Abstract

CHEN, YIZHUO. The Role of Social CRM in Brand Marketing: A Perspective of Consumers’ eWOM. (Under the Direction of Dr. Yingjiao Xu).

Companies used to fully control the relationship between brands and their customers. Nowadays, the control of the relationship has shifted to customers who have the power to influence each other using User Generated Content (UGC) and Electronic Word of Mouth (eWOM). Therefore, incorporating customer relationship management via social media into existing brand marketing framework is of strategic importance. By integrating social media marketing, Customer Relationship Management (CRM) and Social CRM, the first objective of this study is to propose a strategic brand marketing model. The integration of these marketing fields could help companies to systematically implement brand marketing campaign from various channels.

Meanwhile, the relationship management based on social influence would also help companies to efficiently leverage the voice of consumers and promptly respond to the voice from consumers. Thus, beyond the conceptual scope, the second objective for this study is to empirically test some important relationships in the proposed model in terms of identifying an individual’s social influence and evaluating the effectiveness of customized brand marketing communication, which are key components for social CRM implementation. They are 1) the relationship between individuals’ social media behaviors and the influence of their UGC; 2) the relationship between the content of UGC and the influence of UGC, and 3) the effectiveness of customized brand marketing communication on affecting individuals’ attitudes toward the brand. Social media data from an online purchasing environment (Amazon.com) and a non-purchasing environment (Runnersworld.com) were collected using web scraping technique. The data was analyzed using text mining, principal component
analysis, log-linear models, and sentiment analysis. The result suggests a significant relationship between individuals’ social media behaviors and the overall influence of their UGC in both purchasing and non-purchasing social media contexts. The result also suggests that brand names and product attributes shown in the content of UGC in non-purchasing social media context could significantly affect the content co-creation. And content co-creation significantly correlates with influence of the UGC. The findings also suggest a significant influence of customized brand marketing communication on reducing individuals’ negative attitude toward the brand.

In summary, this research expands the scope of brand marketing by integrating social CRM into brand marketing research framework. Managerially, this study could provide great implications to the brand marketing practitioners in their effort of understanding and serving their target consumers. The findings of this study would also provide useful suggestions for designing brand loyalty program at the age of social media.
The Role of Social CRM in Brand Marketing: A Perspective of Consumers’ eWOM

by
Yizhuo Chen

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Textile Technology Management

Raleigh, North Carolina
2015

APPROVED BY:

_______________________________  ________________________________
Dr. Yingjiao Xu                        Dr. Lori Rothenberg
Chari of Advisory Committee

_______________________________  ________________________________
Dr. Marguerite Moore                    Dr. Peter Bloomfield
Biography

Yizhuo Chen was born in Zhengzhou, Henan, China in 1986. He received the bachelor’s degree in Marketing from Zhengzhou University, China. After that, he attended the University of Texas at Austin, and he received the degree of Master of Science in Textiles and Apparel Technology. From there, Yizhuo headed straight to Raleigh, North Carolina, where he participated in the Ph.D. program of Textiles Technology Management at the College of Textiles at North Carolina State University.
# Table of Contents

List of Tables........................................................................................................................................................................ vi
List of Figures ........................................................................................................................................................................... vii

Chapter 1 Introduction ................................................................................................................................................................. 1
  1.0 Introduction ........................................................................................................................................................................ 1
  1.1 Statement of the Problem ...................................................................................................................................................... 2
  1.2 Research Objectives ............................................................................................................................................................ 8
  1.3 Research Approach ............................................................................................................................................................. 11

Chapter 2 Literature Review and Conceptual Model Development ......................................................................................... 12
  2.0 Introduction ........................................................................................................................................................................ 12
  2.1 Brand and Brand Equity ..................................................................................................................................................... 13
    2.1.1 Defining Brand Equity from Customers’ Perspective .......................................................................................... 13
    2.1.2 The Management of Brand Equity ......................................................................................................................... 15
    2.2 Social Media and Electronic Word of Mouth ............................................................................................................... 17
      2.2.1 The Domain of Social Media ................................................................................................................................. 17
      2.2.2 Electronic Word-of-Mouth .................................................................................................................................. 21
  2.3 Customer Relationship Management and Social CRM .................................................................................................... 24
    2.3.1 Customer Relationship Management .................................................................................................................... 24
    2.3.2 The Development of Social CRM .......................................................................................................................... 26
  2.4 Development of a Conceptual Strategic Model ................................................................................................................ 28
    2.4.1 Orientation of Business Goal in Brand Marketing .................................................................................................. 29
    2.4.2 Marketing Strategies on Different Relationship Building Channels ........................................................................... 31
    2.4.4 The Behavioral Characteristics of Social Value ....................................................................................................... 37
    2.4.5 Conceptual Strategic Model ....................................................................................................................................... 41
  2.5 Objectives of Empirical Studies and Hypotheses ............................................................................................................... 44
    2.5.1 The Relationship between Individuals’ Social media Behaviors and the influence of their UGC ................................................................................................................................. 45
    2.5.2 The Relationship between Content of UGC and influence of the UGC ....................................................................... 47
    2.5.3 The Relationship between Customized Brand Marketing Communication and Individual’s Attitude toward the Brand ............................................................................................................... 49
Appendix A - Example Data Analysis Flow Chart in SAS Enterprise Miner .......................... 147
Appendix B - Examples of Brand Position Map Obtained from the Data............................. 148
Appendix C- Python Code Example: Web Scraping Data ..................................................... 151
List of Tables

Table 1. Descriptive Statistics of Variables in the Amazon.com Dataset .............................................. 70
Table 2. Descriptive Statistics of Variables in Runnersworld.com Dataset ............................................. 72
Table 3. Summary of Running Shoes Brand Name Keywords ................................................................. 75
Table 4. Summary of Running Shoes Product Characteristic Keywords .................................................. 76
Table 5. Results of Log-linear Models for H1.1 .......................................................................................... 91
Table 6. Results of Log-linear Models for H1.2 ........................................................................................ 94
Table 7. Results of Log-linear Models for H2.1 ....................................................................................... 97
Table 8. Summary of Results of Log-linear Models for H2.2 ................................................................. 103
Table 9. Summary of Results of Log-linear Models for H2.2 ............................................................... 105
Table 10. Results of Log-linear Models using Brand Name Clusters ................................................... 108
Table 11. Results of Log-linear Models using Product Characteristic Clusters ..................................... 111
Table 12. Paired T-test & Wilcoxon Signed Rank Test for Paired Difference in Average Sentiment ................................................................. 115
Table 13. Paired T-test & Wilcoxon Signed Rank Test for Paired Difference in the Proportion of Non-Negative Sentences ........................................................................................................ 118
List of Figures

Figure 1 Conceptual Strategic Model for Brand Marketing ............................................. 44
Figure 2. Graphical Illustration of H 1.1-1.2 ................................................................. 46
Figure 3. Graphical Illustration of H 2.1-2.3 ................................................................. 48
Figure 4. Graphical Illustration of H 3 ........................................................................... 50
Figure 5. Overview of Research Methodology .............................................................. 53
Figure 6. Web scraping and text classification process example ...................................... 65
Figure 7. Data Manipulation and Analysis Procedures for H 3 ........................................ 68
Figure 8. Histograms of Total Number of “helpful” Votes in Amazon.com Dataset .... 71
Figure 9. Histograms of the Total Number of Views in Runnersworld.com Dataset .... 73
Figure 10. Distribution of the Number of Postings by Member ID ................................. 77
Figure 11. Histogram of Members’ Total Number of Unique Brand Names Mentioned .................................................................................................................................................................................. 78
Figure 12. Average Daily Posting Amount by Hours ...................................................... 79
Figure 13. Average Weekly Posting Amount by Days ..................................................... 80
Figure 14. UGC Creation in the Company-Created Posting Threads ............................... 81
Figure 15. UGC Creation for Each ID Participated in the Company-Created Posting
Threads ......................................................................................................................... 82
Figure 16. Regression Lines and Residual Plots for H 1.1 (Main dataset) ................. 92
Figure 17. Comparison of Model Performance for H 1.1 ............................................ 93
Figure 18. Regression Lines and Residual Plots for H 1.2 (Main dataset) ................... 95
Figure 19. Comparison of Model Performance for H 1.2 ............................................ 96
Figure 20. Regression Lines and Residual Plots for H 2.1 ............................................ 99
Figure 21. Joint Density Estimation and Heat Maps ..................................................... 100
Figure 22. Comparison of Model Performance for H 2.1 ............................................ 101
Figure 23. Result of Variable Clustering for Brand Names ........................................ 107
Figure 24. Comparison of Model Performance: Brand Name Clusters ....................... 109
Figure 25. Result of Variable Clustering for Product Characteristics ......................... 110
Figure 26. Comparison of Model Performance: Product Characteristic Clusters ....... 111
Figure 27. Distributions of Paired Difference in Average Sentiment (After-Before)... 113
Figure 28. Each Participant’s Change in Average Sentiment ...................................... 114
Figure 29. Distributions of Paired Difference in Proportion of Non-Negative Sentences (After-Before) ........................................................................................................... 116
Figure 30. Each Participant’s Change in Proportion of Non-Negative Sentences....... 117
1.0 Introduction

As a distinguished name or symbol intended to identify products or services of one seller from others, brand plays a significant role for both buyers and sellers in the marketplace (Farquhar, 1989). While branded products make the selection easier for consumers, companies receive much benefit from the added value endowed by the brand to their products (Farquhar, 1989). When facing market rivals, well established brands could be a key factor for a company to avoid both price competition that can seriously damage profitability and substitute offerings that can lure customers away (Porter, 2008). Nowadays, however, due to the abundant market offerings, consumers can easily find a similar alternative at a cheaper price for many consumer products such as electronics, apparel, and shoes (The Economist, 2014). Some markets could be so competitive that even the strongest brands have to strive for winning more customers. For example, even for the leading companies in the clothing retail market in the US, each only accounts for a 2 percent to 5 percent share according to a recent market report (Euromonitor International, 2014). For those highly competitive product categories, manufactures not only need to establish their brands but also have to be able to manage their brands successfully.

The field of brand management has been developed for many years with solid theoretical foundations from the fields of consumer behavior and marketing. However, the technological advancement in recent years has been reshaping people’s life in many aspects. Internet and mobile technologies, for example, largely increased information availability and consumers’
technological aptitude, dramatically changing how people find products they want and help each other in making purchasing decisions (D.-H. Park, Lee, & Han, 2007). Meanwhile, on the other hand, technology also drives the evolution of business thinking and strategy. Nowadays, companies are moving toward information-intensive environments by digitalizing business practices and customer knowledge. For example, business intelligence, database marketing and other database-driven approaches are widely adopted in many industries in recent years (Kumar & Reinartz, 2012). Those changes could bring both opportunities and challenges to the brand marketing field. A critical question is how the new concepts, such as social media, could be integrated into brand marketing strategy. The present research concentrates on addressing this critical strategic question.

1.1 Statement of the Problem

In brand management science, one of the most important metrics is brand equity which is defined as the total value of a brand as a separable asset, a measure of the strength of consumers’ attachment to a brand, and a description of the associations and beliefs the consumer has about the brand (Feldwick, 1996). As a classical view of brand management, Wood (2000) postulates that brands should be managed as valuable, long-term corporate assets, thus the growth of brand equity is the major objective of marketing activities. Conversely, another view of brand management is that companies should focus on growing customer equity rather than brand equity as the value of brands depends on customers (Rust, Zeithaml, & Lemon, 2004). With the rise of consumer centric view in marketing fields, the notion of brand equity evolved from a brand-centered and holistic concept to an individual customer based concept called Customer Based Brand Equity or CBBE (Rust, Lemon, &
Zeithaml, 2004; Christodoulides & De Chernatony, 2010). CBBE is defined by Keller (1993) as a memory based identity that is associated with brand identity, meaning, response and relationships, representing the value of a brand that leads to marketing effort uniquely attributable. While traditionally defined brand equity sees the value of brands as the total asset primarily for financial purposes, CBBE is more useful for brand managers by specifically characterizing the psychological antecedents of customer’s loyal behavior such as brand judgments and feelings (Keller, 1993).

One leading force of the competitiveness of a market is savvy customers since they can force down prices by playing the supplier and it’s rivals against one another (Porter, 2008). It has been largely recognized that the source of consumers’ brand knowledge includes both marketer-generated contents (e.g. commercial information from TV or Internet) and user generated contents, such as Word-of-Mouth (WOM) from other consumers (Chevalier & Mayzlin, 2003). In the past, traditional mass media channels such as TV and magazines were largely used for marketing campaigns (Joachimsthaler & Aaker, 1996). Recently, as the number of social media adopters has been increasing exponentially, the role of socially connected customers in the market place became more important. Popular social networking sites (SNS) such as Facebook and Twitter greatly boosted the use of social media among younger generations. For instance, by 2010, there were more than 500 million active users on Facebook and more than 10 billion messages sent through Twitter since its launch in 2006 (Baird & Parasnis, 2011b). In the social media age, interactive online platforms could offer more powerful alternatives for consumer marketing. Yet, the potential of social media in marketing can’t be fully tapped without realizing the social value of customers’ online social
interaction. Such social value of customers is broadly referred to as the monetary consequences of social interactions or WOM with other customers (Libai, Muller, & Peres, 2013).

