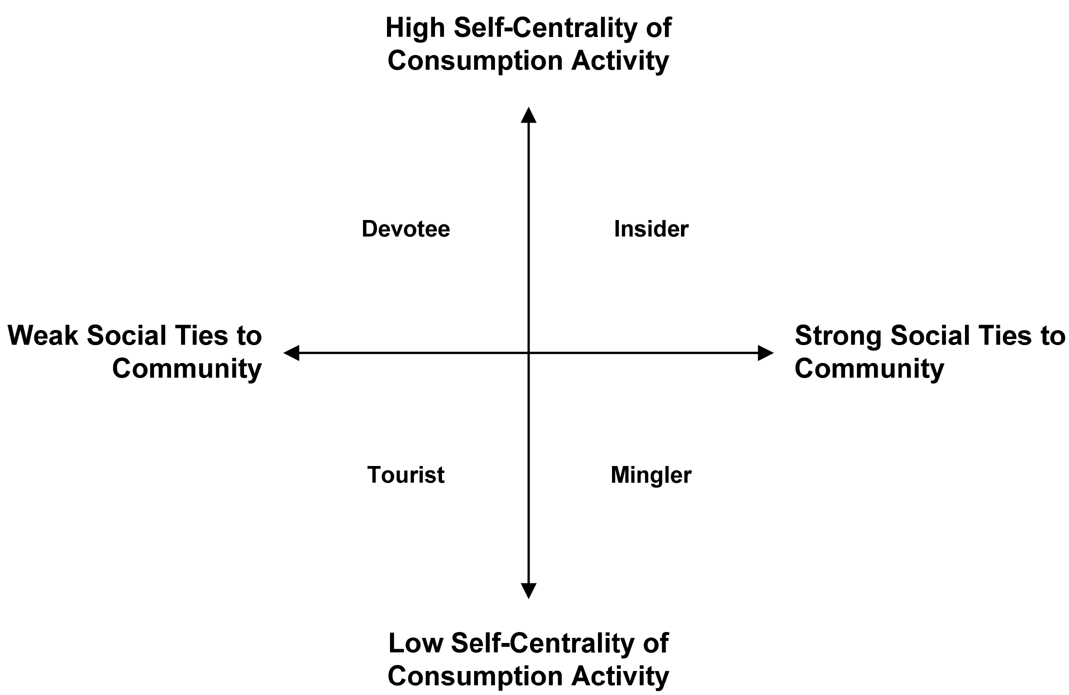


Figure 4.1 Types of virtual consumer community (Kozinets, 1999, p. 255)

Kozinets (1999) differentiated interactions into four modes — informational mode, transformational mode, recreational mode, and relational mode — based on communication orientation and objective (Figure 4.2). Devotees and tourists are mainly in the information mode and are not interested in socializing with others. They tend to collect information and keep it to themselves. On the contrary, minglers and insiders are more interested in communicating with others. They are in the relational mode and interested in spreading word-of-mouth about brands or products. The recreational mode is mainly used by minglers and tourists. They usually enjoy online communication for individual or short-term purposes. They are unlikely to spread the word for the brand or product unless they can gain from this process. Transformational mode is mainly used by devotees and insiders, who value long-term social gain. They own much knowledge about brands or products and also like to share it with others in the community. Kozinets' (1999) framework is very useful for understanding community members and their interactions from a theoretical perspective but very hard to apply in practice due to lack of the definition of some specific measurements.

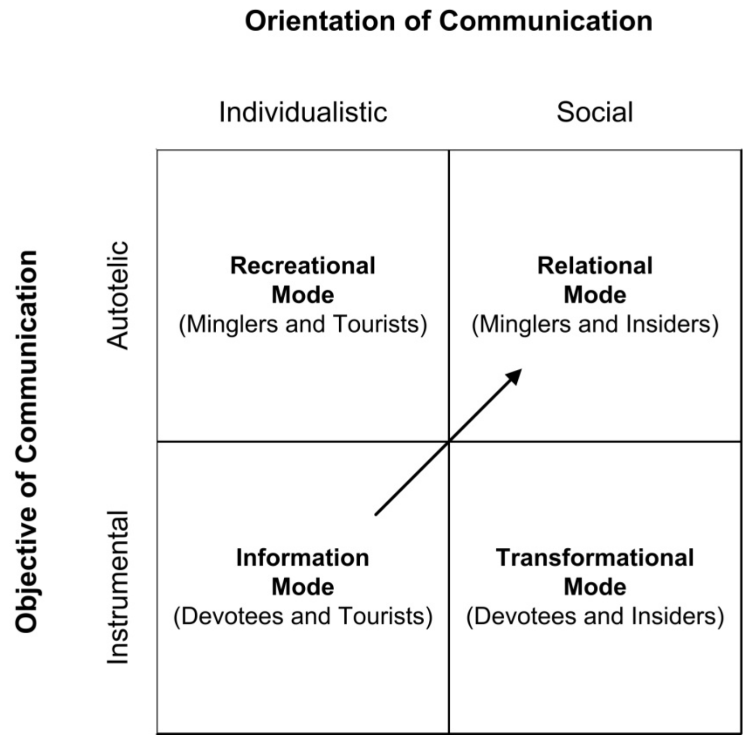


Figure 4.2 Interaction modes for online consumption community (Kozinets, 1999, p. 255)

Shang, Chen, and Liao (2006) divided community members into two groups based on their activities: lurking and posting. They tested the relationship between the participation in online brand community and brand loyalty. They found that posting did not increase brand loyalty but, on the contrary, lurking did contribute to brand loyalty. Their findings indicate that online community extends beyond involvement. They tried to explain the reason why posting did not increase brand loyalty by differentiating loyalty towards brand and community, suggesting that posting might be more relevant to community loyalty rather than brand loyalty. They also highlighted that "online consumer communities should be better understood as a forum for strangers to meet and exchange weak-tie WOM (word-of-mouth) messages, rather than as a form of brand communities" (Shang, et al., 2006, p. 412).

On data driven research, researchers classify community members based on the data. The group number and label are unknown beforehand. De Valck, van Bruggen, and Wierenga (2009) designed a survey to study a virtual community. By performing cluster analysis on five variables, they identified six member types: core members, conversationalists, informationalists, hobbyists, functionalists, and opportunitists. The five variables are frequency of visits, duration of visits, retrieving information, supplying information, and discussing information. Core members are the most active in the community with all of their measurements on all five variables above average. They are also the smallest group of members. Conversationalists participate more in supplying and discussing information during their short visit to the community. Informationalists are active in retrieving and supplying information but not active in discussing information. They also tend to visit the community somewhat less frequently. Hobbyists are the community's most frequent visitors and stay with long duration. They do not focus on retrieving, supplying, or discussing information in the community. Functionalists are the largest group of members. They only actively retrieve information from the community. Opportunitists are the least active members with the measurements on all five variables below average.

On this research question of my dissertation, I also conducted similar cluster analysis to de Valck, van Bruggen, and Wierenga (2009), segmenting consumers connecting to the brand's Twitter account into groups by using real-world tweets. The problem with theory driven consumer classification method is that some group may exist on theoretical level but not in real word. In addition, these classifications in the previous research are for general online community users, not specifically directed at those who engage with brands online. Therefore, there are several unanswered questions. How do potential consumers engage brands in social networking sites like Twitter? How do they exchange information? What is their level of engagement? These are some of the questions that drive my research.

Research Question

Motivated by the previous research on marketing and brand related online communities, my research question is "*what are the characteristics of consumers connecting to brands in the Twitter community?*"

With the help of Twitter, businesses can know their customers at individual level, which makes it possible to thinly slice the market segmentation. Small market segmentation enables highly personalized community, communication, marketing, branding, and product management strategies. The goal of addressing this research question is to

- provide a description of the consumers connecting with the brand on Twitter,
- develop a set of measurements (Table 4.1) to describe these consumers,
- segment the consumers based on these measurements, and
- provide recommendations for the business interested in owning a brand community on Twitter.

Variables for Analysis. I use 49 variables in 5 dimensions to describe consumers connecting with the brand on Twitter (Table 4.1). The 5 dimensions are *profile*, *connection*, *experience*, *activity*, and *impression*, which capture the demographic, relational, and communicative aspects of consumers. These aspects can reflect the levels of consumers' engagement in the community and with the brand.

In the *profile* dimension, I analyze consumers on 5 aspects including *user name*, *picture*, *location*, *Web*, and *biography length*. This gives the demographic information of consumers. It demonstrates how much personal information users disclose and how much effort they spend on creating their Twitter identity. Lampe, Ellison, and Steinfield (2007) drew upon three theories to explain how profile construction can potentially influence participation in the online communities, namely signaling theory, common ground theory, and transaction cost theory. Signaling theory explains the type of information users include in profiles, suggesting that profile elements act as signals which demonstrate the identity of the users. These signals can be manipulated by the senders to convey certain personal qualities and be interpreted by the receivers to judge the characteristics of the profile owners. Common ground theory explains that users fill out profiles and send out signals to facilitate the formation of common ground and enhance mutual understanding. Transaction cost theory bridges signaling theory and common ground theory, and suggests that certain profile elements or certain level of profile disclosure can reduce the cost to build common ground, engage in communication, and form mutual understanding. Therefore, I categorize *user name*, *picture*, *location*, and *Web* with labels based on different levels of personal disclosure. In addition, I measure one's *biography length in word and character* that may also reflect different levels of personal disclosure.

In terms of *connection*, I use 7 variables to demonstrate the social network status of users in the Twitter community, including *follower number*, *following number*, *follower/following ratio*, *follower number speed*, *following number speed*, *number of list including the user*, and *number of list created by the user*. I include *follower number* and *following number* to reflect how social one can potentially be and how large one's network is. I also measure *the ratio of follower number to following number*. A ratio much larger than 1 may indicate the account holder uses Twitter as a broadcasting channel. An extreme example is the Dalai Lama. He has a large number of followers, but he doesn't follow anyone. His purpose on Twitter is to spread information. The ratio around 1 may indicate the account holder is likely to have a potentially balanced interaction with his/her followers and followings. A ratio much less than 1 may indicate that the account holder uses Twitter as an information resource to pull information. I normalize *follower number* and *following number* by week to increase the comparability. I chose to use week as a unit because there is an obvious weekly pattern for people sending tweets (Jansen, et al., 2009). It is reasonable to understand this finding as the way users spend time. Therefore, it is reasonable to make inference that all the activities on Twitter follow this pattern including following people on Twitter. *Number of list including the user* and *number of list created by the user* have similar meanings to follower number and following number. They demonstrate the size of one's network, engagement in the Twitter community, and knowledge about the community.

To explore *experience*, I identify 10 different variables relevant with time measuring how long one uses Twitter and how one connects with the brand to demonstrate the user's experience with Twitter and the brand. They are *Twitter age*, *local age*, *local experience age*, *local experience index*, *follower age*, *follower index*, *local follower age*, *following age*, *following index*, and *local following age*. I use *Twitter age* and *local age* to show one's experience with Twitter. Local here indicates during the study period. I use *local experience age* and *local experience index* to show one's real experience with Twitter on being active and sending out tweets. Index here refers to comparison with overall experience evaluation. For example, *local experience index* is a supplement to local age by taking local age into consideration. It evaluates one's experience in a comparative way. I also have *follower age*, *follower index*, *local follower age*, *following age*, *following index*, and *local following age* to describe the relationship between account holders and brands on the time level. *Follower age*, *follower index*, *following index*, and *following age* describe how motivated one is to connect with the brand. *Local follower age* and *local following age* show how much experience one has with the brand during the study period.

To measure *activity*, I look at one's tweet crafting activity as well as interaction with others on Twitter. I have 7 variables with regards to *the count of tweet*, *tweet without addressing to anyone*, *tweet mentioning someone*, *tweet with hash-tags*, *tweet with hyperlinks*, *retweet*, and *tweet mentioning brand name*, which measure how active one is in sending out different types of tweets.

I have 4 variables including *contact*, *people*, *user interactivity*, and *tweet interactivity* to capture the communication between Twitter users. *Contact* is used to evaluate how many interactions one has with others and *people* is used to evaluate how many unique individuals one interacts with. I employ *user interactivity* to assess how interactive one can be with other Twitter users by measuring how many contacts one has with one individual on average. I evaluate *tweet interactivity* by measuring how many individual Twitter users on average are mentioned in one tweet. For example, if user A just mentions user B in one tweet, then the *tweet interactivity* is 1. If user A mentions user B and C in one tweet, then the *tweet interactivity* is 2.

I have 9 variables normalized by week. Seven variables are *the frequency of tweet*, *tweet without addressing to anyone*, *tweet mentioning someone*, *tweet with hash-tags*, *tweet with hyperlinks*, *retweet*, and *tweet mentioning brand name*. These variables are employed to measure how comparatively active the users are on tweet production. The frequency is evaluated at the week level since previous research found a strong weekly pattern of Twitter usage. The users tweeted more in the middle of the week and less during weekend and the beginning of the week (Jansen, et al., 2009). Therefore, it makes more practical sense to measure tweet frequency by week, which is comparatively stable for individual users. I also normalize *contact* and *people* by week to increase the comparability.

Impression is an online marketing term that refers to the potential process of one seeing an ad (Boone & Kurtz, 2008, p. 516). I borrow this term to define *tweet impression* as the potential process of one reading a particular tweet. There are two ways to increase *tweet impression*, namely increasing the number of followers reading the tweet or increasing the number of tweets sent. To evaluate impression, I develop 6 variables: *the impression of tweet*, *tweet without addressing to anyone*, *tweet with hash-*

tags, *tweet with hyperlinks*, *retweet*, and *tweet mentioning brand name*. I do not measure *the impression of tweet mentioning someone* because his type of tweet is for interpersonal communication purposes rather than making impression on the followers of the tweet sender. Table 4.1 is an overview of all these 49 variables. Those 15 variables in italics are used to calculate other variables since they are less comparable between individuals. I have 34 variables for cluster analysis.

Table 4.1 Variable descriptions

Dimension	Variable[a]	Unit	Definition
Profile	User name	N/A	What kind of user name one has on one's Twitter account? 0: No real name; 1: Nick name; 2: Proper name
	Picture	N/A	Does one include the picture on one's Twitter profile? If so, what kind of picture? 0: No picture; 1: No real person picture; 2: Partial real person picture; 3: Real person picture
	Location	N/A	Does one display the location on one's Twitter profile? If so, what kind of location? 0: No location; 1: Abstract description; 2: Country; 3: State/Province; 4: City
	Web	N/A	Does one display the URL address on one's Twitter profile? If so, what kind of URL address? 0: No link; 1: Company/Business Website; 2: Multimedia or social sharing (e.g. Youtube, Flickr, Digg) 3: Social network (e.g. Facebook, Linkedin, MySpace); 4: Blog; 5: Personal Website
	Biography length by word	Word	How many words are used in the biography on one's Twitter profile?
	Biography length by character	Character	How many characters are used in the biography on one's Twitter profile?
Connection	*Follower number*	Person	How many users one follows on Twitter?
	Following number	Person	How many users follow one on Twitter?
	Follower/following ratio	N/A	How balanced is one's talking to audience and listening to sources on Twitter? $Follower\ /\ FollowingRatio = \frac{FollowerNumber}{FollowingNumber}$
	Follower number speed	Person/week	$FollowerNumberSpeed = \frac{FollowerNumber}{TwitterAge} \times 7$
	Following number speed	Person/week	$FollowingNumberSpeed = \frac{FollowingNumber}{TwitterAge} \times 7$
	Number of lists including the user	List	How many lists one are included in on Twitter?
	Number of lists created by the user	List	How many lists one creates on Twitter?
Experience	Twitter age	Day	How many days has one used Twitter from the day one signs up to May 31, 2009?
	Local age	Day	How many days has one used Twitter from May 1, 2008, to May 31, 2009?
	Local experience age	Day	How many days does one send out tweets from May 1, 2008, to May 31, 2009?