The innovation of social media has built a virtual connection between internet users. Mobile technology, on the other hand, enables seamless and fast social network connection regardless of location and availability of computers. As consumers’ interpersonal communications have been greatly changed by social media, their WOM behaviors have also evolved from face-to-face talks among friends to social interactions in online communities. With the widely available e-commerce and online shopping websites, customers can purchase almost any brands and products conveniently through the internet. To facilitate the purchase decision, many internet based applications such as social blogs, online shopping sites and virtual communities, allow users to generate and exchange reviews and product related contents (Kaplan & Haenlein, 2010). In general, User Generated Contents (UGC) are described as the various forms of media contents that are publicly available and created by end-users rather than organizations (Daugherty, Eastin, & Bright, 2008). Research shows that the UGC and electronic Word-of-Mouth (eWOM) have similar effect as WOM in terms of providing explicit product information and tailored solutions. And the perceived credibility of eWOM is much higher than marker-induced communications (Hung & Li, 2007). Due to the importance of integrating social functions in online shopping context, online social network and social media have become the new battle ground for brand owners and marketers.
For companies, predicting customer behaviors is getting more difficult in the social media age as customers intertwine together and affect each other’s buying choice over the Internet. Moreover, with the intensive market competitions and sophisticated information technologies, loyal customers become more demanding and difficult to maintain (Kumar & Reinartz, 2012). As a result of more diversified demand of consumers, strategic solutions have been proposed by marketing scholars from customers’ point of view (Kotler, 1967). For example, direct marketing, a process of identifying the most likely buyers of certain products and promoting the products specifically for their needs, has been widely used in retailing industry (Ling & Li, 1998). Recently, the consideration of cost efficiency of marketing activities such as advertisements and promotions has been added to the business strategy (Kumar & Reinartz, 2012). Namely, to maximize marketing returns with limited cost, marketing programs should focus on serving a group of clearly defined target customers who are loyal and profitable (Kumar, Pozza, Petersen, & Shah, 2009).

Regarding the business question of how to manage the long-term customer relationship, marketing theorists and practitioners have began moving away from product centric view to customer centric view since 1990s (Kumar & Reinartz, 2012). However, the adoption of customer value perceptive into marketing practices with available customer behavioral and transactional data has just begun few years ago. Traditionally, satisfied customers are believed to have higher likelihood to buy products from the same brand and create more profit for the company (Armstrong & Kotler 2008). However, some argue that satisfaction level and attitudinal loyalty doesn’t always help the company to increase profitability (Reichheld, 1996; Kumar, Pozza, Petersen, & Shah, 2009). Even if customers display
behavioral loyalty to the brand by repeated purchase, they may not worth heavy marketing investments if they keep purchasing deep discounted products and contributes less on profit margin. For this reason, Kumar, Pozza, Petersen, and Shah (2009) proposed the reversed logic theory, advocating a discriminated strategy that marketing spending should be allocated on only loyal customers with high expected future monetary value. The value measurement is called Customer Lifetime Value (CLV) defined as the expected total future discounted value of a customer (Hennig-Thurau, Gwinner, & Gremler, 2002). The marketing strategy of developing the appropriate relationships with key customers to maximize CLV utilizing customer data is defined as Customer Relationship Management (CRM) (Payne & Frow, 2005). Research suggests that CRM initiatives can effectively increase company’s Return-On-Investment (ROI) (Ryals, 2005). In the past years, the usefulness of CRM in effectively enhancing marketing performance has been proved by business practitioners from different industries (Boulding, Staelin, Ehret, & Johnston, 2005). Jayachandran, Sharma, Kaufman, and Raman (2005) reported several business-to-business (B2B) and business-to-consumer (B2C) companies with significant growth of unit profit after implementation of CRM strategy.

It used to be that companies can fully control the relationship between brands and their customers so CRM strategy works well at traditional channels such as call centers and brick and mortar locations (Baird & Parasnis, 2011a). However, in the social media era, the relationship between companies and customers has extended to the online social domain. Therefore, control of the relationship has shifted to the virtually connected customers who have the power to influence each other through social networks (Baird & Parasnis, 2011a).
From the point of view of innovation diffusion, the adoption rate of users who have large influence over others largely affects the success of spreading out new ideas in a social system (Rogers, 1995). Hence, a customer’s social influence can play an important role in brand marketing (Dwyer, 2011). Similar to CRM, a business strategy of engaging customers through social platforms with a goal of building trust and brand loyalty is referred to as Social CRM (Malthouse, Haenlein, Skiera, Wege, & Zhang, 2013).

While CRM helps companies to appropriately make marketing decisions based on the knowledge obtained from their customers purchasing behaviors, consumers’ information consumption behaviors acquired in social CRM, on the other hand, gives brand managers the opportunity to effectively manage the diffusion of brand knowledge among customers. By combining CRM strategy and social value perspective together, brand performance could be better improved with a well maintained relationship with customers that are both profitable and socially influential. In addition to the many existing studies on CRM, social CRM research is emerging. Some of them focus on solving traditional CRM problems with measurements of social media behaviors (Neslin, Gupta, Kamakura, Lu, & Mason, 2006; Nitzan & Libai, 2011). Others went beyond the traditional scope and discuss the social value in addition to ordinary CLV (V. Kumar, Petersen, & Leone, 2010; Schmitt, Skiera, & Van den Bulte, 2011). Nevertheless, up to date, there is no existing research framework concentrating on the role of social CRM for brand marketing strategy in social media environment. Consequently, this study will further explore this research area.
1.2 Research Objectives

As technological innovation of internet moves forward and consumer lifestyles keep evolving, by integrating the lasted development in social media, CRM and Social CRM into one strategic model, these seemingly isolated marketing research fields may be able to help companies to better respond to the changes in the market. Therefore, the first objective for this study is to propose a conceptual strategic model for brand marketing that integrates relationship building and management with Social CRM approach. To be useful for business practitioners, the conceptual strategic model will focus on brand marketing in terms of both the relationship building and relationship management processes for different marketing channels.

Beyond the conceptual scope, the second objective for this study is to empirically test several important relationships proposed in the conceptual strategic model regarding the implementation of social CRM. First of all, the usefulness of social media in marketing relies on the monetary consequences of eWOM behavior (e.g. Hung & Li, 2007). Social media behavior involves a variety of activities, ranging from consuming content to participating in discussions, sharing knowledge with other consumers, and contributing to other consumers' activities (Heinonen, 2011). The influence of social media behaviors on consumers’ purchase decisions has been largely measured by attitudinal measurements (Wang, Yu, & Wei, 2012). But little evidence about the correlation between an individual’s social media behavior and one’s impact on other consumers, such as influencing purchase decision and brand knowledge, was shown from behavioral observations. Therefore, using behavioral observation method, this study aims to examine the relationship between individuals’ social
media behaviors and the influence of their UGC on other consumers in purchasing social media settings.

In addition, individuals participating in social media or social platforms could be driven by different motivations. Among others, concern for other members and advice seeking are the most common reasons regardless of level of social media engagement (Hennig-Thurau et al. 2004). Even if the social interaction happens in non-purchasing setting, a members’ social media behavior could still be critical as brand related information may be spread out to potential buyers when they read the relevant social media content. From the point of view of innovation diffusion, the contribution of a member in non-purchasing social media setting to a brand could be well indicated by his or her influence on the process of brand knowledge diffusion among others. Therefore, this study also attempts to examine the relationship between individual’s social media behavior and the influence of their UGC on other consumers in non-purchasing social media settings.

Next, to implement marketing communication on social media, brand managers need to understand what features indicated by customers regarding a product category is useful in growing brand value. A significant amount of such information could be embedded in the content of UGC. Empirical studies show that consumers are often loyal to several brands in a product category (Jacoby, 1971). So brand managers need to understand the relationship between individuals’ brand preferences/product preferences and their social influences. Socially influential individuals could be targeted based on these preferences so that marketing strategy can be effectively tailored and customized. Therefore, this study
examines the relationship between choices of brand names and product characteristics mentioned in a UGC and the influence of the UGC on other individuals.

Lastly, for any CRM strategies, the relationship between the individual and the firm takes the form of an interactive dialog, primarily through customized brand marketing activities (Kumar & Reinartz, 2012). Previous research also suggests that consumers’ established attitude toward a brand has a large impact on how they respond to marketing activities (Czellar, 2003). Namely, customers are more likely to facilitate the diffusion process of a brand in a positive manner if their positive attitudes toward the brand can be improved by the marketing actives. Therefore, this study also explores whether an individual’s attitude toward a brand would change after participating in customized brand marketing communication initiated by the brand on social media.

In summary, this research is conducted with two major objectives. The first objective is to propose a conceptual strategic model for brand marketing that integrates relationship building and management with Social CRM approach. The second objective is to conduct empirical studies to test several important relationships proposed in the conceptual strategic model in terms of identifying an individual’s social influence and evaluating the effectiveness of customized brand marketing communication, which are key components for social CRM implementation.
1.3 Research Approach

To develop the strategic model, a qualitative review of literature on related research fields was performed. To justify the relationships introduced above in the proposed model, several empirical examinations of relationship were conducted using quantitative methods. Each of the examination was conducted in a carefully designed procedure including four major sections: defining research hypotheses, measurement of constructs, raw data collection and transformation, and statistical analysis. Worth mentioning, unlike other consumer behavior research on this topic, psychological measurements are not used in this study. Since the proposed business model focuses on brand management in online digital environment, the constructs were quantified by numerical and textual data collected from two online social platforms. In business practice, manually recording the large amount of online data could be extremely time-consuming. Given the difficulty of acquiring large amount of behavioral data from social media web pages, the data was captured by using a web scraping algorithm written in Python programming language. Python is a multi-purpose programming language that has efficient high-level data structures and effective approach to object-oriented programming. It is often considered an ideal language for scripting and rapid application development in many areas on most platforms (Python software foundation, 2014). To analyze individual behavior and attitude, the raw data was decomposed into sentences and summary statistics were then aggregated by user ID. Information embedded in the reviews and postings were extracted by text mining approach.
Chapter 2 Literature Review and Conceptual Model Development

2.0 Introduction

In this chapter, relevant literatures are reviewed to provide a theoretical foundation for the development of the proposed conceptual business model. The major research topics in marketing and management sciences discussed in the chapter include brand equity, social media, Electronic Word of Mouth (eWOM), Customer Relationship Management (CRM) and Social Customer Relationship Management (Social CRM).

In the first section, as the cornerstone of brand management, the concept of brand equity is introduced and major theoretical perspectives regarding brand equity management are reviewed. Next, in the second section, the development of social media is briefly discussed and research findings regarding social media behaviors are summarized. In the third section, the review of literature focuses on marketing research from the customers’ perspective, namely, relationship marketing and CRM. Moreover, social CRM is reviewed as an extension of CRM strategy implemented in social media environment, building connections between the concepts introduced previously. Finally, a brand management model integrating social CRM strategy is conceptually developed based on the important findings from the literatures in the corresponding marketing fields. The conceptual strategic model as well as objectives of the empirical studies is presented in the remaining sections of this chapter,
2.1 Brand and Brand Equity

2.1.1 Defining Brand Equity from Customers’ Perspective

In marketing, brands often provide the primary points of differentiation between competitive offerings, and as such they can be critical to the success of companies (Wood, 2000). Originally, the idea of brand came from marks on cattle that used to ensure they were not stolen in the wild west of the U.S (Kapferer, 2012). The tradition definition of brand, therefore, is a distinguished name or symbol intended to identify products or services of one seller from other competitors (Aaker, 2009). Besides the indication of product quality, there are other meanings indicated by a brand. For example, some scholars defined brand as a product that add other dimensions to differentiate it in some way from other products designed to satisfy the same need (Keller, Heckler, & Houston, 1998). Later, the brand was characterized in terms of mental associations that add value to existing benefits created by the product (Kapferer, 2012). While brand is a symbol that helps consumers to differentiate products, brand equity is usually characterized as the outcomes result from the marketing of a product or service because of its brand name that would not occur if the same product or service did not have that name (Keller, 1993). Due to the many different views of brand equity, no single definition of brand equity is universally accepted (Christodoulides & De Chernatony, 2010). As Farquhar (1989) argued, the only consensus about brand equity is that it refers to the added value to the product endowed by the brand. Keller (1993) suggests these two general motivations for studying brand equity: a financially based motivation to estimate the value of a brand for accounting purposes and a strategy-based motivation to improve marketing productivity.
Many believe the brand equity could be measured financially. The financial perspective (e.g. Simon & Sullivan, 1993) primarily concerns financial performance of the brand. However, even though brand equity has been considered as a repository of future profits or cash flows that results from past marketing investment, branding strategy based on the sheer economic view of brand equity can lead to long-term failure (Ambler, 2003). In recent years, as customer centric marketing approaches received increasing attention, fundamentally product centered concept such as traditional brand equity has been challenged by customer centric concepts (Rust, Lemon, & Zeithaml, 2004). The customer centric view point can be found in early literatures in marketing. For example, a shift of marketing theory toward a customer’s perspective has been proposed by Kotler in 1967. However, the emphasis on long term value rather than short term transactions was not fully established in theory and practiced until recently (Rust, Lemon, et al., 2004). There are several important concepts that drive the development toward a customer centered marketing theory. For example, Bolton (1998) examined the link between customer satisfaction and retention by modeling the duration of customer relationship. His findings show the importance of adopting customer value measurement to improve company performance.

Researchers with strategy-based motivations tend to define brand equity from individual customer’s point of view. Because market response to a brand also depends on individual customer’s mental process regarding a brand, it is necessary to consider the psychological response of individuals for a given brand (Wood, 2000). Thus, customer based perspective of brand equity has been proposed with regard to the antecedent and consequence of memory structure change (Ambler, 2003; Keller et al., 1998). Customer Based Brand Equity, or
CBBE, is defined as the differential effect of brand knowledge on consumer response to the marketing of the brand (Keller, 1993). Therefore, the memory structure is referred to as brand knowledge that will alter the consumer response. Brand knowledge is defined in terms of two dimensions. One is brand awareness defined as brand recall and recognition performance by consumers. The other one is brand image which refers to the set of associations linked to the brand that consumers hold in memory (Keller, 1993). The majority of early studies on CBBE are conceptual, followed by mostly empirical studies. They mainly agreed that brand awareness and brand image (associations) are important CBBE components (Christodoulides & De Chernatony, 2010). In addition to brand knowledge, major component of CBBE includes perceived quality, brand loyalty, and other proprietary brand assets (Aaker, 2009).