	Local experience index	N/A	$LocalExperienceIndex = \frac{LocalExperienceAge}{LocalAge}$
	Follower age	Day	After how many days does one start connecting with the brand as afollower on Twitter from the day one signs up?
	Follower index	N/A	$FollowerIndex = 1 - \frac{FollowerAge}{TwitterAge}$
	Local follower age	Day	How many days has one been the brand's follower on Twitter from May 1, 2008, to May 31, 2009?
	Following age	Day	After how many days does one start connecting with the brand as following on Twitter from the day one signs up?
	Following index	N/A	$FollowingIndex = 1 - \frac{FollowingAge}{TwitterAge}$
	Local following age	Day	How many days has one been the brand's following on Twitter from May 1, 2008, to May 31, 2009?
Activity	*Tweet*	Tweet	How many tweets did one send out from May 1, 2008, to May 31, 2009?
	Announcement	Tweet	How many tweets did one send out without mentioning any Twitter user in them from May 1, 2008, to May 31, 2009?
	@mention	Tweet	How many tweets did one send out mentioning some Twitter user in them from May 1, 2008, to May 31, 2009?
	#hash-tag	Tweet	How many tweets did one send out with hash-tagged term(s) from May 1, 2008, to May 31, 2009?
	HTTP	Tweet	How many tweets with hyperlink(s) did one send out from May 1, 2008, to May 31, 2009?
	Retweet	Tweet	How many retweets did one send out from May 1, 2008, to May 31, 2009?
	Branded tweet	Tweet	How many tweets did one send out mentioning brand, product, or service from May 1, 2008, to May 31, 2009?
	Tweet frequency	Tweet/week	How often does one send out tweets? $TweetFrequency = \frac{Tweet}{LocalAge} \times 7$
	Announcement frequency	Tweet/week	How often does one send out tweets without mentioning any Twitter user? $AnnouncementFrequency = \frac{Announcement}{LocalAge} \times 7$
	@mention frequency	Tweet/week	How often does one send out tweets mentioning some Twitter user? $@mentionFrequency = \frac{@mention}{LocalAge} \times 7$
	#hash-tag frequency	Tweet/week	How often does one send out tweets with hash-tagged term(s)? $\#hash-tagFrequency = \frac{\#hash-tag}{LocalAge} \times 7$
	HTTP frequency	Tweet/week	How often does one send out tweets with hyperlink(s)? $HTTPFrequency = \frac{HTTP}{LocalAge} \times 7$
	Retweet frequency	Tweet/week	How often does one retweet? $RetweetFrequency = \frac{Retweet}{LocalAge} \times 7$
	Branded tweet frequency	Tweet/week	How often does one send out tweets mentioning the name of brand, product ,or service? $BrandedTweetFrequency = \frac{BrandedTweet}{(LocalFollowerAge + LocalFollowingAge)/2} \times 7$
	Contact	Contact	How many contacts to people one has on Twitter from May 1, 2008, to May 31, 2009?
	People	Person	How many unique people one mentions on Twitter from May 1, 2008, to May 31, 2009?
	User interactivity	Person/Contact	How many contacts one makes per person on Twitter? $UserInteractivity = \frac{People}{Contact}$
	Tweet interactivity	Person/tweet	How many people on average does one mention in a tweet? $TweetInteractivity = \frac{Contact}{@mention}$
	Contact frequency	Contact/week	$ContactFrequency = \frac{Contact}{LocalAge} \times 7$
	Interaction frequency	Person/week	$InteractionFrequency = \frac{People}{LocalAge} \times 7$

Impression	Tweet impression	Tweet×person/week	$TweetImpression = TweetFrequency \times FollowerNumber$
	Announcement impression	Tweet×person/week	$AnnouncementImpression = AnnouncementFrequency \times FollowerNumber$
	#hash-tag impression	Tweet×person/week	$\#hash-tagImpression = \#hash-tagFrequency \times FollowerNumber$
	HTTP impression	Tweet×person/week	$HTTPImpression = HTTPFrequency \times FollowerNumber$
	Retweet impression	Tweet×person/week	$RetweetImpression = RetweetFrequency \times FollowerNumber$
	Branded tweet impression	Tweet×person/week	$BrandedTweetImpression = BrandedTweetFrequency \times FollowerNumber$

a. Variables in italics are only used for calculation and are not used for clustering.

Methodology

To address this research question, I used Data Set 2 on information of consumers connecting to the brand Twitter account and Data Set 3 on profiles and tweets of sampled consumers in the brand's immediate social network on Twitter (Table 1.3). I mainly analyzed the content of tweets by using descriptive statistics and cluster analysis.

Data processing

I had 2,700 consumers' Twitter accounts for this study. First, I profiled the purposes they used Twitter, based on the tweet content and what they did for living by checking on their Websites on profile or LinkedIn accounts (Table 4.2). I manually checked every account and performed analysis. I did this about one year after I selected the accounts and collected their tweet data. I found 300 accounts (11.11%) were not available on Twitter (Table 4.2). The reasons for unavailability may be that either they closed their accounts, or their accounts were shut down by Twitter. I dropped these accounts from further analysis since they disconnected from Twitter and the brand, and they are most likely outlier Twitter accounts. I identified 448 account holders (16.59%) as organizations or companies (Table 4.2). Among them, 31.47% accounts (5.22% of all accounts) are news media, including TV channel, radio channel, Website, blog, and such. I identified 1,952 accounts (72.30%) used primarily for personal or both personal and business purposes (Table 4.2). Among them, the majority of identifiable jobs fell in marketing, public relation, and advertising (14.15% of all accounts). Interestingly, none of these users whose jobs were identifiable are the employees of the businesses I investigated. However, I was unable to identify what the account holders' are doing for a living for 49.49% personal accounts (35.78% of all accounts). I dropped the organization accounts from further analysis since they cannot represent the individual consumers and reflect individual consumers' behavior. Thus, I have 1,952 individual consumers' Twitter accounts and 1,387,058 tweets for further exploration.

Table 4.2 Account type and job/industry

Account type	**Job/Industry**	***n***	**%**	**% within group**
Personal and mixed	Unidentifiable	966	35.78%	49.49%
	Marketing, public relation, and advertising	382	14.15%	19.57%
	Student	96	3.56%	4.92%
	Information technology and service	85	3.15%	4.35%
	Other	423	15.67%	21.67%
	Subtotal	1,952	72.30%	100.00%
Organization	News media	141	5.22%	31.47%
	Marketing	42	1.56%	9.38%
	Ecommerce	33	1.22%	7.37%
	Other	232	8.59%	51.79%
	Subtotal	448	16.59%	100.00%
Unavailable		300	11.11%	
Total		2,700	100.00%	

Second, I analyzed users on those 6 *profile* variables listed in Table 4.1. *User name*, *picture*, *location*, and *Web* are categorical variables. I manually coded them (Table 4.3). I found 44.47% of accounts did not use real names while 24.08% accounts did use the account holders' proper names. On the *profile picture*, 58.91% accounts used real person pictures, but 20.90% accounts had no picture. 50.92% accounts put city level address on their profiles, and 31.56% did not have any address. 56.40% accounts did not show *Web address* on profile. 16.34% used the organization or company's Website, and 13.22% used their own personal Website address. I also calculated the biography length of their profile.

Table 4.3 Summary statistics for categorical variables

Variable	**Value**	***n***	**%**
User name	0: No real name	**868**	**44.47%**
	1: Nick name	614	31.45%
	2: Proper name	470	24.08%
	Total	1,952	100.00%
Profile picture	0: No picture	408	20.90%
	1: No real person picture	344	17.62%
	2: Partial real person picture	50	2.56%
	3: Real person picture	**1,150**	**58.91%**
	Total	1,952	100.00%
Location on profile	0: No location	616	31.56%
	1: Abstract description	67	3.43%
	2: Country	104	5.33%
	3: State/Province	171	8.76%
	4: City	**994**	**50.92%**
	Total	1,952	100.00%
Web address on profile	0: No link	**1,101**	**56.40%**
	1: Company/Business Website	319	16.34%
	2: Multimedia or social sharing	16	0.82%

	3: Social network	107	5.48%
	4: Blog	151	7.74%
	5: Personal Website	258	13.22%
	Total	1,952	100.00%

Third, I collected the data on *connection* dimension in Table 4.1. I did this in May, 2010, almost a year after the end date of the study period. However, these data still capture the social nature of the accounts given that social people remain the same.

Fourth, I calculated the experience related variables (Table 4.1) based on the data I downloaded. Fifth, I performed analysis of tweet content and labeled tweets if they mentioned someone, if they had hash-tagged terms, if they contained hyperlinks, or if they mentioned the name of the brand they connect to. Based on these labels, I generated the data for the variables in *activity* and *impression* dimensions (Table 4.1). Table 4.4 presents the summary statistics of all these continuous variables.

Table 4.4 Summary statistics of continuous variables

Dimension	**Variable**	**Mean**	**SD[a]**	**Min**	**Q1[b]**	**Q3[c]**	**Max**
Profile	Biography length by word	7.71	8.45	0	0	14.00	34.00
	Biography length by character	49.62	53.22	0	0	90.00	165.00
Connection	*Follower number*	3,589.00	41,489.00	0	15.00	964.00	1,753,688.00
	Following number	2,384.00	11,524.00	0	27.00	1,095.00	254,027.00
	Follower/following ratio	41.30	1,725.90	0	0.40	1.10	76,247.30
	Follower number speed	39.70	557.00	0	0.20	10.10	24,117.50
	Following number speed	25.09	108.93	0	0.42	11.73	1,925.90
	Number of lists including the user	68.20	515.10	0	0	15.00	17,067.00
	Number of lists created by the user	4.08	31.39	0	0	1.00	1,067.00
Experience	Twitter age	225.46	218.70	1.00	68.00	343.75	1,053.00
	Local age	185.75	139.45	1.00	68.00	343.75	396.00
	Local experience age	75.40	103.88	0	2.00	105.00	396.00
	Local experience index	0.32	0.32	0	0.03	0.56	1.00
	Follower age	120.80	180.31	0	0	188.75	992.00

	Follower index	0.66	0.33	0.00	0.37	1.00	1.00
	Local follower age	95.82	81.67	1.00	41.25	123.00	396.00
	Following age	117.57	182.32	0	0	184.75	995.00
	Following index	0.68	0.33	0.01	0.41	1.00	1.00
	Local following age	88.45	82.26	1.00	27.00	122.00	396.00
Activity	*Tweet*	710.60	1,815.50	0	3.00	501.30	21,887.00
	Announcement	316.60	808.70	0	2.30	267.80	18,926.00
	@mention	394.00	1,195.80	0	0	192.80	19,013.00
	#hash-tag	49.23	213.52	0	0	14.00	5,192.00
	HTTP	167.00	598.10	0	0	96.80	12,475.00
	Retweet	40.66	153.58	0	0	14.00	3,137.00
	Branded tweet	0.85	4.88	0	0	0	102.00
	Tweet frequency	19.54	50.52	0	0.28	17.83	1,049.38
	Announcement frequency	9.46	32.65	0	0.22	10.06	907.41
	@mention frequency	10.08	27.48	0	0	6.41	336.09
	#hash-tag frequency	1.39	6.73	0	0	0.44	160.60
	HTTP frequency	5.02	26.99	0	0	3.14	697.39
	Retweet frequency	1.39	8.57	0	0	0.42	274.49
	Branded tweet frequency	0.08	.38	0	0	0	4.79
	Follower number speed	39.70	557.00	0	0.20	10.10	24,117.50
	Following number speed	25.09	108.93	0	0.42	11.73	1,925.90
	Contact	468.80	1,425.80	0	0	218.00	21,292.00
	People	131.58	368.79	0	0	86.00	5,039.00
	User interactivity	0.36	0.33	0	0	0.59	1.00
	Tweet interactivity	0.75	0.56	0	0	1.09	4.81
	Contact frequency	12.61	45.32	0	0	7.13	1,330.00
	Interaction frequency	3.86	12.64	0	0	2.70	257.15
Impression	Tweet impression	310,130.00	3,186,059.00	0	4.00	12,574.00	77,600,516.00
	Announcement impression	157,902.00	2,237,084.00	0	3.00	6,760.00	66,976,003.00
	#hash-tag impression	23,604.00	302,972.00	0	0	306.00	10,268,740.00
	HTTP impression	102,663.00	1,479,669.00	0	0	2,445.00	44,146,974.00
	Retweet impression	20,302.00	215,086.00	0	0	269.00	6,305,801.00
	Branded tweet impression	99.60	813.40	0	0	0	19,481.50

a. SD: Standard deviation
b. Q1: First quartile
c. Q3: Third quartile

TwoStep cluster approach

"Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters)" (Jain, Murty, & Flynn, 1999). The group labels are unknown beforehand but will be derived from the data. Thus, this is a purely data-driven grouping technique. It is useful for exploratory data analysis. I chose cluster analysis as the consumer profiling technique in this project to classify consumers connecting with the brand on Twitter into groups.

I selected the TwoStep cluster analysis procedure for the clustering analysis using SPSS. I opted for this method mainly because it can process both categorical and continuous/interval data. This method has two steps: "(1) pre-cluster the cases (or records) into many small sub-clusters; (2) cluster the sub-clusters resulting from pre-cluster step into the desired number of clusters" (SPSS Inc, 2004, p. 1).

According to SPSS documentation (2004, p. 2), the pre-cluster step is implemented by constructing a modified cluster feature (CF) tree (Zhang, Ramakrishnon, & Livny, 1996). The CF tree includes two basic components: nodes at different levels and a number of entries contained in each node. The entry in the leaf node represents the final sub-cluster. The non-leaf nodes and their entries are used to guide a new record to the leaf node. To insert a new record, it starts from the root node, is recursively guided by the closest entry in the node to find the closest child node, and descends along the CF tree until reaching a leaf node. Then it finds the closest leaf entry in the leaf node. If the record is within the distance threshold of the closest leaf entry, it is included in the leaf entry. Otherwise, it starts a new leaf entry in the leaf node. If there is no space for a new leaf entry in the leaf node, the leaf node is split into two. This process is repeated until all the records are included in the CF tree. For more details of CF tree construction, see the BIRCH algorithm (Zhang, et al., 1996). For distance calculation, SPSS (2004, pp. 3-4) adopts a probability based distance, the log-likelihood distance, to evaluate the similarities and dissimilarities between two clusters when handling both continuous and categorical variables.

In the cluster step, the system employs an agglomerative hierarchical clustering method to group the sub-clusters resulting from the pre-cluster step (SPSS Inc, 2004, p. 3). To determine the optimal number of clusters automatically, SPSS (2004, pp. 4-5) adopts a two-stage evaluation. In the first stage, Bayesian Information Criterion (BIC) is calculated for each number of clusters. BIC is a likelihood criterion penalized by the complexity of the model, which is measured by the number of parameters in the model (Chiu, Fang, Chen, Wang, & Jeris, 2001). The smaller the BIC is, the better the model is. If BIC is the smallest when the number of clusters is 1, the second stage is skipped. Then the cluster number of 1 is the optimal number for this data set. Otherwise, in the second stage, the system is trying to find the largest relative increase in BIC between the two closest clusters to determine the optimal cluster number.

The TwoStep cluster approach assumes that the continuous variables are independent and normally distributed. It also assumes that categorical variables are independent and follow a multinomial distribution. However, empirical testing indicates that the violations of the independence and distributional assumptions are not critical. (Chiu, et al., 2001) Therefore, I skipped testing the independence and distribution on my data considering the TwoStep cluster approach is robust to the violation of these assumptions.

Variables selection in analysis

Variables selections are critical for finding structure in a high dimensional variable space during cluster analysis since the non-informative variables can undermine the group structure. To address the research question, I employed a forward-selection process by using silhouette coefficient to evaluate the variables.

I am motivated by the variable selection procedure from Fowlkes, Gnanadesikan, and Ketternring (1988). They proposed a variable selection procedure in conjunction with the clustering process. It is a forward-selection process. A backward elimination approach is found to be ineffective since the noise variables distort the analysis based on all variables, and it is impossible to differentiate the non-informative variables from the informative ones. The major idea behind their method is to select a variable subset that produces the most "significant" partition.

Fowlkes, Gnanadesikan, and Ketternring (1988) evaluated the significance of the variable selection in the complete linkage hierarchical cluster analysis procedure by using a scaled version of Pillai's trace statistic. Since I use the TwoStep cluster analysis procedure in SPSS, there is no such statistic available. Instead, SPSS provides the silhouette coefficient to measure the performance of cluster analysis. Therefore, I will use the silhouette coefficient (Kaufman & Rousseeuw, 1990) to evaluate the variable performance.

According to Kaufman and Rousseeuw (1990), silhouette coefficient is a statistic used to measure the amount of clustering structure that has been discovered by applying a certain classification algorithm on a group of variables. Silhouette coefficient for each case in the data set is defined as

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

a(i) is the average distance of i to all other objects of A, the cluster to which i has been assigned to; b(i) is the average distance between i and the nearest neighbor cluster. The larger s(i) is, the better i is classified. Average silhouette coefficient for all the cases is an overall quality evaluation for clustering analysis. Based on their experience, Kaufman and Rousseeuw (1990) provided some rules of thumb to understand the silhouette coefficient value in practice, presented in Table 4.5.

Table 4.5 Subjective interpretation of the silhouette coefficient

(Kaufman & Rousseeuw, 1990, p. 88)

Silhouette Coefficient Value (SC)	**Proposed Interpretation**
(0.70, 1.00]	A strong structure has been found
(0.50, 0.70]	A reasonable structure has been found
(0.25, 0.50]	The structure is weak and could be artificial; please try additional methods on this data set
[-1, 0.25]	No substantial structure has been found

I used silhouette coefficient to measure how much noise the variable contains. Assuming one variable is very informative and does not contain much noise, the silhouette coefficient value should be very high for the cluster analysis based on this variable. It indicates that this variable is significant to the cluster structure. Therefore, an iterative forward variable selection process is proposed as follows:

Step 1. Calculate average silhouette coefficient for cluster sizes of 3 to 9 for every variable. Choose the variable with the maximum average silhouette coefficient value and comparatively balanced cluster structure as the initial clustering variable. I evaluate variable performance on cluster sizes of 3 to 9 because it is reasonable to get a cluster size in this range. In addition, the minimum BIC for all variables in the data set is around cluster size of 9, which indicates that I can have the most parsimonious model by clustering the consumers into 9 groups.

Step 2. Calculate average silhouette coefficient for cluster sizes of 3 to 9 by adding one variable at a time. Do this for every remaining variable. Choose the variable yielding the maximum average silhouette coefficient value and comparatively balanced cluster structure as the new clustering variable.

Step 3. Repeat Step 2 until there is no variable whose silhouette coefficients values for cluster sizes of 3 to 9 are mostly above 0.50.

Results

Variable selection

I used the forward-selection method in 24 steps to select the 23 variables for clustering analysis. The variable selected in each step and the corresponding silhouette coefficients are presented in Table 4.6. In general, the average silhouette coefficient decreases from step to step (Figure 4.3). In Step 23, all silhouette coefficients for *local experience index* are slightly above 0.50. However, SPSS presents the silhouette coefficient with one decimal point, but the cluster quality chart presenting silhouette coefficient shows that 0.5 is achieved by rounding a value above 0.50. It shows these 23 variables together can present a reasonable cluster structure for cluster sizes of 3 to 9. These 23 variables include all of the variables in *connection*, *activity*, and *impression* dimensions, and one *experience* variable called *local experience index*.

Table 4.6 Variable selected in each step

Step	**Variable**	**Silhouette Coefficient For Different Cluster Sizes**							
		3	**4**	**5**	**6**	**7**	**8**	**9**	**Mean**
1	User interactivity	0.7	0.8	0.7	0.8	0.8	0.8	0.8	0.77
2	Follower/following ratio	0.7	0.7	0.8	0.7	0.8	0.8	0.8	0.76
3	Follower number speed	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.71
4	Tweet interactivity	0.7	0.7	0.7	0.7	0.7	0.6	0.7	0.69
5	Branded tweet frequency	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.70
6	Branded tweet impression	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.69
7	Announcement frequency	0.6	0.7	0.6	0.6	0.7	0.7	0.7	0.66
8	Number of lists created by the user	0.6	0.7	0.7	0.6	0.7	0.7	0.6	0.66
9	Tweet impression	0.6	0.6	0.7	0.6	0.6	0.7	0.7	0.64
10	#hash-tag impression	0.6	0.6	0.7	0.6	0.7	0.7	0.6	0.64
11	Announcement impression	0.6	0.6	0.7	0.6	0.6	0.7	0.7	0.64
12	#hash-tag frequency	0.6	0.6	0.7	0.6	0.6	0.7	0.7	0.64
13	HTTP frequency	0.6	0.6	0.7	0.7	0.6	0.6	0.6	0.63
14	HTTP impression	0.6	0.6	0.7	0.7	0.6	0.5	0.6	0.61
15	Number of lists including the user	0.6	0.6	0.5	0.7	0.6	0.6	0.6	0.60
16	@mention frequency	0.3	0.6	0.6	0.6	0.6	0.6	0.6	0.56
17	Contact frequency	0.6	0.6	0.5	0.5	0.5	0.5	0.6	0.54
18	Interaction frequency	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.53
19	Following number speed	0.3	0.5	0.5	0.6	0.6	0.6	0.6	0.53
20	Tweet frequency	0.3	0.5	0.5	0.5	0.6	0.6	0.6	0.51

Step	Variable	Silhouette Coefficient For Different Cluster Sizes							
		3	4	5	6	7	8	9	Mean
21	Retweet frequency	0.3	0.5	0.5	0.5	0.5	0.5	0.6	0.49
22	Retweet impression	0.3	0.6	0.5	0.6	0.6	0.6	0.5	0.53
23	Local experience index	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.50

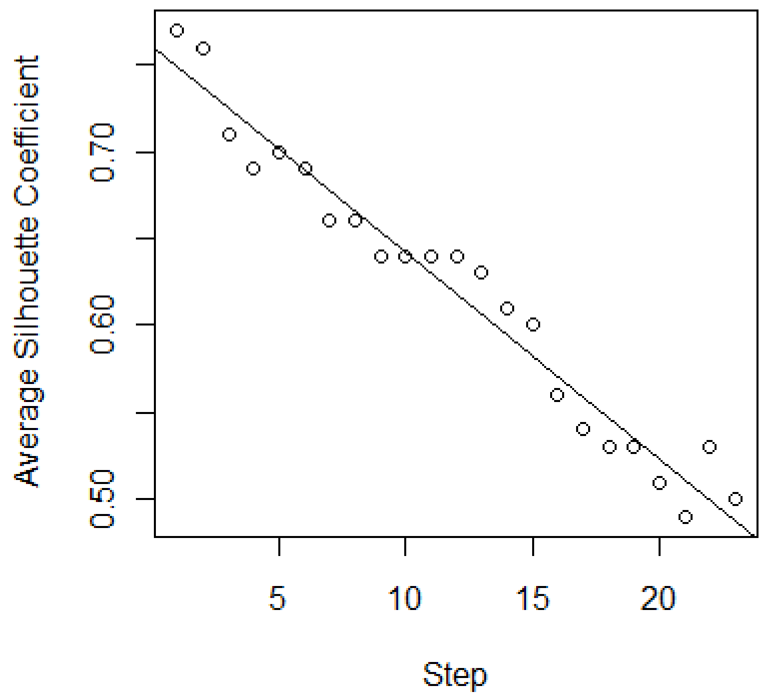


Figure 4.3 Plot of average silhouette coefficient on each step with regression line

Silhouette coefficients for variables in Step 24 are presented in Table 4.7. None of them can be qualified to combine with the selected variables to present a reasonable cluster structure. Interestingly, it includes all of the *profile* variables as well as all of the *experience* variables except *local experience index*.

Table 4.7 Variable performance in Step 24

Variable	Silhouette Coefficient For Different Cluster Sizes							
	3	4	5	6	7	8	9	Mean
User name	0.3	0.4	0.3	0.4	0.4	0.4	0.4	0.37
Picture	0.3	0.4	0.3	0.3	0.4	0.4	0.4	0.36
Location	0.5	0.4	0.3	0.3	0.3	0.4	0.4	0.37
Web	0.3	0.2	0.3	0.3	0.3	0.3	0.3	0.29
Biography length by word	0.5	0.4	0.5	0.5	0.4	0.5	0.4	0.46
Biography length by character	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.39
Twitter age	0.5	0.5	0.5	0.5	0.4	0.4	0.5	0.47
Follower age	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.49
Follower index	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.39
Following age	0.3	0.5	0.4	0.5	0.5	0.4	0.5	0.44
Following index	0.3	0.4	0.4	0.3	0.4	0.4	0.4	0.37

Consumer segmentation

Since the silhouette coefficients for 23 variables with cluster sizes of 3 to 9 are all above 0.50 (Table 4.6), a reasonable cluster structure can be achieved based on these 23 variables. Table 4.8 presents the BIC information for cluster sizes of 1 to 15 to assist in deciding the best cluster size. Since cluster size with the lowest BIC value is the best, cluster size of 7 is the best option. However, BIC for cluster size of 7 is not dramatically different from cluster size of 6. Therefore, I chose cluster size of 6 given that it leads to a smaller cluster and avoids getting fractional groups. I opted to skip the second step, the system used to evaluate by using ratio of BIC changes and ratio of distance measures (SPSS Inc, 2004, pp. 4-5), since these two statistics only evaluate the current number of clusters against the previous number of clusters. It may be true that one cluster size increase does not gain much information, but it is not necessarily true that two or more cluster size increases do not gain much information either. Therefore, I chose cluster size of 6, whose BIC value is 40.84% smaller than that of cluster size of 2, the system's choice of cluster size. Table 4.9 and Figure 4.4 present the distribution information of clusters. I have 3 big fractions and 3 small fractions.

Table 4.8 Cluster analysis

Number of clusters	Bayesian Information Criterion (BIC)	BIC Change[a]	Ratio of BIC Changes[b]	Ratio of Distance Measures[c]
1	31,456.56			
2	13,125.01	-18,331.55	1.00	7.12
3	10,848.93	-2,276.08	0.12	1.45
4	9,388.30	-1,460.64	0.08	1.20
5	8,228.68	-1,159.62	0.06	1.86
6	7,765.04	-463.64	0.03	2.12
7	7,730.12	-34.92	0.00	1.30
8	7,783.98	53.87	-0.00	1.08
9	7,859.38	75.40	-0.00	1.09
10	7,957.34	97.96	-0.01	1.19
11	8,095.71	138.37	-0.01	1.03
12	8,240.86	145.15	-0.01	1.29
13	8,431.31	190.45	-0.01	1.19
14	8,646.67	215.36	-0.01	1.13
15	8,876.90	230.22	-0.01	1.05

a. The changes are from the previous number of clusters in the table.
b. The ratios of changes are relative to the change for the two cluster solution.
c. The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

Table 4.9 Cluster distribution

Cluster	***n***	**%**
1	805	41.24%
2	627	32.12%
3	368	18.85%
4	82	4.20%
5	59	3.02%
6	11	0.56%
Total	1,952	100.00%

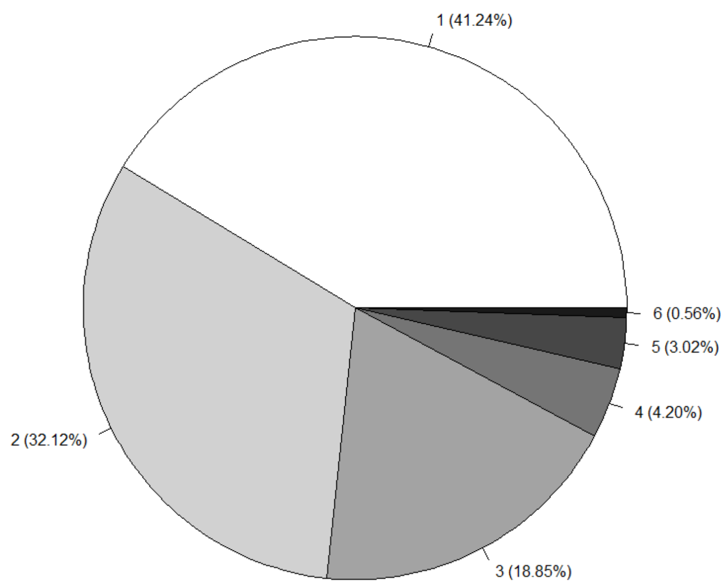


Figure 4.4 Cluster pie chart

Table 4.10 presents the profile of 6 clusters achieved by running cluster analysis on 23 selected variables. Cluster 2 has the smallest group mean value on 22 variables and is relatively low on *follower/following ratio*. This group of consumers was in a small social network. They knew a few people, and a few people knew them. They were not active in engaging in tweeting. They did not make big impressions via tweeting.

Cluster 6 has the maximum group mean on 21 out of 23 variables. This group of consumers was in a big social network. They knew a lot of people, and a lot people knew them. They actively tweeted about 90% of the time. Moreover, they were productive in sending tweets and interacting with other Twitter users. They also made huge tweet impressions. They had low *branded tweet frequency*. However, this group of consumers was still able to make great branded tweet impressions since they had a very large audience. Most of the members in this consumer sector were active celebrity Twitter users.

Cluster 1 has relative small group means on the majority of variables. The consumers in this group had a relatively small social network but not as small as those in Cluster 2. They were not active in tweeting, but they were more active compared with those consumers in Cluster 2. They did not make big tweet impressions, but their impressions were bigger than those made by the consumers in Cluster 2. One thing very interesting about this group is that they had the maximum group mean on *user interactivity*, meaning that they contacted many different individuals.

Cluster 3 has relatively big group means on most of the variables. The consumers in this group had a relatively small social network with strong ties since they had a relatively high group mean on *number of lists including the user* and *number of lists created by the user*, meaning they knew many people, and many people knew them. They were active 80% of the time. They were active in crafting different types of tweets except branded tweets. They made frequent interaction with others. However, they did not make big impressions due to the small social network they owned. It is very likely that they used Twitter more as a personal communication tool.

Cluster 4 and Cluster 5 are two groups very similar to each other. These two groups tweeted a lot about the brand and made relative big branded impressions. They both owned relatively big networks. While both groups were actively engaged in tweet communication, Cluster 5 was more active than Cluster 4 was. They both made relatively big tweet impressions, but Cluster 5 made bigger impressions due to the higher level of tweet communication engagement.

Table 4.10 Cluster profile on selected variables

Dimension	**Variable**		**Cluster[a]**						
			1	**2**	**3**	**4**	**5**	**6**	**Total**
Connection	Follower/following ratio	Mean	3.23	1.36	1.87	1.25	2.44	**6932.98**	41.32
		SD	37.44	8.48	9.41	1.05	8.21	22,988.96	1725.91
	Follower speed	Mean	10.55	1.37	26.56	**188.13**	**181.58**	**2,933.13**	39.72
		SD	27.92	9.96	49.57	284.66	239.18	7,050.73	557.03
	Following speed	Mean	10.15	2.17	22.12	**176.88**	**152.46**	**709.65**	25.09
		Mean	10.55	1.37	26.56	**188.13**	**181.58**	**2,933.13**	39.72
	Number of lists including the user	Mean	17.80	1.90	**75.23**	**226.57**	**470.41**	**3968.36**	68.23
		SD	72.00	31.99	120.04	530.62	618.25	5281.37	515.15
	Number of lists created by the user	Mean	1.90	0.14	**6.59**	2.05	**51.69**	**63.55**	4.08
		SD	5.65	1.47	11.92	4.70	162.44	110.74	31.39
Experience	Local experience index	Mean	0.27	0.04	**0.76**	**0.58**	**0.81**	**0.88**	0.32
		SD	0.19	0.10	0.18	0.27	0.23	0.17	0.32
Activity	Tweet frequency	Mean	6.04	.53	**46.11**	**29.51**	**156.22**	**394.57**	19.54
		SD	5.84	1.51	29.36	25.87	72.98	313.24	50.52
	Announcement frequency	Mean	3.98	0.53	**21.69**	**16.96**	**46.00**	**258.98**	9.46
		SD	4.21	1.51	14.86	17.11	30.22	317.57	32.65
	@mention frequency	Mean	2.06	0.00	**24.42**	**12.55**	**110.22**	**135.59**	10.08
		SD	2.64	0.00	20.60	12.84	63.33	104.51	27.48
	#hash-tag frequency	Mean	0.26	0.00	**2.76**	1.23	**16.74**	**35.93**	1.39
		SD	0.71	0.05	3.83	1.85	21.81	49.60	6.72
	HTTP frequency	Mean	1.41	0.07	**9.70**	**7.84**	**34.09**	**218.28**	5.02
		SD	2.48	0.47	10.36	12.43	30.62	269.36	26.99
	Retweet frequency	Mean	0.23	0.00	**2.69**	**1.65**	**15.70**	**43.31**	1.39
		SD	0.54	0.00	3.92	2.75	25.73	80.03	8.57
	Branded tweet frequency	Mean	0.03	0.00	0.07	**1.14**	**0.23**	0.04	0.08
		SD	0.11	0.05	0.15	1.33	0.61	0.06	0.38
	User interactivity	Mean	**0.64**	0.00	0.32	**0.50**	0.33	**0.42**	0.36
		SD	0.26	0.00	0.15	0.28	0.20	0.16	0.33

	Tweet interactivity	Mean	**1.07**	0.00	**1.13**	**1.25**	**1.34**	**1.45**	0.75
		SD	0.14	0.00	0.17	0.59	0.54	0.88	0.56
	Contact frequency	Mean	2.26	0.00	**28.22**	**15.08**	**141.91**	**254.32**	12.61
		SD	2.96	0.00	24.16	15.50	85.51	379.06	45.32
	Interaction frequency	Mean	1.07	0.00	**8.08**	**6.24**	**40.75**	**71.52**	3.86
		SD	1.31	0.00	7.31	7.08	37.22	72.91	12.64
Impression	Tweet impression	Mean	9,939.85	129.26	152,019.35	**484,786.61**	**2,756,127.65**	**30,816,710.65**	310,129.92
		SD	40,637.89	1,590.88	306,473.44	889,398.46	3,646,832.94	28,633,294.58	3,186,059.23
	Announcement impression	Mean	6,208.78	129.26	61,778.14	**256,744.33**	**853,502.10**	**19,000,215.67**	157,902.41
		SD	25,363.44	1,590.88	119,413.60	485,870.82	1,226,240.98	23,831,787.97	2,237,084.18
	#hash-tag impression	Mean	576.93	0.46	9,651.50	**25,075.25**	**217,969.21**	**2,467,532.25**	23,604.34
		SD	3,167.21	6.19	20,574.50	74,815.68	297,971.23	3,233,301.91	302,972.10
	HTTP impression	Mean	3,373.50	13.01	35,146.22	**198,282.02**	**629,839.02**	**11,938,190.09**	102,662.58
		SD	16,704.47	118.59	73,672.08	483,062.65	912,121.36	16,210,926.77	1,479,668.59
	Retweet impression	Mean	421.60	0.03	10,909.40	**43,367.93**	**196,821.26**	**1,827,906.86**	20,302.08
		SD	1,968.08	0.76	30,838.31	119,557.24	352,827.89	2,075,909.64	215,085.93
	Branded tweet impression	Mean	9.58	0.11	80.02	**770.99**	**1,318.23**	**1,474.48**	99.61
		SD	54.27	1.61	216.76	1,290.47	3,932.04	2,849.96	813.36

a. Bold indicates the value is larger than overall mean. Underline indicates the value is the maximum among all the group means. Dotted underline indicates the value is the minimum among all the group means.