2.1.2 The Management of Brand Equity

The management of a brand functions as a strategic way to manage the brand equity (Wood, 2000). Therefore, the first step of brand management is to specifically identify the quantifiable measurements of brand equity. In the relevant literatures, brand equity has been suggested to be measured by direct and indirect ways. Direct approaches attempt to measure the brand equity directly from the change of consumers’ preferences or utilities and indirect approaches measure brand equity through intermediates (Christodoulides & De Chernatony, 2010). Direct method has been discussed in many early studies from the perspective of behavioral outcomes (Srinivasan, 1979; Park and Srinivasan, 1994; Jourdan, 2002, Leuthesser, Kohli, and Harich, 1995). Measurements of brand equity in term of value was suggested in a later study (Christodoulides & De Chernatony, 2010; Swait, Erdem, Louviere,
The major components in this measurement include tangible and intangible brand value that can be directly attributed to product (Kamakura & Russell, 1993). For example, Ailawadi, Lehmann, & Neslin (2003) proposed that brand equity can be evaluated by revenue premium a brand generated compared with what a private label can generate. However, most direct approaches, such as price premium, can’t provide insight into the customer based source of brand equity, which can be critical for diagnostic of any equity change (Ailawadi et al., 2003).

According to Keller (1993), conceptualizing and measuring brand equity from the customers’ point of view are helpful for marketers for two reasons. One is the various effects of marketing activities on brand knowledge. The other one is the effect of short term marketing effort on long-term brand success. These understandings can facilitate marketing programs to build a successful brand. Therefore, instead of measuring consumer preference or utilities, indirect approaches evaluate mainly evaluate brand equity through its sub-dimensions in consumers’ minds (Christodoulides & De Chernatony, 2010), such as brand awareness, brand association, perceived quality and brand loyalty (Aaker, 1991; Keller, 1993; Yoo, Jeon & Park, 2011).

After brand equity is quantified by appropriate measurements, more importantly, companies need to identify important factors in marketing that can affect their brand equities. The Brand Value Chain proposed by (Kevin Lane Keller & Lehmann, 2003) provides a holistic and integrated approach to understand the value created by brands. In this model, the value creation, including customer mindset, brand performance, and shareholder value, begins with
the marketing activities and marketing program investment. Several elements in marketing mix such as product, communication, trade and employees are suggested to be important factors for brand value creation. Effective marketing communication is critical in the diffusion process of brand knowledge as suggested by innovation diffusion theory (Rogers, 1962). Nowadays, due to the information technology, the scope of marketing channels has been largely extended (Kumar & Reinartz, 2012). Therefore, communication strategy with customers through various channels would be one important part in brand management and marketing. In addition to the above factors in brand equity management, creating community and fans of the brand through relationship building becomes more important for brand building (Kapferer, 2012). Therefore, not only acquiring new customers but also retaining current customers through marketing channels is critical (Jain & Singh, 2002). For companies, growing brand equity is one of the key objectives that can be achieved through gaining more favorable associations and feelings from target consumers (Christodoulides & De Chernatony, 2010). Nowadays, as more people stay connected with friends using social media, companies can build interactive relationship with customers.

2.2 Social Media and Electronic Word of Mouth

2.2.1 The Domain of Social Media

The internet not only allows users to view the contents but also provides a platform for them to create contents and share with others. User Generated Contents (UGC) are described as the various forms of media contents that are publicly available and created by end-users rather than organizations (Daugherty et al., 2008). This definition highlights that UGC are created
by individual users, regardless of the online platforms the contents are posted on. Any online platforms where UGC can be posted and exchanged are called social media. Formally, social media is defined as Internet-based platforms built on the ideological and technological foundations of Web 2.0, that allow the creation and exchange of UGC (Kaplan & Haenlein, 2010). Sharing in social media represents the extent to which users exchange, distribute, and receive content (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011). In addition to sharing contents, other important features of social media include identification, conversation, presence and relationship (Kietzmann et al., 2011). In terms of user behaviors, many social media research concentrates on factors driving the use of social media. Many scholars have explored the mechanism of social media behaviors such as posting stories and sharing contents (Hargittai & Walejko, 2008; Cranshaw et.al., 2010). For instance, Lewis et.al.(2008) suggested that people sharing social relationships as well as demographic traits tend to share a significant number of cultural preferences. Similarly, Bakshy, Karrer and Adamic (2009) reported that social network adoption rates increase as more friends adopt the network, indicating that sharing among close friends occurs more rapidly than sharing among strangers. In addition, they also found that some users play a more active role in sharing contents than others.

The domain of social media includes a wide range of web based applications. In development of social media, there are several important milestones: Classmates.com (1995), SixDegrees.com (1997), Friendster (2002), MySpace, Bebo, and Facebook (2004), Flickr (2004) and Youtube (2005) (Giudice, Peruta, & Carayannis, 2014). While most people would agree that
Wikipedia, Youtube, and Facebook are social media, there are many other types of applications that can be viewed as social media according to the definition. Generally speaking, social media include collaborative projects, virtual social world, content communities, virtual game world and blogs (Kaplan & Haenlein, 2010). In addition to the above categories, microblogs, social bookmarking, product reviews, discussion boards were also considered as social media in recent studies (Hoffman & Fodor, 2010). Among others, social networking sites (SNS) is a web-based services that allow individuals to construct a public or semi-public profile, articulate a list of other users with whom they share a connection, and view or traverse their list of connections and those made by others within the system (Ellison, 2007). Unarguably, this type of social media is more popular than others (Choi, 2006; boyd & Ellison, 2008). Contents shared by SNS is more trustworthy in general than those from other sources, however, privacy rises as an important issue when more public information and shared contents are stored online (Chow & Chan, 2008; Gross & Acquisti, 2005; Fogel & Nehmad, 2009). Because SNS is primarily based on private friendship rather than product interest, marketing activities using SNS is limited from the perspective of users’ privacy concerns.

Another popular category of social media is virtual communities. Unlike SNS, virtual community refers to a network of people with common interests who use electronic means to communicate and share interests, valuable resources, experiences and knowledge (Kardaras, Karakostas, & Papathanassiou, 2003). Despite being widely considered as UGC platforms, virtual communities originally came from the idea of spontaneously collaborating and sharing resources when others requesting them (Rheingold, 1993). Therefore, a distinctive
feature of virtual communities is the high level of social interaction. However, the social interactions on virtual communities are usually not in the form of one-to-one conversations among friends. Instead, users post messages or photos in order to create as many discussions as possible within the community. Therefore, unlike comment box on a shopping website or Youtube page where everyone can post their ideas, the participants in the virtual community can obtain knowledge and develop relationships with others who share similar interests (Bickart & Schindler, 2001). Although building personal relationship on virtual communities is not as prevalent as on SNS, users on virtual communities tend to be more homogeneous with respect to certain topics. As a result, the perceived creditability of virtual community is even higher for consumers with strong common interest (Ridings, Gefen, & Arinze, 2002).

Overall, among business practitioners, it is not a mystery that social media can help their business. Instead, social media becomes a common marketing platform in many industries. However, marketers are still struggling with how to effectively measure the performance of social media because existing performance measurements have not been widely accepted for the interactive media environment (Hoffman & Fodor, 2010). For instance, according to a recent survey conducted in 2009, about one third of the 1700 executives worldwide reported that their social media applications had yet to provide measurable benefits either when used internally or with customers or business partners (McKinsey, 2009). Although discrepancy exists, it has been well recognized business activities that can potentially supported by social media including branding, sales, customer service, and product development (Culnan, McHugh, & Zubillaga, 2010). For example, small companies can potentially use SNS to find new customers, build online communities to maintain good relationship with their customers,
and investigate customers’ demographic information (Pattison, 2009). Many studies show that brand marketing on social media can not only drive traffic and eWOM but also increase customer loyalty and retention (e.g. Culnan et al., 2010; Smith, Fischer, & Yongjian, 2012).

### 2.2.2 Electronic Word-of-Mouth

Consumers’ word of mouth (WOM) is a person-to-person communication between a receiver and a communicator, which is perceived as non-commercial message regarding a brand, product, or service (Arndt, 1967). It has been largely discussed in terms of the motives, social structure, and the subsequent effect on consumer behaviors (Godes & Mayzlin, 2004). In general, research shows that consumers who have purchased an item are willing to offer their experiences for many psychosocial and social motivations, such as helping the community, express negative feelings, or advice seeking (Hennig-Thurau et al., 2004; Sundaram and Webster, 1999). According to Dichter (1966), positive WOM communication is motivated by product involvement, self involvement, other involvement and message involvement. In addition to these motivations, altruism, helping the company, anxiety reduction, vengeance, and advice seeking were proposed to influence both positive and negative WOM (Sundaram and Webster, 1999). Unarguably, one of the most important explanations for the high influence of WOM on consumers purchasing behaviors is the credibility. In both academia and business fields, the recognition of WOM as one of the most crucial forces affecting purchase decisions and the leading indicator of a product’s success has been well established in early studies (Brooks, 1957; Dichter, 1966).
The online counterpart, electronic word-of-mouth (eWOM) shows even more significant impact and relevancy to customers (Bickart & Schindler, 2001). While WOM can only take place in conversations, eWOM is stored electronically. It is supported by many platforms such as web-based opinion platforms, discussion forums, boycott Web sites, and news groups (Hennig-Thurau et al. 2004). Unlike offline WOM, eWOM has a much larger scope of influence on other consumers because of the large number of connections facilitated by the internet. Motivations for eWOM behavior are similar to those for offline WOM. Hennig-Thurau et al.( 2004) suggest that consumers who post eWOM are motivated by eight factors: platform assistance, venting negative feelings, concern for other consumers, self-enhancement, social benefit, economic intensives, helping company, and advice seeking. Based on these factors, the customers can be segmented into: self interested helper, multiple motive consumers, consumer advocates, and true altruists. Many marketers believe that SNSs represent an ideal platform for eWOM, as consumers freely create and disseminate brand-related information (Chu & Kim, 2011). However, virtual community could be a better option where users can both sharing eWOM and discuss interactively with other consumers with similar interests. Virtual communities have been largely studied from the perspective of social behavior (Armstrong & Hagel, 2000). However, limited exploration on product-related communication and the implication for marketing were found (Hennig-Thurau et al.2004).

Research on motivations for participating in virtual communities shows that people join groups in the online environment of virtual communities for both feelings of affiliation and belonging as well as for information and aid in goal achievement (Ridings & Gefen, 2004).
Therefore, common interests and information exchange are two important features. And the value of virtual communities from business’s perspective mainly comes from highly involved users and their eWOM effect since many virtual communities are built around common interests such as sports, product, or social activities (Kardaras et al., 2003). Rather than private network, this type of social media is considered as public space for experience sharing regarding a specific topic. Virtual communities can be classified as commercial or non-commercial communities, depending on whether pursuing commercial interest (Leimeister, Sidiras, & Krcmar, 2004). Although both of them are product centric, participating members are more active in non-commercial communities than in those operated by entities with commercial purposes (Leimeister et al., 2004).

There are two major challenges for integrating virtual communities into marketing strategy. On one hand, with the large amount of discussions among existing users and potential buyers for specific product categories, the valuable market information is rich on virtual community. However, the challenge is how to extract the themes of the information shared among the members and translate them into actionable strategies for companies. On the other hand, managing the influence of eWOM effectively can also be challenging given the complexity and dynamicity of the community environment. The important role of social interaction in virtual communities was first noticed by Balasubramanian and Mahajan (2001). They propose that virtual communities presents an excellent opportunity to explore the economic and social aspects of consumer decisions, but the successful economic leverage of virtual communities must achieve a harmonious interplay between the social and economic motivations. Similarly, as pointed out by Kozinets et al. (2010), eWOM marketing requires
more careful considerations of the social behaviors and communication patterns among consumer networks. For example, any virtual community intervention can be largely influenced by how well the messages comply with the ecosystem of the community. In recent years, the perception of how to integrate eWOM into business model has evolved from organic inter-consumer influence to the network coproduction (Kozinets, De Valck, Wojnicki, & Wilner, 2010). Nevertheless, the notion of the importance of eWOM will need to be transformed into marketing strategies aiming at establishing interaction with key consumers and managing eWOM activities effectively.

2.3 Customer Relationship Management and Social CRM

2.3.1 Customer Relationship Management

Traditional marketing theories consider market as a whole or segment the market into some sub-populations (Smith, 1956). The reason behind this approach is the cost efficiency in mass production. However, many limitations have been shown in recent years. As information technology and the flexible manufacturing practices are becoming mature, market offerings are vastly increased and consumers can easily obtain market information in almost every product category (Kumar & Reinartz, 2012). As a result, customers can easily switch to another provider. One important task for marketers is to define and manage the relationship with customers to maintain long-term profitability. Therefore, meeting individual customer’s need with customized brand marketing strategy has become an important competitive advantage for businesses.
As a derivative strategy from relationship marketing, Customer Relationship Management, or CRM, was developed to manage such one-to-one marketing activities. As defined by Payne and Frow (2005), CRM is a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. CRM brings value to business through enhanced opportunities to use data and information to both understand customers and co-create value with them. CRM is the outcome of the continuing evolution and integration of marketing ideas and newly available data, technologies, and organizational forms (Boulding et al., 2005). Because CRM is closely associated with technology improvement and also integrated with business strategies, the definition and description used by business practitioners vary considerably (Boulding et al., 2005). According to a study on CRM among different companies, the concept can be referred to as narrow as a particular technology solution or as broad as a holistic approach to manage customer relationships (Payne & Frow, 2005). In this study, the CRM is mainly considered as a strategic framework rather than just building the technical infrastructures.

In order to manage the relationship with individual customers, each customer is differentiated by their Customer Lifetime Value, or CLV, which could be calculated as a function of the predicted contribution margin, the propensity for a customer to continue in a relationship, and the marketing resources allocated to the customer (Venkatesan & Kumar, 2004). Accordingly, there are four CRM initiatives for companies to manage the relationship with customers: customer acquisition, customer retention, customer churn, and customer win-back (Kumar & Petersen, 2012). Customer acquisition is the process of acquiring new customers
and customer retention is the process of keeping and developing relationships with the customers after the company acquires. They are the foundation step of the whole CRM process (Kumar & Petersen, 2012). Therefore, customer acquisition and customer retention are more emphasized in this study. These CRM initiatives will largely depend on the analytical system in CRM, which provides knowledge about customers’ behaviors that can help targeting key customer groups and implementing strategic marketing activities. In other words, the data is first recorded in marketing channels and then organized and stored along with other internal data in the company’s database. After the data has been processed and analyzed, the resulted knowledge or model is then applied into marketing programs to enhance marketing campaign performance (Kumar & Reinartz, 2012). After years of implementation among business practitioners, CRM has been proved in many industries to effectively enhance marketing performance (e.g. Boulding et al., 2005; Jayachandran et al., 2005; Ryal, 2005).

### 2.3.2 The Development of Social CRM

Despite the many advantages, many have discussed the weakness of utilizing CLV as the only value metrics in CRM. For example, because CRM differentiates customers based on their total value contribution to maximize total profit, fairness needs to be carefully considered or it will cause many trust issues and even customer loss (Boulding et al., 2005). Furthermore, when customers feel being treated unfairly, their negative WOM will seriously damage brand image (Gruen, Osmonbekov, & Czaplewski, 2006). Because of the widely used social media, customers are more powerful by quickly responding to marketing campaigns and affecting other customers with influential eWOM. For companies, the
traditional model of managing the customer relationship is insufficient in the era of social media and the business strategy needs to adapt to the reality that the customer is now in control (Baird & Parasnis, 2011b). As social media emerged into a major communication platform in recent years, many companies have started incorporating social media as an interactive marketing tool. Therefore, Social CRM strategy has been proposed and is defined as a business strategy of engaging customers through social media with a goal of building trust and brand loyalty (Woodcock, Green, & Starkey, 2011). Unlike other marketing channels such as retailing channels, social media create value for both companies and customers in terms of informational exchange rather than transactions. Recently, Social CRM is one of the newest and fastest growing field in CRM (Kumar & Reinartz, 2012).

A thorough understanding of why consumers are attracted and how they influence other consumers’ attitude and behavior is important for any form of social media marketing (Thorsten Hennig-Thurau et al., 2010). However, the business practitioners’ understanding about the important social media functional features may be perceived differently by consumers. For example, some of the most important reasons ranked by consumers, namely, discount, purchase and product rankings, were in fact the least important reasons ranked by executives (Baird & Parasnis, 2011b). Such perception gap will have a large impact on the effect of social CRM campaign and should be checked by companies regularly. The technical cornerstone of Social CRM is the capability of continually capturing and analyzing customer data from social interactions to help reduce risk and improve all aspects of the customer relationship (Baird & Parasnis, 2011a). Hence, Social CRM could also be built around existing CRM concepts and technologies such as CRM systems, availability of interfaces and
services, and data analysis and mining methods (Reinhold & Alt, 2011). But, in addition to technological components in CRM, the basis for Social CRM strategy also includes customers’ social media engagement and social value instead of solely relying on the prediction of customers’ future economic value, namely, the CLV (Woodcock et al., 2011).

The study of Social CRM is still an emerging field. Some of the existing literatures on this topic focus on the concept of social value. Social value is defined as the economic value generated due to social interactions and WOM (Malthouse et al., 2013). Research shows that social value could play a major role in marketing activities such as referral program (Kumar, Petersen, & Leone, 2010). In practice, referred customers tend to have higher contribution margin and higher retention rate (Schmitt, Skiera, & Van den Bulte, 2011). Because of the importance of social value, other studies attempt to incorporate the social value into CRM (e.g. Nitzan & Libai, 2011; Delen & Crossland, 2008). This study will keep exploring the topic of integrating social value into marketing strategic model but extend it to brand management field by attempting to study the role of managing consumers’ social media behavior in growing brand equity.

2.4 Development of a Conceptual Strategic Model

One of the main objectives of this research is to propose a conceptual strategic model for brand marketing that integrates relationship building and management with Social CRM approach. Therefore, the purpose of the following sections is to address the conceptual relationships of the constructs shown in the literature reviewed and built a strategic model by integrating marketing and relationship management in social media context. To develop the
conceptual model, three important components of the model are suggested and discussed. First, as a brand marketing strategic model, the first component is the orientation of business goal. To be specific, should product centric or customer centric business goal be established? Next, relationship in marketing is characterized as the interactive dialogs between customers and the firm (V. Kumar & Reinartz, 2012; Fournier & Yao, 1997). The nature of such relationship would vary for different marketing channels. For example, customer relationship built on media channels is brand centric while the relationship on retailing channels is transactional and value based (Kumar & Reinartz, 2012). Therefore, different marketing communication channels would also need to be taken into account in the brand marketing strategy. So, the second component need to be addressed is the marketing strategies on different channels that could be used to build customer relationship for brand marketing. Lastly, because of the interactive communication and large amount of connections in social media, the management of customer relationship on social media could be better addressed from CRM perspective. Thus, the third critical component is to identify the behavioral characteristics of social value that can be used for targeting key consumers in CRM. These three components are specifically addressed in details in the following sections.

2.4.1 Orientation of Business Goal in Brand Marketing

As discussed before, brand equity is traditionally described as a separable asset of the total value of a brand, a measure of the strength of consumers' attachment to a brand, and a set of the associations and beliefs the consumer has about the brand (Aaker, 1995; Feldwick, 1996; Rust, Zeithaml, & Lemon, 2004)). Since the brand equity can be valued by either the financial outcome or behavioral outcome induced by a brand, the management of brand
equity has been seen differently (Wood, 2000). Brand equity can also be shown in many other forms and evaluated by different measurements. (Leone et al., 2006; Ailawadi, Lehmann, & Neslin, 2003). Nevertheless, the generic view of brand equity serves mainly for financial purposes rather than help the management of brand building (Keller, 2003). Alternatively, the value of a brand can be viewed from a customer’s point of view. Namely, the life time value of a consumer of a brand was introduced with revenue and cost considerations related to the effect of consumer marketing (Leone et al., 2006). Rust, Zeithaml, and Lemon (2004) defined customer equity as the discounted lifetime values (e.g. Customer Lifetime Value or CLV) of a firm’s customer base. Both brand equity and customer equity emphasize the importance of customer loyalty to a brand, and they have the same notion that value is created by having as many customers as possible and creating profit as much as possible. (Ambler, 2002). However, brands primarily serve the role of attracting customers. Conversely, customers can eventually help brands to monetize their brand value (Keiningham, Aksoy, Perkins-Munn, & Vavra, 2005).

The individual customer based brand equity, CBBE, would be more useful for brand management by integrating both perspectives (Rust, Lemon, & Zeithaml, 2004; Christodoulides & De Chernatony, 2010). Brand knowledge is about the thoughts, feelings, perceptions, images, and experiences that linked to the brand in the minds of customers, rather than the facts about the brand (Leone et al., 2006). So CBBE can characterize the psychological antecedents of customer’s loyal behavior such as brand judgments and feelings. As a psychological measurement, CBBE could be affected by both the marketing communications to the target customers and the influence of these customers’ word of mouth.
behaviors on the brand knowledge diffusion process. In traditional media channels (e.g. TV and radio), such diffusion is difficult to observe so the influence of each customer is not differentiable. Yet, in interactive online channels, customers’ online conversations are observable and could be used for understanding the influence of a customer at each customer level. Consequently, the advantage of CBBE comes from its compatibility with both generic and individual perspectives.

The business goal in this model not only needs to reconcile the measurement of brand equity but also needs to be managerially useful. Therefore, regarding the establishment of business goal, this model proposes that brand marketing strategy needs to be oriented toward customers rather than products. From the above discussion, the growth of CBBE would be the business goal for the brand marketing strategic model.

2.4.2 Marketing Strategies on Different Relationship Building Channels

Relationship in marketing is characterized as the interactive dialogs between customers and the firm (Kumar & Reinartz, 2012; Fournier & Yao, 1997). To open such dialogs and build relationship with customers, a brand relies on various marketing strategies and channels. The first type of strategy is mass marketing on one-way communication channels such as traditional media. In the past, advertising on traditional media has been the cornerstone of most brand building efforts and most companies rely on traditional media advertising as their primary brand-building device (Joachimsthaler & Aaker, 1996). Traditionally, mass marketing strategies are effective due to the fact that customers used to have limited media choices. For instance, if the audience’s involvement level with the TV program is relatively
high, commercials could effectively promote products (Park & McClung, 1986). Since consumers have more information sources from the internet and social media, it’s more challenging for brand managers to solely rely on mass marketing through traditional media such as TV, radio and magazines. Nowadays, other media types allow customers to view contents from various sources without interruption of commercials. For example, in 2009, 74 percent of the U.S. adults and 93 percent of the teenagers use internet and 47 percent of U.S. adults and 73 percent teenagers use social media and SNS (Lenhart et al., 2010). Meanwhile, the use of traditional media such as newspapers and television news dropped significantly since 2005 (Althaus & Tewksbury, 2000). Given the many choices, only contents that are relevant and interesting would be viewed and retained. Thus, the traditional model of brand management needs to adapt to the reality that the customer is now in control (Baird & Parasnis, 2011a). However, traditional media channels are still largely used for marketing campaigns since there are some circumstances for which traditional media advertising is more effective. For example, it is suggested that traditional media is one of the most popular channels for companies to initiate the relationship with customers because of advertising effectiveness (Breuer, Brettel, & Engelen, 2011). By advertising, brand knowledge is transferred to customers through various media channels while customers demonstrate certain level of psychological and/or behavioral responses. The advertising strategy has been shown to have long-term effect on consumer brand choice (Mela, Gupta, & Lehmann, 1997). To maintain customer brand loyalty and profitability, companies usually play a dominant role in actively introducing their brands to customers. As opposed to immediate adoption, the knowledge of brand is diffused by WOM process over time (Quirmbach, 1986). But due to
the non-interactive nature of traditional media, brand marketing is slow in response because companies disseminate brand knowledge to a large group of customers without immediate feedbacks from customers and follow-up marketing activities to manage the diffusion process.

As the use of internet and social media became prevalent, interactive online platforms could become a more powerful alternative for marketing. This new type of marketing communication channel could connect with customers individually and interactively (Hennig-Thurau et al., 2004). Namely, interactive communications are highly customizable for each individual customer and the messages can be sent between companies and customers. Therefore, the second type of brand marketing strategy involves both direct marketing and interactive marketing on two-way communication channels. Direct marketing is a relational marketing process of prospecting, conversion, and maintenance that involves information feedback and control at the individual level by using direct response advertising with tracking codes (Bauer & Miglautsch, 1992). According to Bauer and Miglautsch (1992), channels used for direct marketing is mainly customizable marketing communications platforms (e.g. e-catalogs or emails). On the other hand, interactive marketing is defined as the immediately iterative process by which customer needs and desires are uncovered, met, modified, and satisfied by the providing firm (Bezjian-Avery, Calder, & Iacobucci, 1998). The advantage of interactive marketing relies on the fact that the iterations occur over some duration, allowing companies to build databases that provide subsequent purchase opportunities tailored to the consumer (Blattberg & Deighton, 1991). However, to successfully implement marketing campaign over interactive social media, the strategy used
for traditional media channels need to be carefully updated according to consumers’ behavioral characteristics in the interactive social platforms.

In the context of interactive marketing and direct marketing, the relationship between companies and customers was mostly interpreted from the perspective of customer lifetime value, or the expected total future discounted value of a customer (Hennig-Thurau, Gwinner, & Gremler, 2002). Thus, CRM is mainly applied based on such one-to-one monetary relationship. When making CRM strategy, the resulting value measurement will be used to determine the most valuable customer segments and identify effective approach to promote brand knowledge to those customers, helping companies to maximize customer values while decrease marketing cost (Kumar & Reinartz, 2012).

Compared with costly and inefficient relationship building through tradition media channels, the key advantage of CRM comes from the efficiency in managing the monetary relationship with customers, because the allocation of marketing resources to a customer is directly associated with the expected future profit from this customer. However, the customer value defined in CLV is criticized as an oversimplified measure (Leone et al., 2006). Beyond the monetary value, customers’ social value drives much attention in recent years (Malthouse et al., 2013). It used to be that companies can fully control the relationship between brands and their customers. As a result, CRM strategy works well at traditional channels such as call centers and brick and mortar locations (Baird & Parasnis, 2011a). Despite the importance of WOM has been well recognized, explicit use of WOM in marketing strategy is less commonly seen due to the difficulties in observing WOM in offline shopping context. With
the widely available e-commerce and online shopping websites, customers can purchase almost any brands and products conveniently through the internet. To facilitate the purchase decision, many internet based applications such as social blogs, online shopping sites and virtual communities, allow users to generate and exchange reviews or product related contents (Kaplan & Haenlein, 2010). Research shows that the eWOM has similar effect as WOM in terms of providing explicit product information and tailored solutions. As eWOM can be recorded electronically and stored in a large scale, the relationship with customers not only extended to the online social domain but also become more manageable given the rich eWOM behavioral data.

Two-way communication channels, such as social media, can be either monetary or non-monetary depending on the type of online social platforms (e.g. Leimeister et al., 2004). For instance, product review pages on many online shopping websites allow consumers to directly show their opinions about whether a product is worth buying to other consumers. Different from online shopping website, users of online social communities (e.g. product virtual community) mainly discuss product related topics to help other consumers to make future shopping decisions (T. Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004). With the exponentially increased number of social media adopters, the important role of socially connected customers in the market place is greatly highlighted. Mobile technology, on the other hand, enables seamless and fast social network connection regardless of location and availability of computers. The widely used information technologies at customer’s end and company’s end drove the development of modern relationship marketing under the large framework of computer-mediated marketing management system environments (e.g. firm-
consumer interactions, consumer-firm interactions and consumer-consumer interactions) (Yadav & Pavlou, 2014).