Table 4.11 presents cluster profile on those unselected categorical variables. There is no strong pattern on the *user name* except for Cluster 2, the majority (53.59%) of whom did not use a real name as their user name. On *profile picture*, about 70% to 80% of consumers in all clusters except Cluster 2 used real pictures of themselves. 55.82% of consumers in Cluster 2 did not have a picture. Roughly 60% to 70% of consumers in all clusters except Cluster 2 included an address at the city level on their profile. 71.13% of consumers in Cluster 2 did not have an address. There is no apparent pattern for *Web address* type except for Clusters 1 and 2. For Cluster 1, 51.43% consumers did not have a Web address, and for Cluster 2, 89.95% consumers did not note a Web address on their profile.

Table 4.11 Cluster profile on unselected categorical variables

Variable	Value		Cluster[a] 1	2	3	4	5	6	Total
User name	0: No real name	n	**332**	**336**	**150**	28	19	3	868
		%	**41.24%**	**53.59%**	**40.76%**	34.15%	32.20%	27.27%	44.47%
	1: Nick name	n	256	207	105	24	18	**4**	614
		%	31.80%	33.01%	28.53%	29.27%	30.51%	**36.36%**	31.45%
	2: Proper name	n	217	84	113	**30**	**22**	**4**	470
		%	26.96%	13.40%	30.71%	**36.59%**	**37.29%**	**36.36%**	24.08%
Profile picture	0: No picture	n	46	**350**	9	3	0	0	408
		%	5.71%	**55.82%**	2.45%	3.66%	0.00%	0.00%	20.90%
	1: No real person picture	n	191	63	63	13	12	2	344
		%	23.73%	10.05%	17.12%	15.85%	20.34%	18.18%	17.62%
	2: Partial real person picture	n	18	14	14	2	2	0	50
		%	2.24%	2.23%	3.80%	2.44%	3.39%	0.00%	2.56%
	3: Real person picture	n	**550**	200	**282**	**64**	**45**	**9**	1150
		%	**68.32%**	31.90%	**76.63%**	**78.05%**	**76.27%**	**81.82%**	58.91%
Location on profile	0: No location	n	143	**446**	16	8	3	0	616
		%	17.76%	**71.13%**	4.35%	9.76%	5.08%	0.00%	31.56%
	1: Abstract description	n	32	15	17	0	2	1	67
		%	3.98%	2.39%	4.62%	0.00%	3.39%	9.09%	3.43%
	2: Country	n	49	30	17	4	2	2	104
		%	6.09%	4.78%	4.62%	4.88%	3.39%	18.18%	5.33%
	3: State/Province	n	92	21	39	10	9	0	171
		%	11.43%	3.35%	10.60%	12.20%	15.25%	0.00%	8.76%
	4: City	n	**489**	115	**279**	**60**	**43**	**8**	994
		%	**60.75%**	18.34%	**75.82%**	**73.17%**	**72.88%**	**72.73%**	50.92%
Web address on profile	0: No link	n	**414**	**564**	92	**24**	7	0	1101
		%	**51.43%**	**89.95%**	25.00%	**29.27%**	11.86%	0.00%	56.40%
	1: Company/Business Website	n	159	31	87	18	20	**4**	319
		%	19.75%	4.94%	23.64%	21.95%	33.90%	**36.36%**	16.34%
	2: Multimedia or social sharing	n	3	3	6	3	0	1	16
		%	0.37%	0.48%	1.63%	3.66%	0.00%	9.09%	.82%
	3: Social network	n	56	12	28	5	4	2	107
		%	6.96%	1.91%	7.61%	6.10%	6.78%	18.18%	5.48%
	4: Blog	n	68	13	54	11	4	1	151
		%	8.45%	2.07%	14.67%	13.41%	6.78%	9.09%	7.74%
	5: Personal Website	n	105	4	**101**	21	**24**	3	258
		%	13.04%	.64%	**27.45%**	25.61%	**40.68%**	27.27%	13.22%

a. Bold indicates the dominate category in one cluster.

Table 4.12 presents cluster profile on those unselected continuous variables. Cluster 2 had incomplete biography information, but the consumers in this group were new to Twitter. However, they connected with the brand at very an early stage of their Twitter life. The other clusters had relatively complete biographies and more experience with the brand and the platform.

Table 4.12 Cluster profile on unselected continuous variables

Dimension	Variable		Cluster[a]						
			1	2	3	4	5	6	Total
Profile	Biography length by word	Mean	**8.50**	2.01	**13.04**	**12.52**	**16.39**	**14.45**	7.71
		SD	8.15	5.23	7.74	7.79	7.57	6.41	8.45
	Biography length by character	Mean	**54.55**	11.58	**85.78**	**81.09**	**107.83**	**100.82**	49.62
		SD	51.07	28.74	48.40	48.04	46.54	47.35	53.22
Experience	Twitter age	Mean	**249.09**	92.58	**361.83**	**264.43**	**389.86**	**334.91**	225.46
		SD	203.57	96.92	257.32	220.41	284.21	313.73	218.70
	Follower age	Mean	**141.99**	21.51	**210.71**	**167.49**	**242.64**	**219.53**	120.80
		SD	176.46	73.86	216.70	196.43	234.44	274.78	180.31
	Follower index	Mean	0.57	**0.90**	0.54	0.47	0.49	0.60	0.66
		SD	0.32	0.22	0.29	0.32	0.29	0.34	0.33
	Following age	Mean	**130.54**	20.05	**215.46**	**182.01**	**257.88**	**219.33**	117.57
		SD	173.77	71.46	222.62	206.08	254.89	272.02	182.32
	Following index	Mean	0.62	**0.90**	0.53	0.49	0.48	0.59	0.68
		SD	0.33	0.20	0.30	0.35	0.30	0.32	0.33

a. Bold indicates the value is larger than overall mean. Underline indicates the value is the maximum among all the group means. Dotted underline indicates the value is the minimum among all the group means.

Table 4.13 summarizes the cluster profiles presented in Tables 4.10, 4.11, and 4.12 with labels added for each cluster. Cluster 1 is named “Opportunistic Mingler”. The major characteristics of these consumers include that they are not active in tweeting but very active in interacting with different Twitter users whenever possible. Cluster 2, called “Rookie User - Brand Enthusiast”, contains consumers who are mainly new Twitter users and thus do not have complete profiles and are not actively involved in tweeting. They do, however, connect with the brand almost as soon as they sign up on the platform, which shows passion and great interest about the brand. Cluster 3, “Personal Conversationalist”, is consumers active in tweeting and interacting with people in their small tight social networks. This communication is most likely personal. Clusters 4 and 5 are named “Active User - Brand Evangelist”. The major characteristics for these consumers are that they are active in tweeting, engaging in the interaction, and owning a big social network. Moreover, they are very involved in the word-of-mouth communication and make big branded tweet impressions. Cluster 6 is the “Celebrity User”. They are famous and social people who actively engage on Twitter and own a very big social network.

Table 4.13 Cluster definition

Dimension	1: Opportunistic Mingler (41.24%)	2: Rookie User - Brand Enthusiast (32.12%)	3: Personal Conversationalist (18.85%)	4&5: Active User - Brand Evangelist (7.22%)	6: Celebrity User (0.56%)
Profile	• Has a picture of oneself on profile • Has a city level location on profile • Does not have a Web address on profile • Relatively complete biography	• Does not use real name as user name • Does not have a picture on profile • Does not have a location on profile • Does not have a Web address on profile • Incomplete biography	• Has a picture of oneself on profile • Has a city level location on profile • Complete biography	• Has a picture of oneself on profile • Has a city level location on profile • Complete biography	• Has a picture of oneself on profile • Has a city level location on profile • Complete biography
Connection	• Small social network	• Very small social network	• Small social network with strong ties	• Big social network	• Very big social network with weak ties
Experience	• Rich platform experience • Active about 30% of time • Connecting with the brand for a long time	• Little platform experience • Inactive most of the time • Connecting with the brand soon after they sign up on Twitter	• Rich platform experience • Active about 60% to 70% of time • Connecting with the brand for a very long time	• Rich platform experience • Active about 80% of time • Connecting with the brand for a very long time	• Rich platform experience • Active about 90% of time • Connecting with the brand for a very long time
Activity	• Not active in tweeting or interacting with	• Not very active in tweeting or interacting with	• Active in tweeting and interacting with other users	• Active in tweeting and interacting with other users	• Very active in tweeting and interacting with

	other users • Very active in interacting with different Twitter users whenever possible • Not active in engaging in word-of-mouth communication	other users • Not very active in engaging in word-of-mouth communication	• Not active in engaging in word-of-mouth communication	• Active in engaging in word-of-mouth communication	other users • Not active in engaging in word-of-mouth communication
Impression	• Very low tweet impression • Very low branded tweet impression	• Minimum tweet impression • Minimum branded tweet impression	• Low tweet impression • Low branded tweet impression	• Big tweet impression • Big branded tweet impression	• Very big tweet impression • Very big branded tweet impression

Discussion

In this chapter, I profiled consumers on Twitter, particularly those with connection to specific brands. I came up with 49 variables in 5 dimensions to describe the consumers (Table 4.1). The 5 dimensions included *profile*, *connection*, *experience*, *activity*, and *impression*. Among the 49 variables, 15 variables were used for calculation, and 34 variables are used for clustering. I applied a forward-selection process by using silhouette coefficient to evaluate each variable in 34 variables. This variable selection procedure gave me 23 most informative variables, including all the variables in the *connection*, *activity*, and *impression* dimensions, and *local experience index* in the *experience* dimension (Table 4.6). Then I used the TwoStep cluster procedure to perform cluster analysis based on these 23 variables. I identified 6 clusters of consumers (Tables 4.8 and 4.9). I put 5 labels on them including Opportunistic Mingler (41.24%), Rookie User - Brand Enthusiast (32.12%), Personal Conversationalist (18.85%), Active User - Brand Evangelist (7.22%), and Celebrity User (0.56%) (Table 4.13).

On relationship with the brand, 3 groups — Brand Enthusiast, Brand Evangelist, and Celebrity User — are close to the business. Brand Enthusiasts are consumers choosing to connect with the brand very early in their Twitter journey, which indicates their deep awareness about the brand and the brand Twitter account, and their preference for the brand. This group of consumers may explain why Shang, Chen, and Liao (2006) found that lurking contributed to brand loyalty. It is likely that consumers like Brand Enthusiasts are already loyal to the brand before joining the community, and connecting with the brand and receiving the messages from the brand can enhance the branded cues and the brand loyalty. These consumers may appear to have weak ties with the community, but they have strong ties with the brand. They will not mind receiving branded messages. In fact, they might expect branded messages to gain knowledge and up-to-date information about the brand. Since this group of consumers is very new to Twitter, some of them may become Brand Evangelists after they get familiar with the platform and accumulate enough brand news.

Brand Evangelists are similar to core members in de Valck, van Bruggen, and Wierenga's (2009) study, and insiders in Kozinets' (1999) model. They are social, active, and engaging in the community. They are the civilian market mavens on Twitter. They also have knowledge about the brand and would like to participate in word-of-mouth communication. They have resources and abilities to be brand evangelists, and thus, they are perfect candidates for spreading brand messages. They are the consumers most likely to talk about the branded experience when using or trying out the product or service. They are the consumers businesses want to send coupons or sample products to. They are the consumers the business owner wants to see at social events like tweet-up.

Two major characteristics can identify Brand Evangelists from other Twitter regular users (Cluster 1-3). They are active in their community and own a big social network. It makes sense that Brand Evangelists are socialable and outgoing since it is almost impossible to find an introverted person who talks about a brand all the time. But Personal Conversationalists are also very social. What differentiates Brand Evangelists from Personal Conversationalists is the size of their social networks. Brand Evangelists own a big social network with mixed weak ties and strong ties. The reason they can possibly maintain a network at such a scale is they offer interesting or profound tweets. In addition, they may enjoy tweeting and commenting on many different things including brand, product, and service. Thus, these two personalities — being socialable and enjoying offering comments — make them natural brand evangelists.

Celebrity users are the star market mavens on Twitter. They are usually domain experts, leaders, or celebrities in the entertainment industry. They own a huge amount of audience, which makes them important to businesses. However, they are not active in word-of-mouth communication. On the other hand, since their identities are known, their word-of-mouth messages may be more effective if tying closely with their background and domain expertise. For example, Brittany Spears comments on the iPhone can only be accepted by her followers when the content pertains to the non-technical aspects. Similarly, an established computer scientist tweeting about Starbucks coffee may not catch his/her followers' attention, but he/she may be more successful when tweeting about the technical aspects of the iPhone.

Opportunistic Minglers and Personal Conversationalists are a little distant from the brand. Opportunistic Minglers are the typical lurking group in the community. These consumers are like tourists in Kozinets' (1999) model. Most of the time, they are simply wandering in the community and receiving tweets. But they may take advantage once in a while to communicate with different people. They are mainly passive information receivers on Twitter. I do not see any possibility where they can contribute in word-of-mouth communication online, but they may be involved in word-of-mouth communication offline.

Personal Conversationalists focus more on interpersonal communication and relationships. This group of consumers owns a social network with strong ties. They are most likely to pay attention to their friends' word rather than strangers. Therefore, friend recommendation message like "you friend like this brand or use that product" from Facebook may work the best on them.

Conclusion

On this research question, I concluded that there are 5 types of consumers connecting with the brand on Twitter, namely Opportunistic Mingler (41.24%), Rookie User and Brand Enthusiast (32.12%), Personal Conversationalist (18.85%), Active User and Brand Evangelist (7.22%), and Celebrity User (0.56%). Twitter can develop systems to help the business user to classify consumers into these different groups. The business can use this classification to develop personalized and effective brand and customer management strategies, and marketing campaigns.

Summary

Maintaining a brand presence on social networks is a critical marketing strategy for businesses, possibly enhancing awareness, increasing conversions, and enlarging the customer base. Understanding consumers in the online community is an integral aspect of maintaining an effective presence in social networks, benefitting businesses' brand and customer management strategies. In this chapter, I profiled consumers on Twitter, particularly those connecting to company brand accounts. I identified 49 variables to describe the consumers in 5 dimensions: profile, connection, experience, activity, and impression. Among the 49 variables, 15 variables were used for calculation and 34 variables were used for clustering. I applied a forward-selection process by using the silhouette coefficient to evaluate each variable. This variable selection procedure gives me 23 most informative variables including all the variables in the *connection*, *activity*, and *impression* dimensions, and *local experience index* in the *experience* dimension. Then I used the TwoStep cluster procedure to perform cluster analysis to achieve 6 clusters. I put 5 labels on the clusters: Opportunistic Mingler (41.24%), Rookie User and Brand Enthusiast (32.12%), Personal Conversationalist (18.85%), Active User and Brand Evangelist (7.22%), and Celebrity User (0.56%).

CHAPTER 5. CONCLUSION

Local, national, and global commercial businesses are increasingly interested in leveraging Twitter to present the brand, manage WOM communication, and interact with consumers. In response to such interests, numerous entrepreneurs developed Twitter business applications, external to Twitter itself. Some applications, such as CoTweet (http://cotweet.com/) and HootSuite (http://hootsuite.com/), assist companies in managing the Twitter communication platform and WOM conversation channel. In addition to these applications, there are also many online publications or printed books pertaining to the topic of using Twitter for business. However, most of these applications and publications are based on intuitive ideas and lack support of solid and comprehensive understanding of the platform.

This dissertation provides an in-depth analysis of WOM communication among consumers and businesses on Twitter and uncovers Twitter community dynamics from a business perspective. The goal of this dissertation is to bridge research and practice, explore social network analysis metrics, and make actionable recommendations for Twitter business users. This dissertation covers three aspects important to current and prospective business Twitter users, namely benefit (i.e., what can a business get from Twitter?), role (i.e., how active should a business be on Twitter?), and audience (i.e., who connects to a business on Twitter?). Addressing these questions make important contributions to current understanding of consumers and WOM conversations, which informs how to design the corresponding marketing and management strategies. In the rest of this final chapter, I present the answers to the research questions, discuss the potential contributions of my dissertation, highlight several insights for business to use Twitter, and describe my future research ideas.

Findings and Answers

In this dissertation, I am pursuing the answers to three fundamental questions with regards to businesses using Twitter for marketing, advertising, branding, and word-of-mouth management purposes. The questions and answers are as follows:

Research question 1: What are the branding influences of social network on word-of-mouth communication on Twitter? Being the brand's followers or the brand's followings have main effect but not interaction effect on the volume of WOM messages and WOM communication with other consumers and the brand. To be more specific, consumers who follow or are followed by a brand send out more WOM messages and participate more in the conversation about the brand with other consumers and the brand. However, being the brand's followers and the brand's followings have no statistically significant influence, no main effect, and no interaction effect on the number of retweet WOM messages from other consumers and the brand, the number of WOM messages with brand hash-tagged, and the number of WOM messages with hyperlinks

Research question 2: What are the influences of brand engagement in word-of-mouth communication on consumers' level of engagement in word-of-mouth *communication on Twitter?* A business that actively engages in the WOM communication process as a participant can increase the engagement level of consumers in its immediate social network, which in turn increases the engagement level of consumers outside the brand's immediate social network. Retweeting, as an explicit way to show consumers reactions to the brand engagement on WOM channel, demonstrates that the brand influence only reaches consumers with a second-degree relationship to the brand.

Research question 3: What are the characteristics of consumers connecting to brands in the Twitter community? There are five types of consumers in the brand's immediate social network on Twitter: Opportunistic Mingler (41.24%), Rookie User and Brand Enthusiast (32.12%), Personal Conversationalist (18.85%), Active User and Brand Evangelist (7.22%), and Celebrity User (0.56%). Opportunistic Minglers actively engage in conversations with other Twitter users whenever possible but do not actively participate in tweeting without addressing specific people. Rookie User and Brand Enthusiast is a group of consumers who are new Twitter users, own incomplete profiles, and do not actively tweet, but they do connect with the brand very shortly after signing up on Twitter. Personal Conversationalist includes consumers who are tweeting and chatting with friends in their small tight social networks. Active Users and Brand Evangelists, as a group, are very social in terms of actively tweeting, actively interacting with others, and owning a big social network. In addition, they are very engaging in the WOM communication and make big branded tweet impressions. Celebrity Users, consumers with big fame, actively tweet and have a very large social network.

Contributions and Implications

My dissertation contributes to both academia and industry. To academia, my dissertation advances our understanding about the dynamics in the Twitter community and other social media sites, the trajectories of information dissemination, and the interactions between members in the Twitter network and other social networks.

My dissertation can also assist client-side companies, both system end users as well as service providers, and Twitter by helping them to uncover effective ways to use Twitter or other similar services in their efforts to interact with consumers in meaningful and profitable ways.

As a service provider, Twitter hosts the platform on which brands and customers have conversations. It provides different features to enable such communications. My dissertation can help Twitter understand this interaction between brands and consumers so that it can provide better support for clients and improve its services.

Six Do's for Business

Throughout the dissertation, I make practical and actionable recommendations for businesses who wish to engage in online WOM, using platforms such as Twitter. These suggestions are based on solid research findings. I highlight six recommendations here.

1. Maintain brand presence on different social media sites

In Chapter 3, I discuss the central role that proximity to the communication channel plays in WOM message diffusion. The brand surely used some other communication channels before Twitter and should continue to use those channels. However, there is a major boost in the WOM volume about the brand on Twitter immediately after it launches its account. This demonstrates the dramatic influence of the brand by getting close to the channel.

2. Connect with as many consumers as possible as followers or followings on Twitter

In Chapter 2, I present that consumers who follow or are followed by brands tweet more about the brand, product, or service than do consumers without connection to the brand. The connections also motivate these consumers to engage in conversations about the brand with their friends and the business.

3. Follow back to consumers requesting to follow the brand on Twitter

In Chapter 2, I discuss that the relationship of the brand following consumers is more influential in motivating consumers to tweet about the brand, product, or service than is the relationship of consumers following the brand. Having the brand to follow consumers creates emotional bonds between the brand and consumers and, from the consumers' perspective at least, brings the brand closer.

4. Tweet every 1.5 to 4 hours

In Chapter 3, I show that 75% of retweeting activity happens within 1.5 hours and almost all of retweeting takes place within 4 hours, indicating that tweet consumption happens within 1.5 to 4 hours. Therefore, businesses who wish to stay active on Twitter should tweet at least every 1.5 to 4 hours in order to seize the consumers' attention.

5. Tweet in a way to spark strong emotions by using humor or inspiration

In Chapter 3, I present a list of highly retweeted messages, which are mostly humorous, anecdotal, and philanthropic. These are the popular tweet themes and suggest the most effective ways to craft tweets.

6. Keep in mind the five types of consumers in the brand's immediate social network

In Chapter 4, I identify five types of consumers in the brand's community on Twitter. These five types are Opportunistic Mingler, Rookie User and Brand Enthusiast, Personal Conversationalist, Active User and Brand Evangelist, and Celebrity User. Opportunistic Mingler has the least value to the brand. Rookie User and Brand Enthusiast has great interest and preference in the brand. Personal Conversationalist likes to be communicated with by using his/her friend's name. Active User and Brand Evangelist is a very important group of consumers to the brand. The consumers from this group are social people and market mavens. Celebrity User is the rock star influential on Twitter.

Future Research

I am interested in continuing this line of research on understanding information diffusion and consumers' interaction from a business perspective. First, I want to refine the measurement and analytics. In this dissertation, I don't differentiate the size of Twitter accounts, which may cause the measurement inaccuracy since it is reasonable to guess that H&R Block with several thousands of people in its community can be less influential than Whole Foods with more than a million people in its community. In addition, taking communication context into consideration can also refine the measurement and analytics.

Second, I want to extend my dissertation work by incorporating experiments and surveys. My dissertation work primarily focuses on analyzing tweets, Twitter trace data. According to Jansen (2009), collecting trace data will not interfere with the natural flow of behavior and events in the contexts. Analyzing data can get the natural human behaviors, which makes very valuable research contributions. However, there are certain behaviors that trace data cannot capture. Experiments and surveys are more effective methods for gathering such behaviors. Participants can be asked directly what they are doing and what they are thinking. Therefore, running experiments and surveys can be a good extension of my dissertation work on information diffusion and consumers' interaction by uncovering consumers' psychological behaviors or behaviors not captured in trace data.

Third, it will be interesting to go beyond Twitter and study other social media platforms. "Not all social networks are equal" (Allsop, et al., 2007, p. 399). There are so many different types of social media platforms, as I list in Chapter 2. I am interested in investigating how information propagates on different platforms and how consumers interact in different online communities.

Fourth, I am interested in developing an analytics system design to make the gathering and analysis of trace data more feasible for businesses. Most of the analyses in this dissertation involve complex statistical methods. In the real world, it will be impossible for practitioners to master all of these methods. An analytics system, which can provide insights about the platform and return analysis results in a reasonable time frame, would be very beneficial to businesses.

REFERENCES

Allsop, D. T., Bassett, B. R., & Hoskins, J. A. (2007). Word-of-mouth research: Principles and applications. *Journal of Advertising Research*, 47(4), 398-411.

Arndt, J. (1967a). Word-of-mouth advertising and informal communication. In D. F. Cox (Ed.), *Risk taking and information handling in consumer behavior* (pp. 188-239). Cambridge, MA: Harvard University Press.

Arndt, J. (1967b). *Word of mouth advertising : A review of the literature* New York: Advertising Research Foundation.

AtGoogleTalks (Producer). (2010, June 4, 2010) @Google & YouTube present a conversation with Conan O'Brien. Podcast retrieved from http://www.youtube.com/watch?v=u7TwqpWiY5s.

Balasubramanian, S., & Mahajan, V. (2001). The economic leverage of the virtual community. *International Journal of Electronic Commerce*, 5(3), 103-138.

Barnes, N. G., & Mattson, E. (2009). *Social media in the 2009 Inc. 500: New tools & new trends*. Dartmouth, MA: Center for Marketing Research at the University of Massachusetts, Dartmouth. Retrieved January 6, 2010, from http://www.umassd.edu/cmr/studiesresearch/socialmedia2009.pdf

Bharati, P., & Tarasewich, P. (2002, May). Global perceptions of journals publishing e-commerce research. *Communications of the ACM*, 45(5), 21-26.

Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, 15(3), 31-40.

bigmouthmedia, & Econsultancy (2009). *Social media and online PR report*. London: Econsultancy.

Boone, L. E., & Kurtz, D. L. (2008). *Contemporary marketing* (13 ed.). Mason, OH: Thomas High Education.

boyd, d. m., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship *Journal of Computer-Mediated Communication*, 13(1), Article 11. Retrieved November 30, 2009, from http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of Interactive Marketing*, 21(3), 2-20.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York: Routledge.

Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM,

Comcast (n.d.). About Comcast Corporation Retrieved January 9, 2010, from http://www.comcast.com/Corporate/About/Corporateinfo/Corporateinfo.html?lid=1AboutCorporateInfo&pos=Nav&

de Chernatony, L. (2000). Succeeding with brands on the Internet. *Brand Management*, 8(3), 186-195.

de Valck, K., van Bruggen, G. H., & Wierenga, B. (2009). Virtual communities: A marketing perspective. *Decision Support Systems*, 47(3), 185-203.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Marketing Science*, 49(10), 1407-1424.

Fowlkes, E. B., Gnanadesikan, R., & Ketternring, J. R. (1988). Variable selection in clustering. *Journal of Classification*, 5(2), 205-228.

Gelt, J. (2009). Kogi Korean BBQ, a taco truck brought to you by Twitter Retrieved January 10, 2010, from http://www.latimes.com/features/la-fo-kogi11-2009feb11,0,4771256.story

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science*, 23(4), 545-560.

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., et al. (2005). The firm's management of social interactions. *Marketing Letters*, 16(3/4), 415-428.

H&R Block (n.d.). H&R Block | our company | H&R Block as a whole Retrieved January 11, 2010, from http://www.hrblock.com/company/index.html

Ha, H.-Y. (2004). Factors influencing consumer perceptions of brand trust online. *Journal of Product & Brand Management*, 13(5), 329-342.

Haigie, R. A., Feick, L. F., & Price, L. L. (1987). Types and amount of word-of-mouth communications about retailers. *Journal of Advertising*, 63(3), 260-278.

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*, 18(1), 38-52.

Holt, D. B. (2004). *How brands become icons: The principles of cultural branding*. Boston, MA: Harvard Business School.

Ionescu, D. (2009). Dell proves that Twitter can be profitable Retrieved January 6, 2010, from http://www.pcworld.com/article/184076/dell_proves_that_twitter_can_be_profitable.html

Israel, S. (2009). 20 ways businesses use Twitter Retrieved January 6, 2010, from http://images.businessweek.com/ss/09/10/1006_twitterville/1.htm

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), 264-323.

Jansen, B. J. (2009). *Understanding user-Web interactions via Web analytics*. San Rafael, CA: Morgan & Claypool.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdhury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of American Society for Information Science*, 60(11), 2169-2188.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis: An introduction to cluster analysis*. New York: Wiley-Interscience.

Keller, E. (2007). Unleashing the power of word of mouth: Creating brand advocacy to drive growth. *Journal of Advertising Research*, 47(4), 448-452.

Kling, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Kozinets, R. V. (1999). E-tribalized marketing?: The strategic implications of virtual communities of consumption. *European Management Journal*, 17(3), 252-264.

Lampe, C., Ellison, N., & Steinfield, C. (2007). A familiar face(book): Profile elements as signals in an online social network. In B. Begole, S. Payne, E. Churchill, R. St. Amant, D. Gilmore & M. B. Rosson (Eds.), *Proceedings of the Annual SIGCHI Conference on Human Factors in Computing Systems* (pp. 435-444). New York,

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458-468.

Lleti, R., Ortiz, M. C., Sarabia, L. A., & Sanchez, M. S. (2004). Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Analytica Chimica Acta*, 515(1), 87-100.

Mangold, W. G., Miller, F., & Brockway, G. R. (1999). Word-of-mouth communication in the service marketplace. *Journal of Service Marketing*, 13(1), 73-89.

McColl, S. (2009). The Bon Appétit Awards 2009 Retrieved January 10, 2010, from http://www.bonappetit.com/magazine/baawards/2009/10/baa_2009

Milstein, S., Chowdhury, A., Hochmuth, G., Lorica, B., & Magoulas, R. (2008). *Twitter and the micro-messaging revolution: Communication, connections, and immediacy—140 characters at a time (O'Reilly Radar Report)*. Sebastopol, CA: O'Reilly Media.

Montgomery, D. C. (2008). *Design and analysis of experiments* (7th ed.). Hoboken, NJ: John Wiley &Sons.

Naked Pizza. (June 1, 2009, January 10, 2010). We thinks we <3 Twitter. http://www.blog.nakedpizza.biz/twitter.htm.

Naked Pizza (n.d.). Naked Pizza New Orleans prebiotics probiotics multi-grain Retrieved January 10, 2010, from http://www.nakedpizza.biz/mission.htm

Nielson (2007). *'Word-of-mouth' the most powerful selling tool: Nielsen global survey*. New York: Author. Retrieved January 6, 2010, from http://asiapacific.acnielsen.com/site/documents/TrustinAdvertisingOct07.pdf

Olobatuyi, M. E. (2006). *A user's guide to path analysis*. Lanham, MD: University Press of America, Inc.

Parr, B. (January 20, 2010). Bill Gates surpasses 100,000 Twitter followers in 8 hours Retrieved January 21, 2010, from http://mashable.com/2010/01/19/bill-gates-100000/

Patti, C. H., & Chen, C. H. (2009). Types of word-of-mouth messages: Information search and credence-based service. *Journal of Promotion Management*, 15(3), 357-381.

Phelps, J. E., Lewis, R., Mobilio, L., Perry, D., & Raman, N. (2004). Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email. *Journal of Advertising Research*, 44(4), 333-348.

Qualman, E. (2009, Social media ROI examples & video. http://socialnomics.net/2009/11/12/social-media-roi-examples-video/.

Raban, D. R., & Rabin, E. (2009). Statistical inference from power law distributed web-based social interactions. *Internet Research*, 19(3), 266-278.

Rao, L. (December 15, 2009). No pick-up in Twitter's U.S. traffic in November Retrieved January 6, 2010, from http://www.techcrunch.com/2009/12/15/no-pick-up-in-twitters-u-s-traffic-in-november/

Reichheld, F. F. (2001). *The loyalty effect: The hidden force behind growth, profits, and lasting value*. Boston, MA: Harvard Business School Press.

Sagolla, D. (2009). *140 characters: A style guide for the short form*. Hoboken, NJ: John Wiley & Sons.

Schmitt, G. (2009). *FEED 2009: The Razorfish digital brand experience report*. San Francisco, CA: Razorfish. Retrieved November 13, 2009, from http://feed.razorfish.com/downloads/Razorfish_FEED09.pdf

Schonfeld, E. (January 18, 2010). Getting a handle on the size of the realtime Web And the Twitter ecosystem Retrieved January 18, 2010, from http://www.techcrunch.com/2010/01/18/realtime-web-size-twitter-ecosystem/

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. New York: John Wiley &Sons.

Shang, R.-A., Chen, Y.-C., & Liao, H.-J. (2006). The value of participation in virtual consumer communities on brand loyalty. *Internet Research*, 16(4), 398-418.

Social Media B2B (2009). B2B lead generation using social media Retrieved January 7, 2010, from http://socialmediab2b.com/2009/07/b2b-lead-generation-social-media/

SPSS Inc (2004). *TwoStep Cluster*. Chicago, Illinois: Author. from http://support.spss.com/ProductsExt/SPSS/Documentation/Statistics/algorithms/14.0/twostep_cluster.pdf

Strauss, N., & Buss, A. (2009). *Online community handbook: Building your business and brand on the Web*. Berkeley, CA: New Riders.

Sun, T., Youn, S., Wu, G., & Kuntaraporn, M. (2006). Online word-of-mouth (or mouse): An exploration of its antecedents and consequences. *Journal of Computer-Mediated Communication*, 11(4), 1104-1127.