Despite less technical challenges associated with marketing activities and communication encountered by companies, in terms of the challenges of brand marketing using social media, the leading issue associated with implementing marketer initiated social media marketing is the privacy concern from customers (Fogel & Nehmad, 2009). Nevertheless, receiving messages from other customers is less problematic. Studies have shown that opinions from other consumers is one of the most important factors that affect the potential buyer’s purchase decisions (Brooks, 1957; Dichter, 1966). Therefore, the potential of social media can’t be fully tapped without realizing the business value of a customer’s influence on other customers’ buying behaviors or eWOM behaviors. Such social value is broadly referred to as the monetary consequences of social interactions or WOM with other customers (Libai et al., 2013). As a result of the speed of internet and the broadness of online social network, the social value of customers using social media may be leveraged to a degree that is much larger than that of offline consumers.

In summary, brand marketing strategies, namely mass marketing, direct marketing and interactive marketing are discussed according to two types of communication channels: traditional one way communication channels and two-way communication channels. Mass marketing strategy could be sufficient for traditional one way communication channels where brand knowledge diffusion is difficult to observe. Yet, for two-way communication channels, direct marketing and interactive strategies with CRM approach could be more efficient for
facilitating the brand knowledge diffusion among potential consumers. Given the more consumption of social media and increasing importance of eWOM, brand managers need to extend their marketing campaign to social media domain. Moreover, it is critical to incorporate marketing strategy and marketing communication channels with a holistic strategic framework.

2.4.4 The Behavioral Characteristics of Social Value

Marketers have recognized that the perceived credibility of WOM is much higher than marker-induced communications (Hung & Li, 2007). In addition, while marketer-generated contents (e.g. commercial information from TV or Internet) can affect consumers’ brand knowledge, user-generated contents (e.g. WOM and customer reviews), especially those posted by consumers with larger social influence, shows even larger influence on other consumers (Chevalier & Mayzlin, 2003). For companies, predicting customer behaviors is more difficult in the social media context, because customers intertwine together and affect each other’s buying choice at a higher degree than in offline context. From a marketing point of view, a carefully targeted customer group needs to be clearly defined and well served to maintain loyalty with limited marketing resources (Kumar et al., 2009). In terms of brand management, the performance of a brand could be better improved by a well maintained relationship with customers who are both profitable and socially influential, rather than a less controlled relationship without appropriate marketing communication (Kumar & Reinartz, 2012). The rich information provided by the social media gives brand managers the
opportunity to effectively target those customers. Social value of a customer is broadly referred to as the monetary consequences of one’s social interactions or WOM with other customers (Libai et al., 2013). In brand marketing, the direct monetary consequences may not be easily assessed. However, a consumers’ impact on the core element in brand management, brand knowledge, could be evaluated by level of social media behavior and social interaction with others. So, to enhance the usefulness of the brand marketing strategic model, the social value is defined in this study in terms of the influence on knowledge diffusion with regard to brand. Moreover, such social value needs to be defined in terms of observable behavioral characteristics.

First, social media engagement level of a consumer could have a large impact on his or her influence on the value creation process. Although there has been disagreement regarding the scope of behaviors that account for social media engagement, it is a consensus that the degree of engagement varies (Malthouse et al., 2013). If the social media engagement level is low, consumers are likely to passively consume the contents on the social media and demonstrate the basic forms of feedback such as “like” a page on Facebook (Malthouse et al., 2013). If the level of social media engagement is high, consumers may actively participate in the content co-creation process and share contents with other users. Because reviews created by highly engaged users tend to have more perceived credibility and can attract more viewers, given a positive attitude toward the brand, customers with high social media engagement could generate more social value through influencing a large number of other consumers (C. Park & Lee, 2009). From another point of view, customers who frequently participate in the discussion regarding a brand are also more likely to gain brand knowledge and expertise.
These consumers could be potentially valuable because experienced content contributors can receive more favorable psychological response and more participations from other users (Bendapudi & Leone, 2003). Consequently, the higher the engagement level in social media using, the more valuable the customer tends to be.

Second, other than the level of social engagement, social status in the users’ social network can also affect the social value of the consumer. A member’s social status in a social network refers to the number of connections in his social network and his influence on these connections. These two attributes are referred to as social network size and social network tie strength (Haythornthwaite, 1996). Research show that the growth of (innovation) adoption rate is contributed most by those influential members in a social group that both spread and are consulted for information (Flynn, Goldsmith, & Eastman, 1996). Therefore, social status could be another important indicator of social value in addition to social media engagement.

In addition to targeting customers with the behavioral characteristics, brand managers aiming at implementing marketing activities on social media also need to understand the impact of the social media content regarding the brand generated through postings and reviews on product forum or online shopping website. On one hand, the contents created by customers with high social media engagement could be both informative and beneficial for companies. Online consumer reviews describe product characteristics in terms of usage situations and measure the product performance from a user’s perspective (J. Lee, Park, & Han, 2008). As end users, customer may have a better understanding of the most needed features of functionality and appealing design. For example, research has suggested that inviting
customers to participate in the product development and designing stages may increase the likelihood of the success of the product (Lovelock & Young, 1979). Therefore, it is beneficial for companies to encourage customers to actively create contents rather than be passive audiences (Wind & Rangaswamy, 2001). On the other hand, consumers are often loyal to several brands in a product category (Jacoby, 1971). Thus, consumers’ eWOM behaviors could be affected by their brand preferences. Unlike in purchasing social media setting where consumers have clear goal of purchasing, consumers in non-purchasing settings such as forums tend to be interested in learning about a product category via open discussions (Bickart & Schindler, 2001). Thus, some product characteristics and brand names are more likely to generate responses than others in non-purchasing settings. Based on these arguments, brand managers also need to understand brand preferences and product preferences of the socially influential consumers so that their marketing strategy can be effectively tailored and customized.

In summary, social engagement, social status and social media content could be useful for predicting the social value of a customer. Information used for targeting customers and knowledge used for implementing marketing activities all come from the social media database. As information regarding consumers’ social media behaviors and brand performance keep updating, the customer targeting and marketing communication strategies would be modified accordingly.
2.4.5 Conceptual Strategic Model

The Brand Value Chain (Kevin Lane Keller & Lehmann, 2003) suggests that company’s marketing investment is the foundation of brand value creation process in customer mindset, brand performance, and shareholder value through customers’ brand loyalty. Thus, as one of the most important marketing investments, the marketing communication and the long-term customer relationships on the corresponding marketing channels are the cornerstone of this model.

By definition, CBBE is the differential effect of brand knowledge on consumer response to the marketing of the brand (Keller, 1993). Brand building blocks, the sub-dimensions of CBBE, are the memory based mindset of individual customers. Therefore, since customer mindset regarding a brand is about the brand knowledge, brand value could be then well characterized by CBBE. Among the many brand building blocks, brand salience, brand imagery, and brand feelings are closely related to the marketing communications because feelings, images and perceptions could be largely influenced by marketing programs without having experience with product (Keller, 2001). So, marketing communication could directly affect a consumer’s brand knowledge. Furthermore, their understandings about the brand can also be affected by the opinions from other consumers (e.g. Hennig-Thurau et al., 2004). Namely, in addition to direct influence, the marketing communication also affect brand knowledge through brand knowledge diffusion process promoted by certain highly influential consumers, such as opinion leaders (Roger, 1995).
Consequently, marketing communication is linked to brand salience, brand imagery, and brand feelings in CBBE by either direct influence or brand knowledge diffusion process among consumers, which is based on the building of customer relationships on different marketing channels. From previous sections, customer relationship building for brand marketing involves two types of media channel. Namely, brand knowledge can be distributed to consumers by either non-interactive relationship on one-way communication channels or interactive relationship on two-way communication channels. In two way communication channels, purchasing and non-purchasing social media platforms are included.

Based on the two types of marketing communication channels, different marketing strategies are then applied. For traditional one-way communication channels, mass marketing communication could disseminate brand knowledge to the mass audience by traditional media to build undifferentiated relationship and generate WOM. Unlike diffusion of brand knowledge on one-way communication channels, diffusion of brand knowledge on two-way communication channels is observable and a consumer’s influence on the diffusion process for interactive relationship can be suggested by consumers’ social media behavioral characteristics, such as social media engagement and social status as well as their created social media contents. For brand managers, the information would serve two purposes: targeting customers to implement marketing communication and understanding key elements that can facilitate the marketing communication. With the above information, social CRM strategy is applied to determine the appropriate CRM initiatives. Then, social media marketing communication is performed on the target customers according to CRM initiatives, to boost the diffusion of brand knowledge and further grow CBBE.
For example, brand managers could decide to acquire or maintain relationships with certain customers, based on the knowledge about consumer behaviors from social media. In terms of the information flow, the behavioral data is collected from the above sources and passed to CRM unit to help making decision whether the relationship is decided to be acquired, maintained or terminated. With identifiable social value indicators in terms of behavioral characteristics such as social media engagement, social status and social media content, company can invest in building relationship with customers who are critical in growing brand equity while reducing the social media marketing cost on customers who are less beneficial. In the next process, tailored and appropriate social media marketing communications would apply according to the knowledge generated from social media content regarding the characteristics that can increase the responses rate of customers. In the long term, it is uncertain that the initially identified high social value customers may continuously respond well to the marketing activities and contribute to the brand diffusion. Therefore, their behaviors need be analyzed periodically to determine the future relationship management actions. Consequently, the healthy relationship building and management cycle could effectively help companies to update knowledge about their customers and grow brand equity in the long run.

In summary, based on the Brand Value Chain, a conceptual strategic model for brand marketing is presented in this study consisting of several major components including marketing communication, customer relationship building, brand knowledge diffusion, Social CRM and CBBE. The conceptual model is proposed and shown in Figure 1.
2.5 Objectives of Empirical Studies and Hypotheses

Beyond the conceptual scope, this study aims at providing empirical evidences to some of the important relationships proposed in the conceptual model, which consist of key assumptions for the implementation of social CRM. They are important for identifying an individual’s social influence and evaluating the effectiveness of customized brand marketing communication. These relationships were tested in purchasing settings and non-purchasing social media settings. In summary, this study empirically examined three relationships. The remaining body of this section describes the relationships to be examined and the hypotheses were developed accordingly.
2.5.1 The Relationship between Individuals’ Social media Behaviors and the influence of their UGC

The first relationship proposed in conceptual model regarding the implementation of social CRM is the relationship between individuals’ social media behaviors and the influence of their UGC. In purchasing social media contexts, the UGC author’s past social media behavior is usually recorded by the social media website and shown in his or her public profile. From the perspective of innovation diffusion, individual people adopt and spread innovations at different rates due to various considerations (Ronger, 1995). Consumers who read product reviews seeking for purchase suggestions could be largely influenced by the author’s past behaviors such as the author’s experience and his or her credibility. Thus, consumers may be more likely to adopt a UGC if the author’s profile indicates higher engagement level and higher social status (Brodie, Hollebeek, Jurić, & Ilić, 2011). Therefore, the following hypothesis is proposed:

H1.1 In purchasing social media contexts, the overall influence of an individual’s UGC is positively associated with his or her engagement level and social status indicated in the profile.

In non-purchasing social media settings, making purchase decision may not be the primary driver for individuals to participate in social media or social platforms. Instead, among others, concern for other members and advice seeking are the most commonly reasons (Hennig-Thurau et al. 2004). In a product community, members could read any content that interests them. But other members’ experiences or opinions regarding a brand or product
could largely affect the viewers’ future purchase behaviors. In this process, community members who have higher engagement level and higher social status may suggest more views from others. Similarly, a hypothesis is proposed for the non-purchasing setting as the following:

**H1.2** In non-purchasing social media contexts, the overall influence of an individual’s UGC is positively associated with his or her engagement level and social status indicated in the profile.

Accordingly, the 2 hypotheses proposed above are illustrated in Figure 2.

*Figure 2. Graphical Illustration of H 1.1-1.2*
2.5.2 The Relationship between Content of UGC and influence of the UGC

The second relationship proposed in conceptual model regarding the implementation of social CRM is the relationship between content of UGC and influence of the UGC. Intuitively, whether a consumer adopts the UGC or not could largely depends on content of the UGC. Product characteristics and brand names are two primary topics in UGC (Malthouse et al., 2013). Therefore, it is hypothesized that choice of brand names and product characteristics could affect the influence of the UGC. Because UGC from purchasing social media settings usually involves more than one product categories that are not comparable, this study only investigate this relationship in non-purchasing social media settings for a specific product category. In non-purchasing social media settings, most of the UGC is co-created. That is, one user creates the original message and others can contribute more content by replying to the original message. The amount of content co-creation would suggest the how informative and how the popular the UGC is in the community. Hence, it would be reasonable to postulate that the total views of a UGC can also be influenced by the amount of content co-creation. In summary, this empirical research objective is formulated into the following hypotheses regarding non-purchasing social media contexts:

**H2.1 In non-purchasing social media contexts, the influence of a UGC is positively associated with the level of content co-creation in the UGC.**

**H2.2 In non-purchasing social media contexts, choices of brand names and product characteristics mentioned in a UGC have a significant impact on the influence of the UGC.**
H2.3 In non-purchasing social media contexts, choices of brand names and product characteristics mentioned in a UGC have a significant impact on the level of content co-creation associated with the UGC.

Accordingly, the above 3 hypotheses are illustrated in Figure 3.

In non-purchasing social media contexts:

Figure 3. Graphical Illustration of H 2.1-2.3
2.5.3 The Relationship between Customized Brand Marketing Communication and Individual’s Attitude toward the Brand

Finally, the third relationship proposed in conceptual model regarding the implementation of social CRM is the relationship between customized brand marketing communication and consumers’ attitude toward a brand. Literature shows that companies have implemented customized brand marketing communitarian on various social media platforms as an approach of marketing intervention (Stead, Gordon, Angus, & McDermott, 2007). But the effectiveness of such approach is not conclusive. However, the social community environment in non purchasing social media settings may generate more favorable marketing effect. Thus, a well implemented customized brand marketing communitarian in non purchasing social media settings may positively affect community members’ attitude toward the brand. And future brand marketing could be largely benefited as individuals’ established attitude toward a brand has a large impact on how they respond to marketing activities (Czellar, 2003). Therefore, the following hypothesis is developed:

**H3. In non-purchasing social media context, customized brand marketing communication is associated with a positive change on individuals’ attitude toward the brand.**

Accordingly, H3 shown above is illustrated in Figure 4.
In non-purchasing social media contexts:

Figure 4. Graphical Illustration of H3
Chapter 3 Methodology

3.0 Introduction

The goal of this chapter is to describe the methodology used for the second objective for this study, including data collection procedures and statistical models used for testing the important relationships proposed in the conceptual model. This chapter is organized to address the procedures involved for testing the three proposed relationships. First, the overall research design is described briefly. Second, measurements for the variables contained in corresponding hypotheses are formulated and discussed. Next, data collection process is described in terms of data source identification, raw data collection procedures, and data preparation for analysis. Sample data is summarized and described in the next section. Finally, the data analysis procedures and methods are briefly discussed.

3.1 Research Design

In marketing research, both quantitative and qualitative research methods have been widely used. Survey and in-depth interview could be the typical instruments for these methods respectively. Despite the advantages of these methods, such as flexibility and easy to interpret, there are drawbacks for making conclusions based on psychological responses from participants. For instance, although survey data is convenient to obtain, especially from online survey, both the quality and validity of the data could be questionable due to non-response bias, sampling bias, subjects’ inability to reflect their true thinking, subjects purposefully providing wrong information, etc. (e.g. Mertens, 1998; Sale, Lohfeld, & Brazil,
2002). On the other hand, interviews are relatively more costly, time consuming, and may be subjective in interpreting the results. Additionally, interviews provide less external validity due to the limited number of participants (Godes & Mayzlin, 2004). In comparison, direct observation of behavioral responses could be not only more cost efficient but also more reliable by reducing subjects’ recall bias (Godes & Mayzlin, 2004). Therefore, observational data will be used to measure the variables of interest in this study.

In this study, variables contained in the hypotheses were defined in terms of quantifiable behavioral characteristics observed on social media platforms. The measurements of the variables were operationalized according to their corresponding numerical or textual descriptions in social media platforms. Prior to data collection, relevant social media was identified. Next, data collection process was implemented on the selected social media. To assure the data quality, raw data captured in the previous step was transformed and prepared for analysis. The resulting datasets were then analyzed in the last step. An overview of the research methodology is illustrated in Figure 5. Each element was discussed in detail in the following sections.
3.2 Measurement of Variables

In order to empirically investigate the research hypotheses, measurements for the variables contained in each hypothesis were clearly defined and operationalized. From previous discussion, the following relationships are hypothesized for both the purchasing social media and non-purchasing social media.

**H1.1** In purchasing social media contexts, the overall influence of an individual’s UGC is positively associated with his or her engagement level and social status indicated in the profile.

**H1.2** In non-purchasing social media contexts, the overall influence of an individual’s UGC is positively associated with his or her engagement level and social status indicated in the profile.

**H2.1** In non-purchasing social media contexts, the influence of a UGC is positively associated with the level of content co-creation in the UGC.
**H2.2** In non-purchasing social media contexts, choices of brand names and product characteristics mentioned in a UGC have a significant impact on the influence of the UGC.

**H2.3** In non-purchasing social media contexts, choices of brand names and product characteristics mentioned in a UGC have a significant impact on the level of content co-creation associated with the UGC.

**H3.** In non-purchasing social media context, customized brand marketing communication is associated with a positive change on individuals’ attitude toward the brand.

### 3.2.1 Measurements for Variables in H1.1 to H1.2

In H 1.1 to H 1.2, three variables are involved: Overall influence of an individual’s UGC, Consumers’ Online Social Status, and Social Media Engagement Level. The overall influence of an individual’s UGC was used to indicate a one’s social influence via the UGC he/she created. Meanwhile, the social status and social media engagement level describe the individual’s social media behaviors. The measurement of each variable is discussed in detail in the following sections.

*Overall influence of an individual’s UGC:* Online retailers have commonly used consumers’ response to a review as the primary way of measuring how consumers evaluate a review and it could be used as a measure of perceived value and influence in the decision-making process (Mudambi & Schuff, 2010). Therefore, the influence of a UGC could be measured
by other consumers’ responses to the UGC, including viewing, replying, indicating it as useful, accepting and using the UGC for their related decision makings. The total number of response an individual had ever received from others across all his/her UGC was used as the measurement of the overall influence of the individual’s UGC. Specifically, in the purchasing social media setting, viewers are usually provided with an option to indicate whether a review is helpful or not. So the influence of an individual’s UGC was measured by the total number of responses indicating “helpful” for his/her UGCs. In non-purchasing setting, viewers can only read the postings without having options to indicate helpfulness. Therefore, the number of views is counted and shown besides the posting. Similarly, the influence of an individual’s UGC in non-purchasing social media setting was measured by the total number of views for all his/her UGCs in this study.

**Social Media Engagement Level:** High social media engagement of an individual has been described as actively processing the role of the brand in their lives or participate in various forms of eWOM or brand related UGC creation, such as writing reviews or creating videos showcasing the product (Malthouse et al., 2013). Thus, in this study, an individual’s social media engagement level was measured by the activity level, or the total number of reviews or postings the individual created.

**Individuals’ Online Social Status:** In real worlds, social status refers to the position or rank of a person or group, within the society (Warner, Meeker, & Eells, 1949). Similarly, in most social media platforms, such as forums or product reviews on online shopping website, it is
common that a rank is placed in a user’s profile to indicate his or her overall importance in the community. For example, in Runnersworld.com, the Community Rank is defined as “the overall contributions to, and influence within, the community”. It is calculated based on the user’s overall contribution to the community from social media behaviors and interaction with followers. Therefore, the community rank suggested by the social media platform was used as a measurement of an individual’s social status.

3.2.2 Measurements for Variables in H 2.1 to H2.3

H2.1 to H2.3 only relates to non-purchasing social media and there are four variables involved: Influence of a UGC, Content Co-creation in the UGC, Choices of Brand Names in the UGC Content, and Choices of Product Characteristics in the UGC Content. These variables are related to the characteristics of each individual forum posting as hypotheses 2 focus on the influence of individual UGC.

Influence of a UGC: As addressed above, the influence of a UGC could be measured by viewers’ responses to the UGC, including viewing, replying, indicating it as useful, accepting and using the UGC for their related decision makings. In non purchasing context, such as forum, this influence consists of mainly the diffusion of product/brand knowledge to community members. Thus, influence of a UGC was measured by the total number of views a posting received.
Content Co-creation in the UGC: Active participation in writing product reviews in social media is one important form of content co-creation (Malthouse et al., 2013). In a product forum, the original message or posting is created by one consumer while the replies are contributed by many other consumers. Therefore, in this study, the number of replies a particular posting received was used to measure the degree of content co-creation associated with that particular posting.

Choices of Brand Names in the UGC Content: Because most of the discussions in product virtual communities are about products or brands, the choices of brand names in the UGC content represent what brand names were mentioned in the posting. Parsing the document collection generates a Term-Document Frequency Matrix. A Document-Term Matrix or Term-Document Matrix is a mathematical matrix that describes the frequency of terms that occur in a collection of documents (SAS Document, 2012). Each entry of the matrix represents the number of times that a term appears in a document. To measure the choices of brand names, binary variables indicating the occurrence of a brand name were constructed. These binary variables were extracted from the Term-Document Frequency Matrix consisting of variables indicating the occurrence of all terms shown in the data. In addition to individual brand names, as there are many brand names and some of them may be similar to each other, brand name clusters were calculated using Principle Component Analysis (PCA) to represent a group of brand names that are often compared together or share similar characteristics. The data containing content of UGC in the product virtual communities could come from two sources: posting title and posting body. They need to be considered individually because
titles are open to all potential viewers on the webpage but content of a specific posting can only be viewed after the viewer choose to enter into the posting page. Therefore, choices of brand names in the UGC content were obtained from the title of the UGC and body of the UGC, respectively.

*Choices of Product Characteristics in the UGC Content:* Similarly, to measure choices of product characteristics in the UGC content, binary variables indicating the occurrence of a product characteristics keyword were constructed. Product characteristics clusters were also extracted from the Term-Document Frequency Matrix consisting of variables indicating the occurrence of all terms shown in the data. As there are many product characteristics keywords and some may be similar to each other, Principle Component Analysis (PCA) was implemented to calculate product characteristics clusters representing a group of brand names that are often compared together or share similar characteristics. This step was performed by using Concept Link in SAS Text Miner (SAS Documentation, 2013). These product characteristics factors were used to fit the model. Also, they were obtained using the title of the UGC and body of the UGC, respectively.

**3.2.3 Measurement for H3**

H3 consists of only one variable: Individuals’ Attitude toward a Brand. Due to the conversational and textual nature of the contents on the social medial, individuals’ attitude toward a brand in the posting was measured by the sentiment of the brand related sentences in the posting, which is a number ranged from -1 to 1 indicating the degree of negativeness or
positiveness of the sentence calculated by Natural Language Processing (NLP). In addition to the magnitude of sentiment, the sign of sentiment was also calculated to indicate if the sentiment is negative (negative sign) or non-negative (positive sign and zero). To test the change on individuals’ attitude after viewing the posting, the attitudes related to the brand before and after replying to the posting thread created by the company representative for that brand were both calculated.

3.3 Data Collection

3.3.1 Identification of Data Source

To collect sample data for testing the proposed research hypotheses, a list of social media platforms was identified first. Accordingly, sample datasets were obtained from both purchasing social media setting and non-purchasing social media contexts. In addition, validation sample datasets were also collected from the same social media context.

For non-purchasing social media setting, an ideal sample source should be a social platform that is well known by consumers without financial affiliation with any particular brand or company, such as product forums by magazines. As one of the most popular and well known magazines for frequent runners in the U.S., Runners’ World concentrates on running related information that informs, advises, and motivates runners of all ages and abilities (Runningsworld.com, 2013). According to a survey conducted by Running USA (2013), Runnersworld.com is one of the most visited running websites in the U.S. by daily runners.
There are several forums on the Runnersworld.com with a variety of topics including training, health, nutrition, and shoes. Beyond the typical features of a product virtual community, Runnersworld.com also allows company representatives of sport goods brands to interact with community users by creating brand specific posting threads. Consequently, the sample data for non-purchasing social media setting was extracted from the shoe forum hosted by Runnerworld.com. The validation dataset was collected from the running apparel and gear forum hosted by Runnerworld.com to compare the results.

For purchasing social media setting, the ideal data source would be large and well known shopping websites that allow consumers to create reviews for the products and make suggestions for other buyers. Amazon.com is the leading shopping website offering diversified product categories in the U.S. (Jopson, 2011). Therefore, consumer reviewers’ behavioral characteristics and the corresponding “helpful” indications were collected from Amazon.com. Because Amazon.com is in the dominate position in the e-commerce market, the representativeness of the sample could be considered as sufficient. A validation dataset was sampled from Amazon.com also.

### 3.3.2 Procedures of Obtaining Data for H1 and H2

While social media websites such as product forums have a large amount of usable data, they usually contain noises as well (e.g. commercial contents, irrelevant photos) Therefore, a screening is needed to only extract needed information from the data. Instead of manually screening each posting, which can be very time consuming, web scraping technique was
adopted for data extraction. Web scraping, or screen scraping, is a process of writing scripts to parse the HTML source to extract the data while ignoring graphic links and explanatory text (Stein, 2002). It is widely used for collecting online resources to create a dataset for further analysis. In this study, web screening was performed using Python 2.7. Python is a multi-purpose programming language that has efficient high-level data structures and effective approach to object-oriented programming. Several modules developed in Python environment (e.g. Urllib, Beautiful Soup) are commonly used to web scraping, so Python is often considered an ideal language for scripting and rapid application development in many areas on most platforms (Python software foundation, 2014).

The implementation of web scraping in this study involved three major steps: 1) analysis of the structure of the website, 2) obtaining and parsing webpage source code, and 3) writing data into file. Forums usually have a nested structure. That is, each page showing postings under one topic are assigned by different URLs and are nested within topic list displayed on the main page on the forum. Moreover, profile of the users who generate the postings is stored and displayed in another web location. Therefore, the list of all URLs of the topics was extracted first via scraping all topic list pages. The titles of the postings and view count were collected in this step. The scraping process then continued for all relevant contents shown on the posting pages directed by each URL to obtain relevant posting level information including posting body, type of posting (Reply vs Posting), and posting created time. To collect the behavioral information regarding the author, the URLs to the author’s profile pages were collected for each posting. When the profile page was loaded, the user level information was collected, including user ID, activity level and community rank of the author.
All of the information is embedded in the HTML source code. To obtain the actual data, the pattern of the HTML tags near the information was identified. With the pattern, information can be captured with Regular Expression. Regular Expression, or Regex, is a programming language that is used for systematically and accurately locating specific character strings embedded in character text (Thompson, 1968).

### 3.3.2.1 Data Preparation

The raw data captured need to be treated and prepared before analysis for two purposes. First, the data directly captured from the social platforms contains only consumers’ response for each individual posting. To obtain the influence of UGC for each consumer shown in hypotheses 1, the data was aggregated by user ID. After postings and unique user IDs are captured, the summary statistics for all unique IDs were compiled based on the available information for the postings to form the user level dataset. Second, as reviews and postings are described in words, these textual information needs to be transformed into numerical format for further quantitative data analysis. Such transformation was conducted by using text mining method. Text mining, or text analytics, is the process of extracting interesting and non-trivial patterns or knowledge from unstructured text documents (Tan, 1999). Text mining as a data analytical techniques has been adopted to a variety of applications such as topic discovery, consumer relationship management, and target advertising (S. Lee, Baker, Song, & Wetherbe, 2010). For each posting, the raw dataset was transformed and classified further to obtain information regarding brand names, product models and the sentiment of the
sentences regarding the brand. Most of the techniques in text mining are to quantify textual
data and transform them into numerical data. But unlike ordinary content analysis, text
mining is capable of handling a large collection of documents much faster than human
coding.