Sundaram, D. S., Mitra, K., & Webster, C. (1998). Word-of-mouth communications: A motivational analysis. *Advances in Consumer Research*, 25(1), 527-531.

Talbot, D. (January 13, 2010). How Google ranks Tweets Retrieved April 20, 2010, from http://www.technologyreview.com/web/24353/?nlid=2664&a=f

The Coffee Groundz (n.d.). About - the Coffee Groundz - Houston, TX - coffee, tea, gelato, spirits, free WiFi and so much more! Retrieved January 9, 2010, from http://thecoffeegroundz.com/about

The Home Depot (2008). Business code of conduct and ethics Retrieved January 10, 2010, from http://ir.homedepot.com/phoenix.zhtml?c=63646&p=irol-govConduct

The Whole Foods (n.d.). Our core values | WholeFoodsMarket.com Retrieved January 11, 2010, from http://www.wholefoodsmarket.com/values/corevalues.php

Thibaut, J. W., & Kelley, H. H. (1975). *The social psychology of groups*. New York: Wiley.

Vignette (2009). *Social media in the enterprise: A step-by-step roadmap for success*: Author. Retrieved January 7, 2010, from http://www.vignette.com/dafiles/docs/Downloads/Social-Media-in-the-Enterprise.pdf

Walker, T., Bradford, R., & Resnicow, J. (2009). *Best practices for business*.

Wangenheim, F. v., & Bayon, T. (2004). The effect of word of mouth on services switching: Measurement and moderating variables. *European Journal of Marketing*, 38(9/10), 1173-1185.

Wauters, R. (December 9, 2009). Twitter spawned 50,000 apps to date, will open up firehose for more Retrieved January 6, 2010, from http://www.techcrunch.com/2009/12/09/twitter-le-web-2009/

Weber, L. (2009). *Marketing to the social web: How digital customer communities build your business* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Weil, K. (February 22, 2010). Measuring tweets Retrieved February 22, 2010, from http://blog.twitter.com/2010/02/measuring-tweets.html

Wikipedia (n.d.-a). Comcast Retrieved January 9, 2010, from http://en.wikipedia.org/wiki/Comcast

Wikipedia (n.d.-b). The Home Depot Retrieved January 11, 2010, from http://en.wikipedia.org/wiki/Home_Depot

Wikipedia (n.d.-c). Zappos.com Retrieved January 10, 2010, from http://en.wikipedia.org/wiki/Zappos

Zappos (n.d.). Zappos core values Retrieved January 10, 2010, from http://about.zappos.com/our-unique-culture/zappos-core-values

Zhang, T., Ramakrishnon, R., & Livny, M. (1996). BIRCH: An efficient data clustering method for very large databases. In *Proceedings of the ACM SIGMOD Conference on Management of Data* (pp. 103-114). New York: ACM,

Zhou, B., & Wong, W. H. (Forthcoming). A bootstrap-based non-parametric ANOVA method with applications to factorial microarray data. *Statistica Sinica*.

APPENDIX A. BRAND BRIEFING

The Coffee Goundz (@CoffeeGroundz)

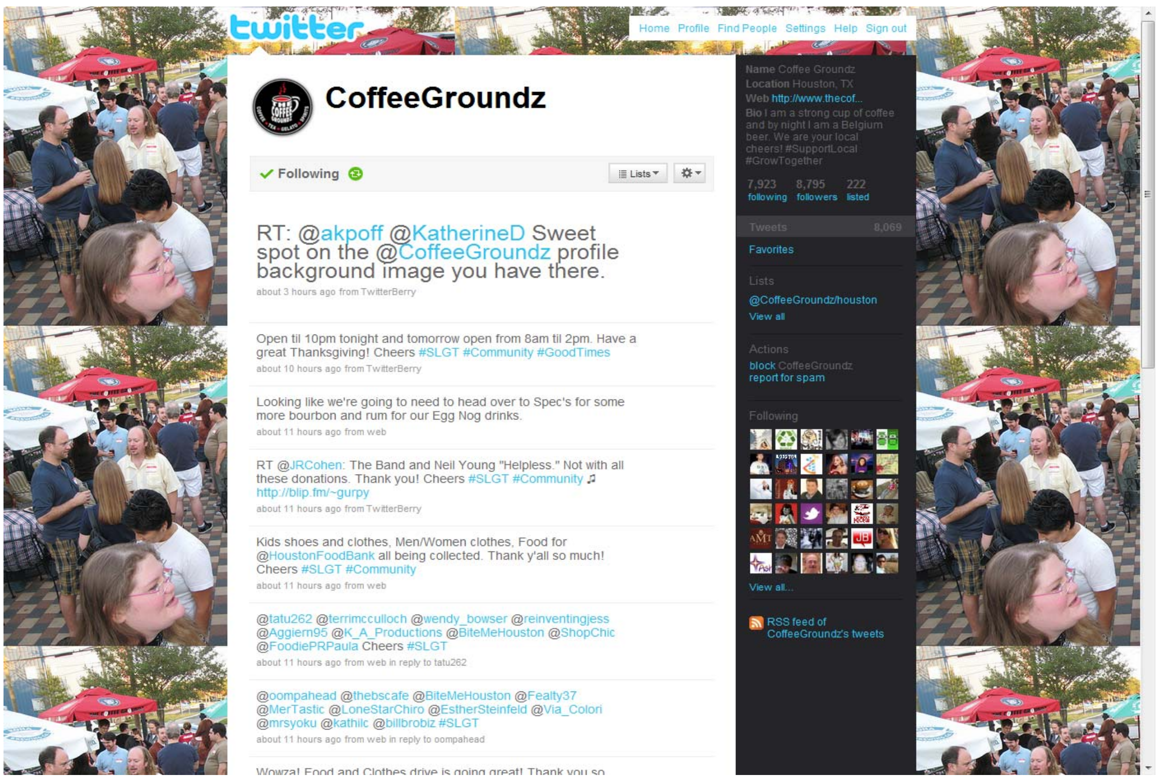


Figure A.1. Screenshot of Coffee Groundz Twitter account (@CoffeeGroundz)

The Coffee Groundz is an independent local coffee house in the Midtown of Houston, Texas, which serves coffee, tea, alcoholic beverages, baker items, gelato, sandwiches, and salads. The brand highlights three concepts: being local, free Wi-Fi, and social media. According to its Website, the business was started on a premise that “the Houston Area needed its own brand of coffee shop – a coffee shop that was inspired by the theme and the spirit of a European Cafe” (The Coffee Groundz, n.d.). Its coffee is locally roasted and even its gelato is locally made. Being local also means local community. Its owners believe their customers should enjoy free Wi-Fi after paying for their products. The brand is very social media savvy and maintains its presence on the major social media channels including Twitter, Facebook, LinkedIn, Flickr, and YouTube. It incorporates these social media channels on the Website. Especially for Twitter, its live tweet stream is displayed on almost all pages on the Website. To the Coffee Groundz, using social media really means to be social and connect with its customers and to create online community. The store manager J.R. Cohen not only interacts with their Twitter followers online but also tries to meet them in person to strengthen the social ties (Walker, Bradford, & Resnicow, 2009). He is a believer of the idea that social media is a platform for people to socialize online and offline (Walker, et al., 2009). On Twitter the shop announces upcoming events like concerts and tweetups, which have about 200 people attending each month and ultimately converts some participants to regular customers (Israel, 2009).

Comcast (@comcastcares)

Figure A.2. Screenshot of Comcast Twitter account (@comcastcares)

Comcast is the largest cable service provider in America. The business claims to focus on product quality: “deliver digital services, provide faster Internet and clearer broadband phone service, and develop and deliver innovative programming” (Comcast, n.d.). However, as a cable company, product quality certainly is important, but customer service or the experience is even more important. There is minimal text about their goals with regard to customer service, company culture, and business concepts on their official Website. The brand loyalty is largely due to their monopoly position in the market. Most of the time, they are the only service provider in an area. The customers have no choice but use them in many situations. The brand has a number of problems in terms of customer service, network neutrality, HDTV quality, signal intrusion and accidental transmission of adult content (Wikipedia, n.d.-a). The brand maintained a very active Twitter account in 2008 and launched a blog named Comcast Voices (blog.comcast.com) to serve as place to communicate with customers in March 2009. On the blog its Twitter account (@comcastcares) stream is displayed. On Twitter customers can watch @comcastcares trying to solve problems. The American Customer Satisfaction Index reports an increase of 9.3% in satisfaction for Comcast and the only change, according to the survey, was using Twitter (Israel, 2009).

Home Depot (@HomeDepot)

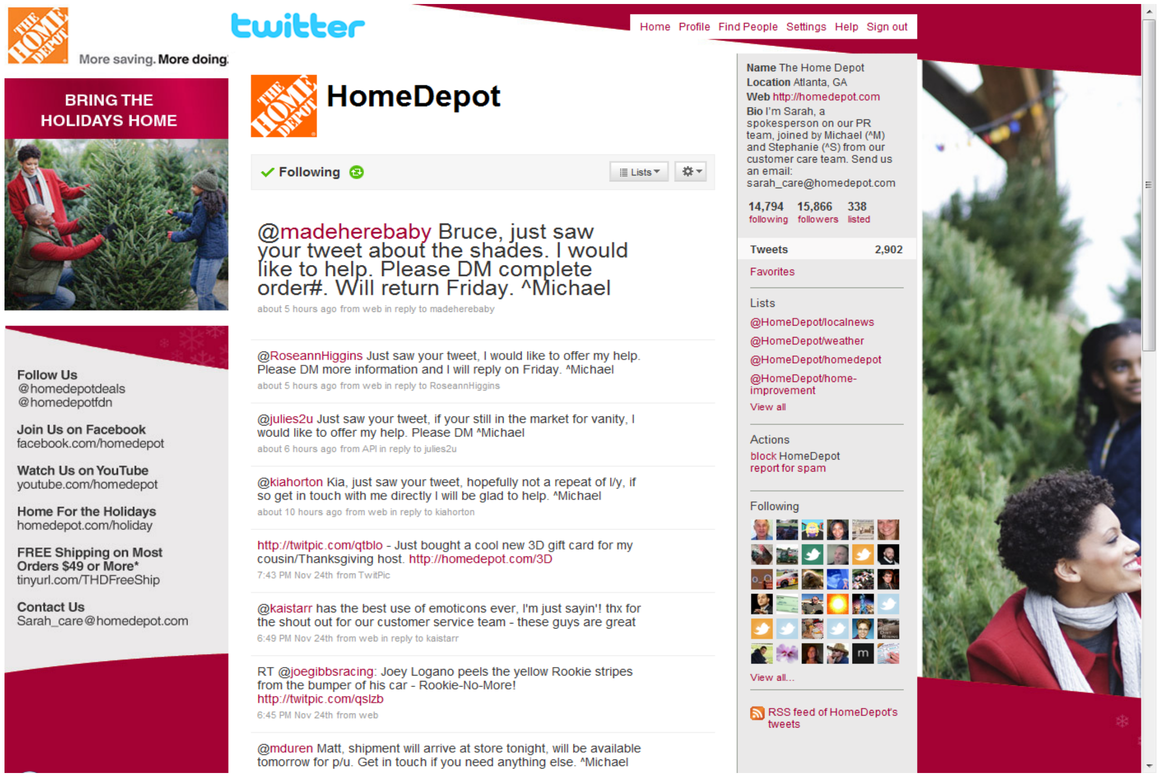


Figure A.3. Screenshot of Home Depot Twitter account (@HomeDepot)

Home Depot is a retailer of home improvement and construction products and services (Wikipedia, n.d.-b). Its slogan is “More saving. More doing.” introduced on March 18, 2009, to replace the old one “You can do it. I can help.” used since 2003 (Wikipedia, n.d.-b). Both of these slogans highlight the do-it-yourself (DIY) spirit. The brand emphasizes the value, community, people and business, which in turn believes its brand value is achieved via “Doing the Right Thing, having Respect for all People, building Strong Relationships, Taking Care of Our People, Giving Back, providing Excellent Customer Service, Encouraging Entrepreneurial Spirit and providing strong Shareholder

Returns" (The Home Depot, 2008). Highlighting people and community as core brand values makes social media an indispensible part of the business. The company has its brand presence on Twitter, Facebook and YouTube. These accounts are displayed on the social media profile pages as well as on its official retail Website.

H&R Block (@HRBlock)

Figure A.4. Screenshot of H&R Block Twitter account (@HRBlock)

H&R Block is a tax preparation service provider, which also provides some finance consulting services. The brand stresses the customer, the quality of service, and the availability and flexibility to provide service. (H&R Block, n.d.) On their official

Website, there is a brand community where the customers can ask questions and the brand shares tax tips. The brand has a presence on most major social media sites like Twitter, Facebook and YouTube.

Kogi BBQ (@kogibbq)

Figure A.5. Screenshot of Kogi BBQ Twitter account (@kogibbq)

Kogi BBQ is a mobile food truck company selling Korean and Mexican fusion food in Los Angeles. The brand highlights fusion, street food culture and social media savvy. The brand's value is recognized by the Bon Appétit Awards 2009 (McColl, 2009). The Los Angeles Times comments on the social effect introduced by the brand's on-the-go style, “The truck and its staff of merry makers have become a sort of roving party, bringing people to neighborhoods they might not normally go to, and allowing for interactions with strangers they might not otherwise talk to.” (Gelt, 2009) The brand has a popular presence on Twitter. In (Walker, et al., 2009), Alice Shin, their creative director, echoed how the brand started using Twitter back in November 2009. On the Thanksgiving weekend, Kogi BBQ catered a restaurant industry conference, but no one came to eat their food. Alice tweeted their location on Twitter. Within 30 seconds, they started to see a crowd coming to their truck while looking at their phones. Alice also commented that 60% of their customers came from Twitter on that day and most of their customers now follow the brand on Twitter.

Naked Pizza (@NAKEDpizza)

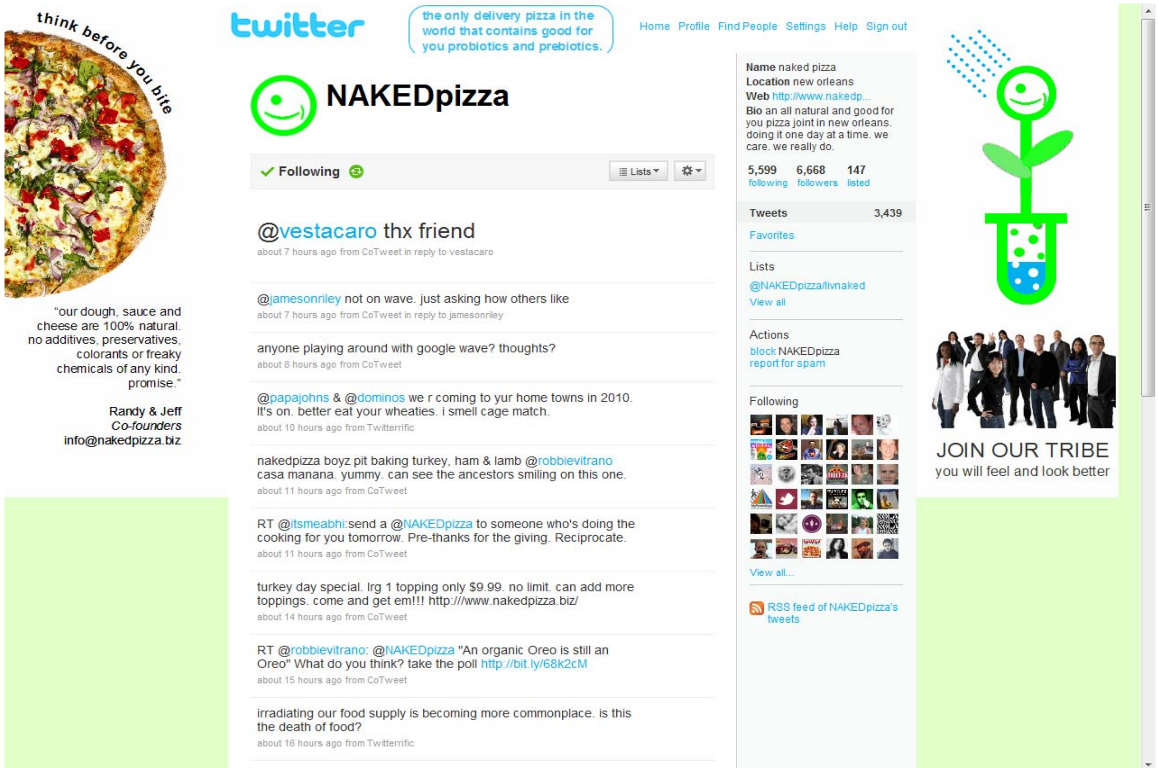


Figure A.6. Screenshot of Naked Pizza Twitter account (@NAKEDpizza)

Naked Pizza is a Pizzeria in New Orleans that sells a healthy version of pizza. The brand emphasizes nutrition and health (Naked Pizza, n.d.). It aims to demonstrate "pizza does not have to be part of the problem in our national epidemic of obesity and chronic disease, but, in fact, can be part of the solution" (Naked Pizza, n.d.). The brand is also very savvy on social media. They have Twitter and Facebook accounts, which are listed on their Website. Twitter plays a big role in their business. The brand did an experiment on May 29, 2009, and found that 68.6% of their sales come from people on Twitter saying "I'm calling from Twitter" and 22 (85%) out of 26 their new customers were from Twitter (Naked Pizza, June 1, 2009; Qualman, 2009; Walker, et al., 2009).

Starbucks (@Starbucks)

Figure A.7. Screenshot of Starbucks Twitter account (@Starbucks)

Starbucks is a world-famous coffeehouse chain (Wikipedia, 2009). English teacher Jerry Baldwin, history teacher Zev Siegel, and writer Gordon Bowker founded Starbucks in 1971, with its first store in Pike Place Market in Seattle, Washington (McGraw Hill Higher Education, n.d.) The company aims to be "the premier purveyor of the finest coffee in the world" (Starbucks, 2008). Its products include coffee, handcrafted beverages, fresh food, coffee-related merchandise and related items (Starbucks, 2008). Starbucks is active on the Web and is keen on building up its online community, as evidenced by accounts on Twitter, Facebook and YouTube. It also built up its own online communities like My Starbucks Idea (mystarbucksidea.force.com) for collecting ideas to improve their products and service, and Starbucks V2V (www.v2v.net/starbucks) for getting people together to volunteer for community work. As such, Starbucks appears to be a company interested in the social networking communities.

@Starbucks is managed by Brad Nelson, who tries to make his tone on Twitter as if people are sitting near you in the coffee shop (Israel, 2009). Nelson believes Starbucks' Twitter accounts are an "extension of CEO Howard Schultz's favorite public statement 'Starbucks is more about people than coffee'" (Israel, 2009).

Whole Foods (@WholeFoods)

Figure A.8. Screenshot of Whole Foods Market Twitter account (@WholeFoods)

Whole Foods Market is a grocery retailer of natural and organic products. The brand values:

- Selling the highest quality natural and organic products available
- Satisfying and delighting our customers
- Supporting team member happiness and excellence
- Creating wealth through profits and growth
- Caring about our communities and our environment
- Creating ongoing win-win partnerships with our suppliers

- Promoting the health of our stakeholders through healthy eating education (The Whole Foods, n.d.)

The brand has been very active on social media and has accounts on Twitter, Facebook, Flickr and a corporate blog.

Zappos (@zappos)

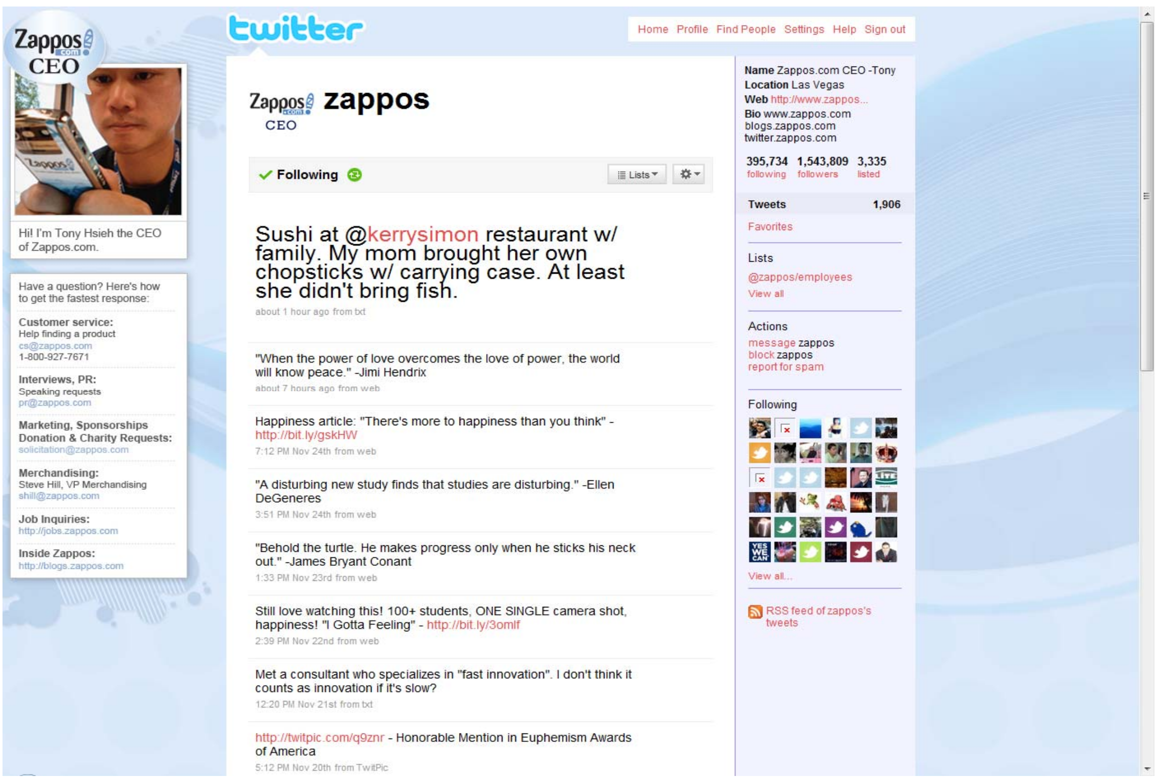


Figure A.9. Screenshot of Zappos Twitter account (@zappos)

Zappos is an online retailer specializing in selling shoes and it is famous for its excellent customer service. The company employs a loyalty business model and relationship marketing strategy (Wikipedia, n.d.-c). The brand heavily emphasizes customer service. Its CEO Tony Hsieh said, “I are a service company that just happens to sell shoes.” (Israel, 2009) The brand aims to “deliver WOM through service” (Zappos, n.d.), which is the primary source of the company’s rapid growth (Wikipedia, n.d.-c). The brand is active on social media sites and uses them to demonstrate their passion about customer service (Israel, 2009). The company also encourages its employees to adopt social media and generate organic WOMs.

APPENDIX B. F VALUE DISTRIBUTIONS FOR BNANOVA TEST

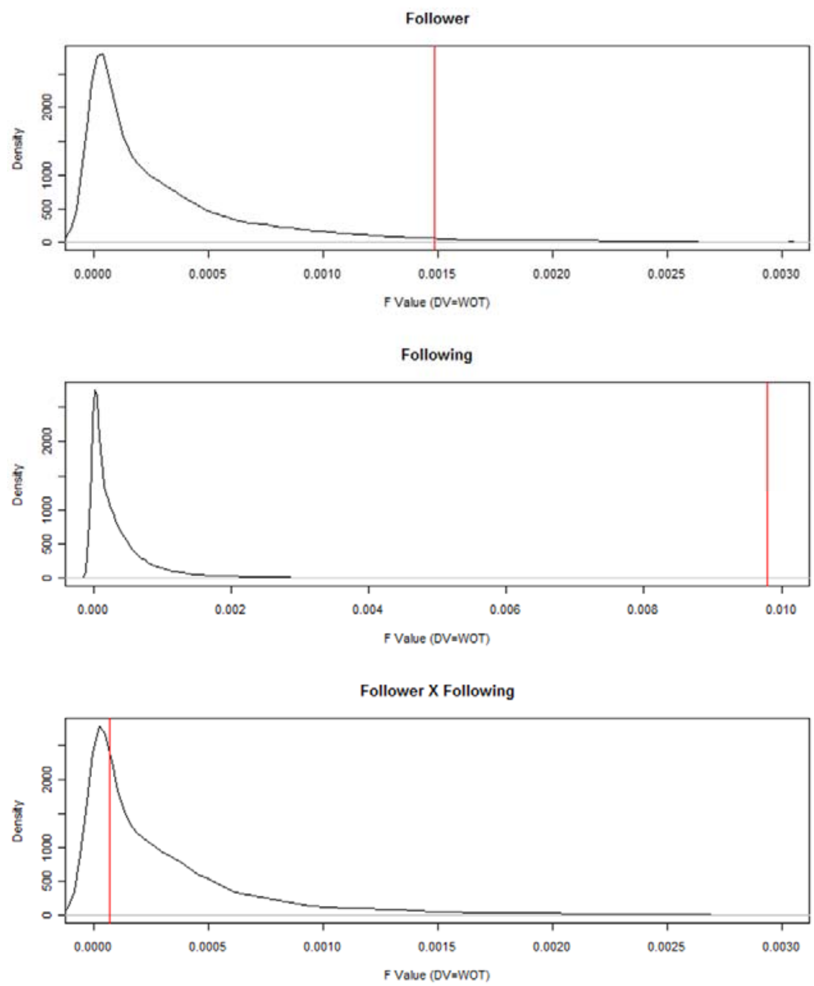


Figure B.1. F value distribution for BNANOVA test on WOT[1]

[1] The vertical red line denotes the observed F value.

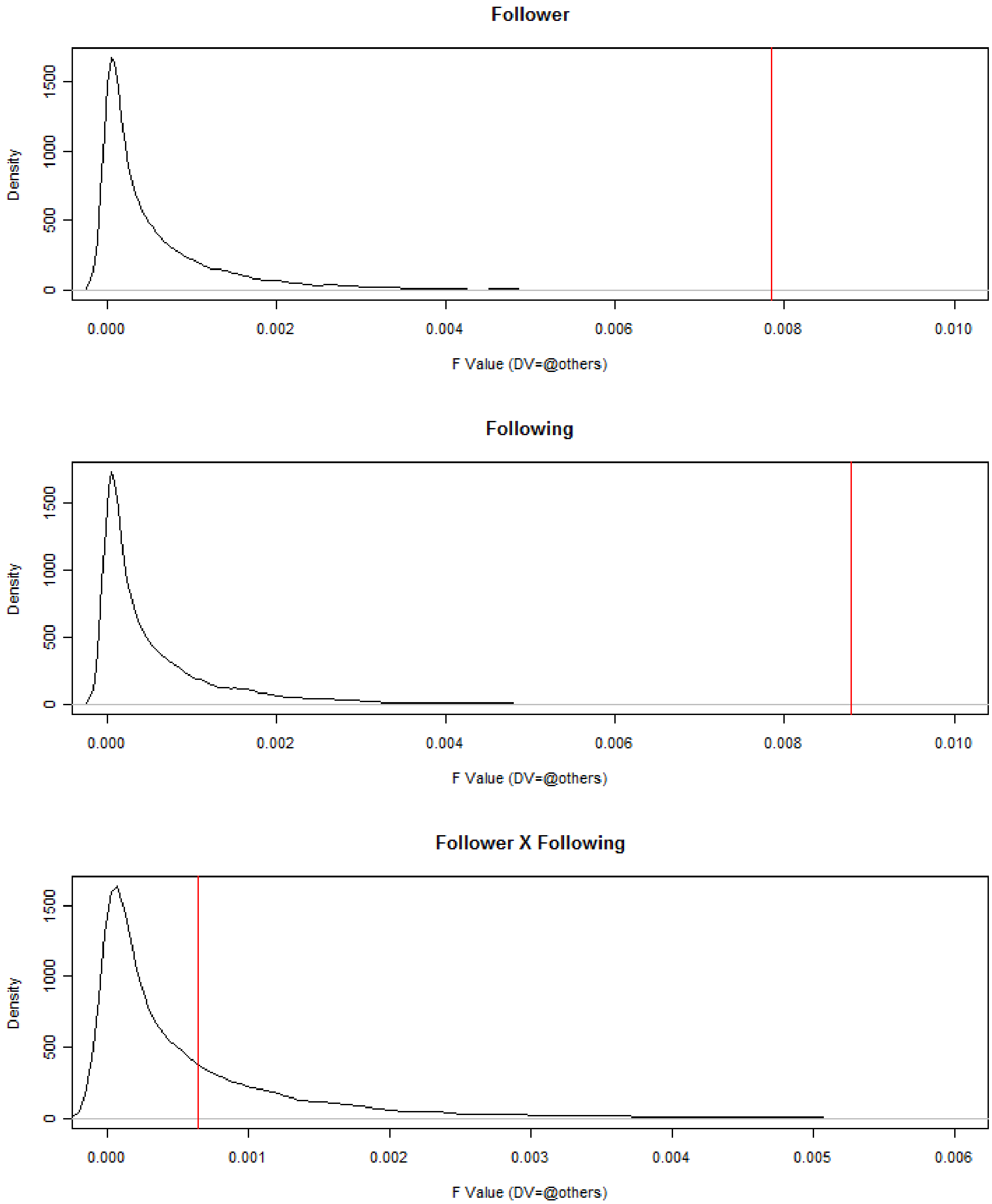


Figure B.2. F value distribution for BNANOVA test on @others[2]

[2] The vertical red line denotes the observed F value.

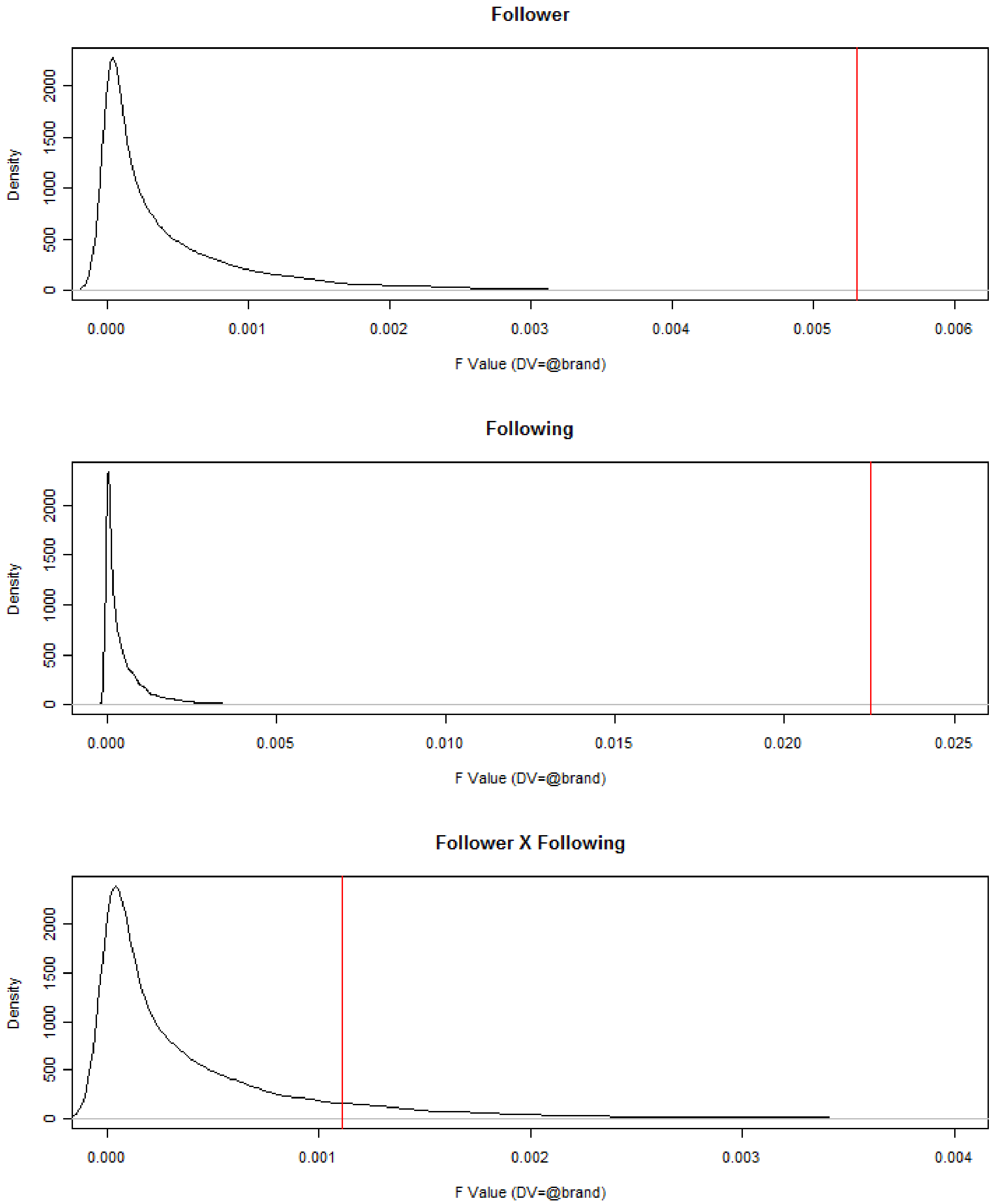


Figure B.3. F value distribution for BNANOVA test on @brand[3]

[3] The vertical red line denotes the observed F value.