To extract information in the textual components in the raw data from Runnersworld.com,
some issues need to be considered with caution. First, the large paragraphs in each posting
often involve multiple topics and meanings. Therefore, the analysis has to be conducted on
more semantically unified elements. In this study, the postings were divided into sentences
which are usually less diversified in terms of topics. Punctuations denoting end of sentences
(Period, Question mark, etc.) are used for the decomposition process, resulting in a collection
of sentences. Second, since forum users can use either brand names or the product names to
refer to one type of running shoes, the recognition of the running shoe brand depends not
only on the brand names but also on the names of each model under the brand. Based on a
recent survey conducted by Runblogger (2013) regarding popular running shoe models in the
market place, a local dictionary containing popular running shoe product models and
corresponding running shoe brands will be created. This dictionary was embedded in Python
script as a basis for processing each word in a sentence. Third, each sentence was also
processed by Python’ NLTK natural language processing package and the build-in classifier
to acquire an estimate of the sentiment of each sentence. As described above, the sentiment is
ranged from -1 to 1, indicating the degree of negativeness or positiveness of the sentence.
Finally, the sentence level data was aggregated by user IDs to obtain individual level data set
for further analysis. An example of web scraping and text classification process is shown in Figure 6.
Figure 6. Web scraping and text classification process example
3.3.3 Procedures of Obtaining Data for H3

Unlike the first two hypotheses, Hypotheses 3 involves comparison of posting history before and after the time when a consumer interacted with the company representative on the community. So the data collection steps are more complex and need to be performed individually according to 1) what is the time of the interaction? 2) Whether posting history exists before the time of the interaction? And 3) whether posting history exists after the time of interaction? From the obtained the sample data, posting history of the discussion threads created by company representatives of two brands (Mizuno and Saucony) as well as the ID and posting time of each forum user who participated in the discussion were identified and recorded into two different datasets, with 1720 observations for Mizuno and 350 observations for Saucony, respectively. The datasets created using the above approach were then further processed to obtain the ready-for-analysis dataset. Several data manipulation steps using Python were described as follows.

First, textual data of postings in the main dataset were separated into sentences in Python. The resulting data contains 20715 sentences. Sentiments were calculated for all sentences using Python Sentiment Analysis package and each one was labeled as “Brand related” or “Non-brand related” by searching the brand name keyword within each sentence. The IDs identified in the discussion threads were used to search for the author’s sentiment and sign of sentiment in posting history of the authors from the dataset containing brand related sentences and corresponding sentiments. The negative sign was labeled when the sentiment is negative and non-negative sign was labeled when sentiment is equal or greater than zero.

Second, for each posting created by a member (identified by ID), the posting was labeled as
content posted before participating in the marketing communication if the time of the posting is before the time of author’s first reply to the company representative’s posting. Similarly, the posting was labeled as content posted after participating into the marketing communication if the time of the posting is after the time of the author’s last reply to the company representative. Third, each of the two variables was aggregated by ID to calculate average sentiment, number of sentences, and average of the sign of sentiment (the proportion of non-negative sentence). The two variables were then joined together based on IDs to form the final dataset for analysis. The data manipulation was performed by using SQL procedure in SAS 9.3. The data manipulation and analysis procedure for H3 was summarized in Figure 7.
Figure 7. Data Manipulation and Analysis Procedures for H3

3.4 Data Description

3.4.1 Description of Amazon.com Data

Using the procedures described above, two samples were randomly collected from Amazon.com separately: one served as the testing dataset and the other one served as validation dataset. The data were obtained from the profile page of reviewers registered with Amazon.com. Variables collected for each sample member (labeled with a user ID) include the consumer’s community rank, the consumer’s total number of reviews posted, and the consumer’s total number of “helpful” votes received from other consumers. In addition to
these variables, the percentage of having one or more “helpful” votes for each consumer
reviews was also collected. The relatively high percentage of having one or more “helpful”
votes for the consumer reviews (Mean=0.86, STD=0.064) suggests the use of “helpful” votes
is more appropriate for measuring individuals’ influence rather than “helpful” percentage. So
the consumer’s community rank, the consumer’s total number of reviews posted, and the
consumer’s total numbers of “helpful” votes received from other consumers are behavioral
indicators of each sample subject. Totally, 2000 observations were collected. Members
whose total number of “helpful” votes is less than 30 were excluded from the sample data.
The resulting main dataset includes 1691 observations. Similarly, another 1000 observations
were randomly collected. After removing members whose total number of “helpful” votes is
less than 30, a total of 947 observations were kept. This dataset is used as validation dataset.
Descriptive analysis was performed on the sample data. Table 1 shows the results for the
three variables in the main and validation data sets.
Table 1. Descriptive Statistics of Variables in the Amazon.com Dataset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td>Total number of “helpfulness” votes</td>
<td>1691</td>
<td>2267.27</td>
<td>5637.02</td>
<td>47</td>
<td>143508</td>
</tr>
<tr>
<td></td>
<td>Total number of reviews</td>
<td>1691</td>
<td>349.6399</td>
<td>534.8147</td>
<td>8</td>
<td>11678</td>
</tr>
<tr>
<td></td>
<td>Rank</td>
<td>1691</td>
<td>5090.39</td>
<td>2901.61</td>
<td>41</td>
<td>9980</td>
</tr>
<tr>
<td>Validation</td>
<td>Total number of “helpfulness” votes</td>
<td>947</td>
<td>2203.06</td>
<td>4889.9</td>
<td>0</td>
<td>59914</td>
</tr>
<tr>
<td></td>
<td>Total number of reviews</td>
<td>947</td>
<td>377.313</td>
<td>477.3067</td>
<td>0</td>
<td>4666</td>
</tr>
<tr>
<td></td>
<td>Rank</td>
<td>947</td>
<td>5157.65</td>
<td>2853.34</td>
<td>21</td>
<td>9960</td>
</tr>
</tbody>
</table>

Distribution of total number of “helpful” votes was highly right skewed in both the main dataset (Mean= 2285.7, SD= 4513) and the validation dataset (Mean= 2267.2, SD= 5637.4). This result suggests possible log transformation for fitting linear model. The histograms of total number of “helpful” votes in the main dataset and validation dataset are shown below in Figure 8.
3.4.2 Description of Runnersworld.com Data

In Runnersworld.com, two sample datasets containing the profile and posting history of registered forum users were collected. The main dataset was obtained from the running shoe forum and the validation dataset was obtained from the apparel and gear forum. The forums only allow public view of the recent two years’ posting history. In these two datasets, the following information was recorded: time when posting was created, the authors’ ID, total number of times viewed by other people, title of the posting, content of the posting,

Figure 8. Histograms of Total Number of “helpful” Votes in Amazon.com Dataset
community rank of the author, and total number of postings created by the author in the past two years. Similarly, only postings receiving more than 30 views were included in the data. As a result, main dataset contains 1615 observations from the running shoes forum and validation dataset contains 558 observations from the apparel and gear forum. Descriptive analysis was performed first and results are shown in Table 2.

Table 2. Descriptive Statistics of Variables in Runnersworld.com Dataset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main</strong></td>
<td>Total number of views</td>
<td>1615</td>
<td>2234.66</td>
<td>4856.08</td>
<td>33</td>
<td>102470</td>
</tr>
<tr>
<td></td>
<td>Community rank</td>
<td>1615</td>
<td>4846.76</td>
<td>3931.36</td>
<td>24</td>
<td>15055</td>
</tr>
<tr>
<td></td>
<td>Total number of postings</td>
<td>1615</td>
<td>684.1358</td>
<td>2070.05</td>
<td>8</td>
<td>19191</td>
</tr>
<tr>
<td><strong>Validation</strong></td>
<td>Total number of views</td>
<td>558</td>
<td>2087.41</td>
<td>5422.19</td>
<td>5</td>
<td>107711</td>
</tr>
<tr>
<td></td>
<td>Community rank</td>
<td>558</td>
<td>4155.59</td>
<td>3932.24</td>
<td>4</td>
<td>15103</td>
</tr>
<tr>
<td></td>
<td>Total number of postings</td>
<td>558</td>
<td>1093.48</td>
<td>2771.85</td>
<td>9</td>
<td>38491</td>
</tr>
</tbody>
</table>

The results indicate that distribution of the community users’ total number of views was highly right skewed in both main dataset (Mean= 2234, SD= 4856, skewness =11.5) and the validation dataset (Mean= 2087, SD= 5422, skewness =13.7). Moreover, the number of replies for the postings in Runnersworld.com is also highly skewed (Mean=6.6, SD=12.98, skewness =15.6 in the main dataset; Mean=7.3, SD=10.1 skewness =5.83 in the validation dataset). These evidences suggest possible log transformation of the data for further
modeling. The histograms of the total number of views in main dataset and validation dataset are shown in Figure 9.

![Histograms of the Total Number of Views in Runnersworld.com Dataset](image)

**Figure 9. Histograms of the Total Number of Views in Runnersworld.com Dataset**

To identify the brand names embedded in the postings, based on the data shown on shoe finder page on the Runners’ World website and other popular running shoes reviews, a local dictionary containing 37 popular running shoe brands and a total number of 186 latest
running shoes series for all 37 brands was created. A search algorithm was then applied using this dictionary to capture the running shoe series in the text. The results were aggregated back to brand name level to generate variables containing the number of mentioning of each brand name. The overall results are shown in Table 3.
To obtain product characteristics shown in the posting content, a group of keywords representing running shoes attributes were collected from several running shoes review
websites. Next, these words were selected from the Noun terms appeared in the texts. Only terms having both higher overall frequency (Term Frequency > 100) and higher appearance in different postings (Document Frequency>100) were kept in a product characteristics keyword list. As a result, 22 keywords were identified and recorded. The results are shown in Table 4.

<table>
<thead>
<tr>
<th>TERM</th>
<th>ROLE</th>
<th>WEIGHT</th>
<th>FREQ</th>
<th>NUM OF DOCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>arch</td>
<td>Noun</td>
<td>0.306</td>
<td>712</td>
<td>309</td>
</tr>
<tr>
<td>comfortable</td>
<td>Noun</td>
<td>0.277</td>
<td>343</td>
<td>246</td>
</tr>
<tr>
<td>cushion</td>
<td>Noun</td>
<td>0.425</td>
<td>2306</td>
<td>694</td>
</tr>
<tr>
<td>fit</td>
<td>Noun</td>
<td>0.34</td>
<td>776</td>
<td>397</td>
</tr>
<tr>
<td>flexibility</td>
<td>Noun</td>
<td>0.155</td>
<td>172</td>
<td>105</td>
</tr>
<tr>
<td>forefoot</td>
<td>Noun</td>
<td>0.309</td>
<td>1179</td>
<td>390</td>
</tr>
<tr>
<td>issue</td>
<td>Noun</td>
<td>0.367</td>
<td>996</td>
<td>434</td>
</tr>
<tr>
<td>material</td>
<td>Noun</td>
<td>0.178</td>
<td>220</td>
<td>112</td>
</tr>
<tr>
<td>midfoot</td>
<td>Noun</td>
<td>0.268</td>
<td>458</td>
<td>212</td>
</tr>
<tr>
<td>midsole</td>
<td>Noun</td>
<td>0.183</td>
<td>338</td>
<td>161</td>
</tr>
<tr>
<td>money</td>
<td>Noun</td>
<td>0.176</td>
<td>161</td>
<td>108</td>
</tr>
<tr>
<td>motion</td>
<td>Noun</td>
<td>0.162</td>
<td>220</td>
<td>105</td>
</tr>
<tr>
<td>outsole</td>
<td>Noun</td>
<td>0.192</td>
<td>222</td>
<td>165</td>
</tr>
<tr>
<td>price</td>
<td>Noun</td>
<td>0.219</td>
<td>378</td>
<td>185</td>
</tr>
<tr>
<td>problem</td>
<td>Noun</td>
<td>0.418</td>
<td>1302</td>
<td>515</td>
</tr>
<tr>
<td>shop</td>
<td>Noun</td>
<td>0.181</td>
<td>192</td>
<td>112</td>
</tr>
<tr>
<td>size</td>
<td>Noun</td>
<td>0.363</td>
<td>1305</td>
<td>404</td>
</tr>
<tr>
<td>sole</td>
<td>Noun</td>
<td>0.215</td>
<td>289</td>
<td>166</td>
</tr>
<tr>
<td>stability</td>
<td>Noun</td>
<td>0.366</td>
<td>1491</td>
<td>432</td>
</tr>
<tr>
<td>store</td>
<td>Noun</td>
<td>0.414</td>
<td>1238</td>
<td>491</td>
</tr>
<tr>
<td>support</td>
<td>Noun</td>
<td>0.636</td>
<td>1171</td>
<td>738</td>
</tr>
<tr>
<td>weight</td>
<td>Noun</td>
<td>0.371</td>
<td>619</td>
<td>435</td>
</tr>
</tbody>
</table>
While consumers on Amazon.com might be as diverse as the general population, the members who participate in the forums in Runnersworld.com were believed to have some shared characteristics, such as their interest in running, that are unique to a particular consumers group. The descriptive analysis of the sample dataset suggests some important behavioral characteristics of the community members. First, the UGC in the community were mostly contributed by some key active community members. The distribution of the number of postings by member ID is shown in Figure 10. Majority of the postings were created by a small number of forum users, indicating a small proportion of high-engagement active users.