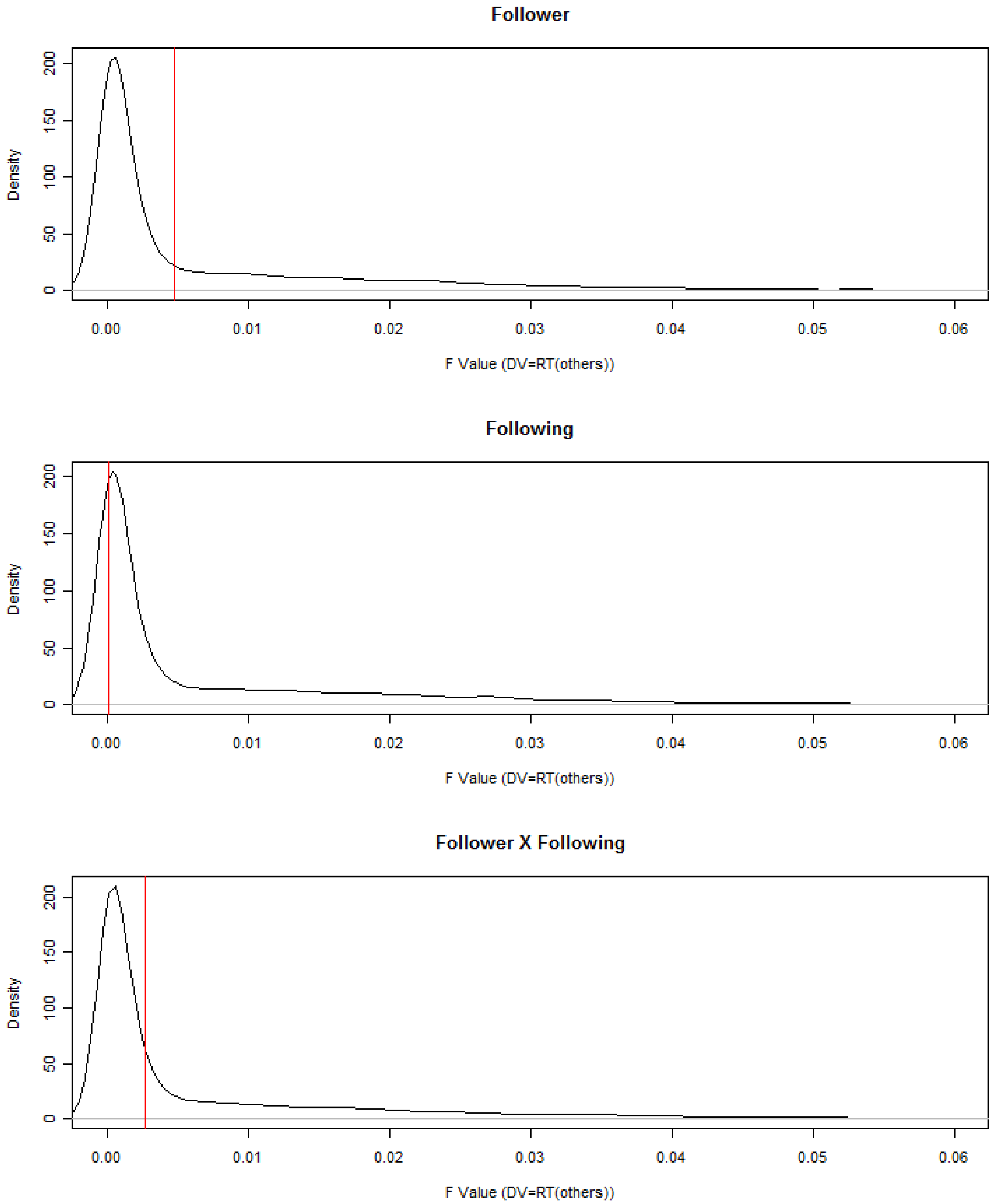


Figure B.4. F value distribution for BNANOVA test on RT(others) [4]

[4] The vertical red line denotes the observed F value.

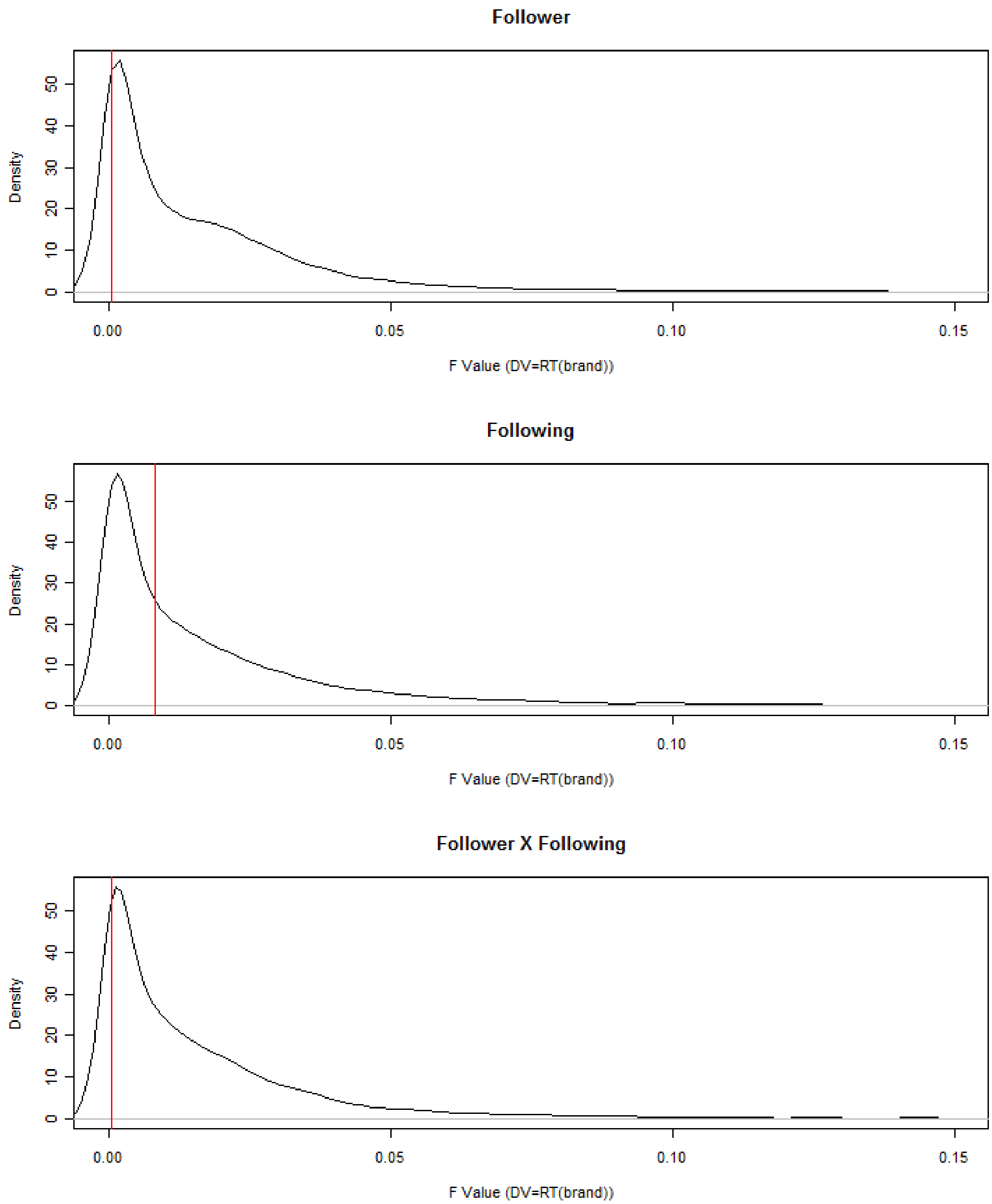


Figure B.5. F value distribution for BNANOVA test on RT(brand) [5]

[5] The vertical red line denotes the observed F value.

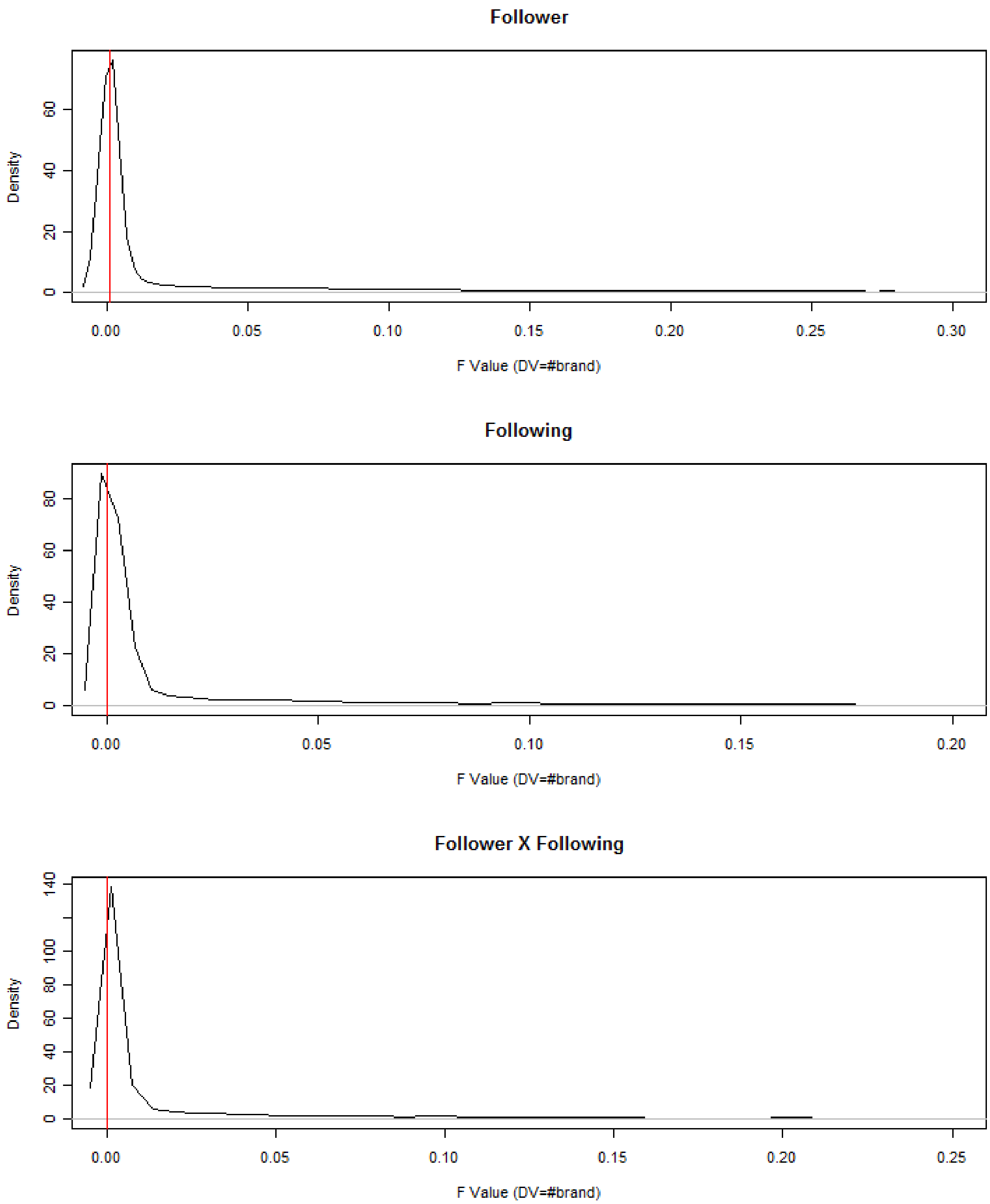


Figure B.6. F value distribution for BNANOVA test on #brand[6]

[6] The vertical red line denotes the observed F value.

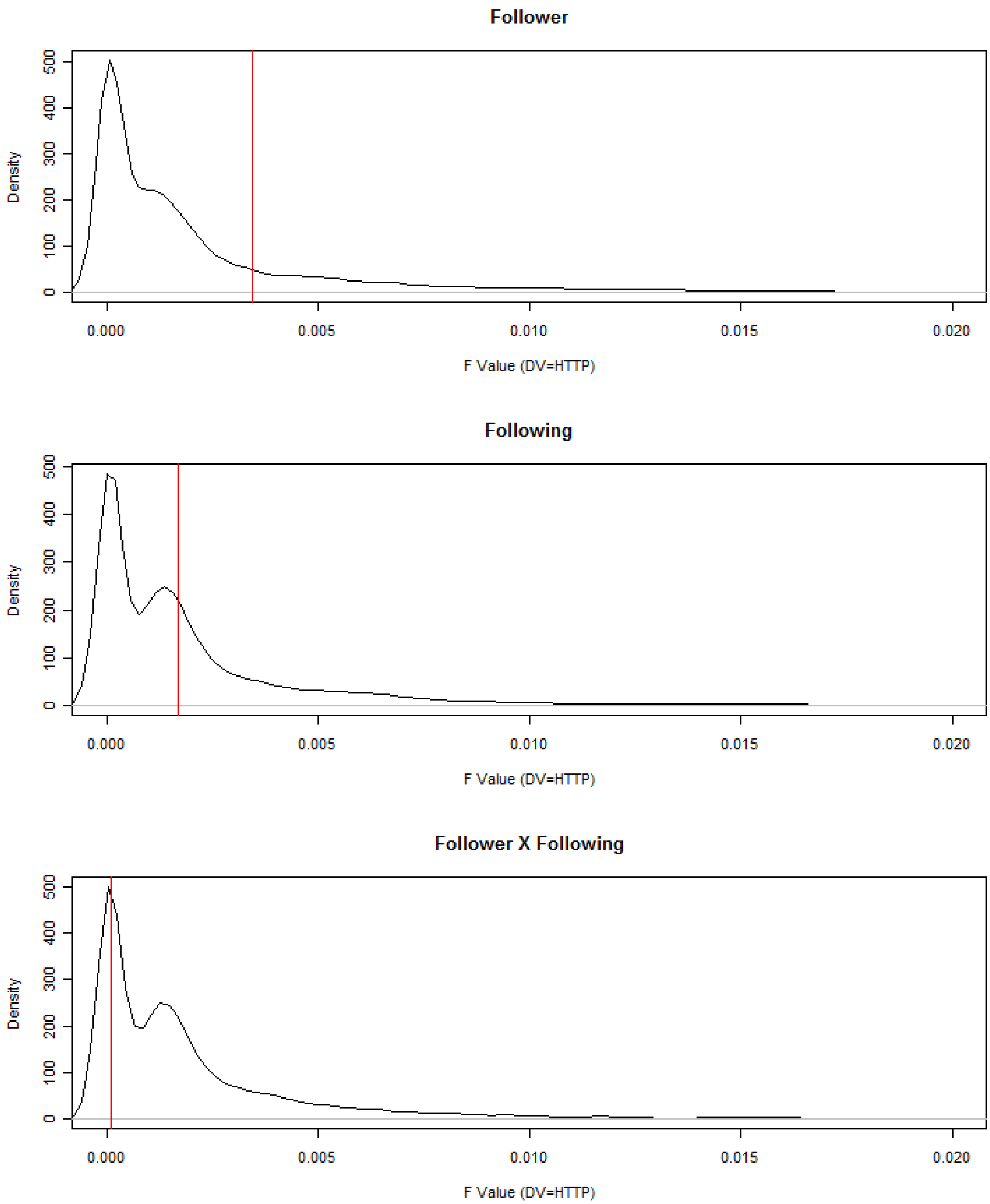


Figure B.7. F value distribution for BNANOVA test on HTTP[7]

[7] The vertical red line denotes the observed F value.

Vita

Mimi Zhang

Mimi Zhang is Ph.D. candidate in the College of Information Sciences and Technology with minor in Statistics at the Pennsylvania State University. Her primary research interest is human information interaction, especially on applying relevant theories from multiple disciplines to gain a better understanding of human information behavior. She has focused on the theoretical and empirical aspects of how people interact with information in both search and social media domains. Her broad research interests include information retrieval, social media, human-computer interaction, computer-mediated communication, economics, and experiment design. She has been involved in a number of projects, ranging from qualitative to quantitative studies, surveys to interviews, and lab studies to transaction log analyses. She specializes in multiple-treatment experiment design and analysis. Mimi has authored or coauthored more than 20 publications in the area of information technology and systems, with articles appearing in a multi-disciplinary range of edited book, journals, and conferences. Her dissertation topic is to understand information dissemination in the branding and marketing landscape. She is studying how commercial companies employ social media to manage and market their brands as well as to communicate with their customers. Mimi received the Bachelor of Management degree in Information Management and System from the School of Management, Hefei University of Technology, China, in 2003, and the Master of Science degree in the Information Knowledge Management from the Department of Information Science, Loughborough University, UK, in 2004.