Figure 10. Distribution of the Number of Postings by Member ID
Second, the brand knowledge varies largely among active community users. Figure 11 shows the histogram of members’ total number of unique brand names mentioned in their postings. The histogram indicates that while most of the users have never mentioned any brand names, some of them mentioned as many as 20 unique brand names. These evidences show that levels of brand involvement and social media engagement differ dramatically among active members of the community.

![Histogram of Members’ Total Number of Unique Brand Names Mentioned](image)

**Figure 11. Histogram of Members’ Total Number of Unique Brand Names Mentioned**
Third, most of the posting activities might occur during members’ leisure time. Figure 12 and Figure 13 show the average daily posting amount by hours and average weekly posting amount by days. These graphs suggest the community members tend to post during working hours (9am-4pm) and between Monday and Friday. There are two increasing periods for posting activities. The posting amount increased dramatically in the morning at 9:00 am and decreased. Then, more UGC were posted starting at 7:00 pm and decreased after 10:00 pm.

![Figure 12. Average Daily Posting Amount by Hours](image)

Figure 12. Average Daily Posting Amount by Hours
Lastly, the data shows that community members not only try to participate in the customized online marketing communication but also constantly involved. After screening, the resulting datasets for testing H3 include 35 unique IDs for one running shoe brand (Mizuno) and 23 unique IDs for another running shoes brand (Saucony). The sample size is small because the screening process requires that consumers must have participated in the company-created posting threads and must have posting history before and after the participation period. The participations activities of the two company-created posting threads from 2013 to 2014 are shown in Figure 14 and Figure 15. These plots show that consumers constantly participated in the company-created postings threads. They also show various levels of posting activities outside the company-created postings threads.
Figure 14. UGC Creation in the Company-Created Posting Threads
3.5 Methods for Data Analysis

3.5.1 Data Analysis Method for H1.1, H1.2 and H2.1

After the data was prepared using the procedures described above, the resulting datasets were analyzed for the corresponding hypotheses. Hypotheses 1.1 and 1.2 proposed that there is a positive association between an individual’s social media behavioral characteristics and the overall influence of the individual’s UGC. Similarly, Hypothesis 2.1 proposed a positive
association between content co-creation of a UGC and the influence of the UGC. According to the descriptive analysis, the overall influence of an individual’s UGC and influence of a UGC were highly skewed, so they were transformed using a logarithm function. Therefore, the data was fit to a log-linear model, or equivalently, a linear regression was fit to the log transformed data. Because the hypothesis was to test their individual effect, social status, engagement level and content co-creation were used as predictors individually in each model. This model is shown as equation (1) and it was applied in testing H1.1, H1.2 and H2.1.

\[
\text{Log (} y_i \mid x_1) = \beta_0 + \beta_1 x_1 + \varepsilon_i 
\]  \hspace{1cm} (1)

\[
y_i \mid x = \exp (\beta_0 + \beta_1 x_1 + \varepsilon_i) 
\]  \hspace{1cm} (2)

\[
y_i \mid x_1 + 1 = \exp (\beta_0 + \beta_1 (x_1 + 1) + \varepsilon_i) = \exp (\beta_0 + \beta_1 x_1 + \beta_1 + \varepsilon_i) = y_i \mid x_1 \times \exp (\beta_1) 
\]  \hspace{1cm} (3)

Where,

\( y_i \) is the overall influence of an individual’s UGC.

\( x_1 \) is the individual’s social status / engagement level / content co-creation.

Equation (1) is the linear model of log transformed \( y \).

Equation (2) is the corresponding model of \( y \) on original scale.

Equation (3) is the relationship between \( y \) given \( x \) and a new \( y \) given that \( x \) increases by 1 unit.
Because of the log transformation, coefficients in these models represent a linear relationship between the predictors and the log transformed response. Therefore, to interpret the effect, the log transformed response needs to be transformed back to original scales, which is a non-linear relationship. The interpretation of the model parameters for the original response variable is illustrated by equation (2) and equation (3), indicating a multiplicative effect. Namely, n unit increase in the independent variable corresponds to exp(\(\beta\)) to the power of n times increase in the dependent variable. Next, main dataset and validation dataset were used to build two models. Consistency of the fitted models was evaluated by comparing the regression estimates of the two models.

### 3.5.2 Data Analysis Method for H2.2 and H2.3

Hypothesis 2.2 proposed a significant impact of the content of a UGC on the influence of the UGC. Hypothesis H2.3 proposed a significant impact of the content of a UGC on the content co-creation of a UGC. Due to the high skewness in the distributions of both influence of the UGC and content co-creation, log-linear models were also fitted to the main dataset and the validation dataset as shown in equation (4).

\[
\log(y_i | X) = \beta_0 + \mathbf{X}_i \mathbf{\beta} + \epsilon_i
\]  

(4)

Where,

\(y_i\) is influence of a UGC or the content co-creation of a UGC.
X is: The dimension reduced full Term-Document Matrix by Singular Value Decomposition (SVD), or Term-Document Matrix containing only brand names/product characteristics, or brand names or product characteristics clusters.

β is the vector of parameters.

To include textual data into the model, UGC content was numerically transformed by the bag-of-words representation using Term-Document Matrix, where each cell indicates the times of occurrence of the terms. Next, there are several steps involved in this method. First, to test if UGC content has any effect on the influence of the UGC or content co-creation of the UGC, the full Term-Document Matrix containing all terms was used as a predictor in a log-linear model. Due a large number of terms (n=3086) in the full Term-Document Matrix, the dimension of the matrix was reduced by Singular Value Decomposition (SVD). Namely, column projections of the Term-Document Matrix that best fit the data were used to represent the distinct concepts in the texts (SAS, 2013). Because of limited computational resources, 100 column projects, or distinct concepts, were used to in SVD. Second, brand names and product characteristics represent an important part of the information contained in the texts, so Term-Document Matrix containing only brand names or product characteristics was used as predictors of a log-linear model. The matrices of brand names and product characteristics were entered into the model individually. Third, due to the relatively large number of brand names (n=37) and product characteristics (n=22), to evaluate their individual effects of them on the influence of UGC and content co-creation of UGC, these keywords were clustered.
using Variable Clustering Node (based on algorithms similar to PCA) in SAS Enterprise Miner 7.1. The resulting clusters were entered into the log-linear models as predictors.

As a result, three regression models were built in this step: 1) a log-linear model including all terms as predictors 2) a log-linear model including only brand names or product characteristics as predictors, and 3) a log-linear model including variable clusters of brand names or product characteristics as predictors. As body of the posting is not visible prior to viewing behavior, only title of the posting would affect viewing decision. Accordingly, body of the posting would affect replying decisions. Thus, only title of the posting was used as UGC content in testing H2.2. And only body of the posting was used as UGC content in testing H2.3. Furthermore, variable selection was performed using stepwise algorithm based on p-valuates smaller than 0.05. As brand names and product characteristics are about running shoes, data collected from apparel and gear forum can’t be applied for validation. Therefore, the running shoes forum dataset itself was partitioned into training (60 percent of total observations) and validation datasets (40 percent of total observations) to avoid overfitting. Models fitted by main dataset and the validation dataset were then compared. The analysis was performed using Text Miner Module provided by SAS Enterprise Miner 7.1.

3.5.3 Data Analysis Method for H3

Hypotheses 3 proposed the customized brand marketing on social media has a positive impact on community members’ attitude toward the brand. Because each data point contains two types of attitudes toward a brand (average sentiment and proportion of non-negative
sentences) and each type contains two observations (before and after interaction with brand representatives) for the same individual, the data was analyzed by paired t-test to determine if the paired difference is significant different from zero. The paired difference was first defined in terms of the average of sentiment. The hypothesis is formulated as:

$$H_0: \mu_{\text{sentiment difference}} = 0 \quad H_a: \mu_{\text{sentiment difference}} \neq 0$$

Where,

$$\mu_{\text{sentiment difference}} = E [E [\text{Sentiment}_{\text{after } i}] - E [\text{Sentiment}_{\text{before } i}]],$$

which is the mean of paired difference in average sentiment.

$\text{Sentiment}_{\text{before } i}$ is the sentiment for participant $i$ before interacting with brand representative.

$\text{Sentiment}_{\text{after } i}$ is the sentiment for participant $i$ for participant $i$ after interacting with brand representative.

The paired difference was also defined in terms of the proportion of non-negative sentences. Paired t-test was also used in this step assuming normal approximation. The hypothesis is formulated as:

$$H_0: \mu_{\text{proportion difference}} = 0 \quad H_a: \mu_{\text{proportion difference}} \neq 0$$
Where,

\[ \mu_{\text{proportion difference}} = E \left[ E \left[ I_{\text{after } i} \right] - E \left[ I_{\text{before } i} \right] \right], \]

which is the mean of paired difference in the proportion of non-negative sentences.

\( I_{\text{before } i} \) is the indicator function of negative sentiment for participant \( i \) before interacting with brand representative such that:

\[
I(x) := \begin{cases} 
1 & \text{if } x \geq 0 \\
0 & \text{if } x < 0 
\end{cases}
\]

\( I_{\text{after } i} \) is the indicator function of negative sentiment for participant \( i \) after interacting with brand representative.

Due to the small sample size, normal assumption may not hold well. So, in addition to paired t-test, Wilcoxon signed rank test, a non-parametric paired difference test based on testing median, was also performed to validate the result of t-test. The test statistics is shown as the following:

\[
W = \left| \sum_{i=1}^{N_f} \text{sgn}(x_{2,i} - x_{1,i}) \cdot R_i \right|
\]  

(5)

Where,

\( x_{1i} \) and \( x_{2i} \) are the average sentiment or the proportion of non-negative sentences before and after interaction for participant \( i \).
Sgn is the sign function such that:

\[
\text{sgn}(x) := \begin{cases} 
-1 & \text{if } x < 0, \\
0 & \text{if } x = 0, \\
1 & \text{if } x > 0.
\end{cases}
\]

\(R_i\) is the rank of pair \(i\).
Chapter 4 Data Analysis and Results

4.1 Results of Hypothesis 1.1 and 1.2

To test H1.1, two linear regressions were fit to the log transformed data collected from Amazon.com. Individual’s social status (measured by the reviewer’s community rank) and engagement level (measured by total number of reviews) were used as predictors individually to model the influence of his/her UGC (measured by total number of “helpful” votes received for all his/her reviews). Diagnostics were performed to check influential points, outliers, and important assumptions used in linear regression model (e.g. normality of the residuals and constant variances). Extreme values in the predictors (larger than 95 percent percentile) were removed from the original data. Observations that are considered as outliers (studentized residuals > 3) and influential (Cook’s Distance >1) were also removed. The model parameters were estimated from the resulting dataset. The same modeling procedures were performed on the validation dataset. The model estimates are summarized in Table 5. The regression lines and residual plots after log transformation are shown in Figure 16.
### Table 5. Results of Log-linear Models for H1.1

**Dependent variable:** Overall influence of an individual’s UGC in purchasing settings

<table>
<thead>
<tr>
<th></th>
<th>Main Data (n=1691)</th>
<th>p-value</th>
<th>Validation Data (n=947)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Status:</td>
<td>-0.000235</td>
<td>&lt;.001</td>
<td>-0.000229</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Intercept:</td>
<td>8.17</td>
<td>&lt;.001</td>
<td>8.18</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Engagement Level:</td>
<td>0.002</td>
<td>&lt;.001</td>
<td>0.00164</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Intercept:</td>
<td>6.35</td>
<td>&lt;.001</td>
<td>6.45</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
The results indicate a significant linear correlation between an individual’s engagement level and the log transformed influence of his/her UGC. The results also indicate a significant linear relationship between an individual’s community rank and the log transformed influence of his/her UGC. To compare the model performance, the mean predictive values and mean actual values were plotted for both the main model and the validation model (as shown in Figure 17). The plot indicates that model performances are consistent for both the main and validation datasets.
Because of the log transformation, the exponential of parameter (i.e. \( \exp (\beta_1) \)) represents the multiplicative effect on the original scale of the dependent variable for one unit increase in the independent variable. Consequently, the model suggests that each 100 points increase in an individual’s social status (higher rank in the community) in purchasing social media settings means that the influence of his/her UGC is approximately 1.02 times higher than before in terms of total number of “helpful” votes. Furthermore, each increase of 100 reviews in an individual’s engagement level in purchasing social media settings means that the influence of the individual’s UGC is approximately 1.2 times higher in terms of total number of “helpful” votes.

**Figure 17. Comparison of Model Performance for H1.1**
To test H1.2, two linear regression models were fit to the log transformed data collected from Runnersworld.com. Similarly, individual’s social status (measured by the member’s community rank) and engagement level (measured by the total number of postings) were used as predictors individually to model the influence of his/her UGC (measured by the total number of views he or she received). The estimates of regression models are summarized in Table 6. The regression lines and residual plots are shown in Figure 18. The results suggest a significant linear relationship between a community member’s engagement level and the log transformed overall influence of his/her UGC. The results also suggest a significant linear relationship between a community member’s social status and the log transformed overall influence of his/her UGC.

Table 6. Results of Log-linear Models for H1.2

<table>
<thead>
<tr>
<th>Dependent variable: Overall influence of an individual’s UGC in non purchasing settings</th>
<th>Main Dataset</th>
<th>p-value</th>
<th>Validation Dataset</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n=1615)</td>
<td></td>
<td>(n=558)</td>
<td></td>
</tr>
<tr>
<td>Community rank:</td>
<td>-0.00002113</td>
<td>0.0135</td>
<td>-0.00004361</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Intercept:</td>
<td>7.19</td>
<td>&lt;.001</td>
<td>7.04</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of postings:</td>
<td>0.0001805</td>
<td>&lt;.001</td>
<td>0.00005189</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Intercept:</td>
<td>7.05</td>
<td>&lt;.001</td>
<td>6.78</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
The Mean predictive values and mean actual values were plotted for the main model and the validation model. The plot is shown in Figure 19. Unlike the plot for Amazon.com data, the plot for Runnersworld.com data shows a large variation between the main dataset and the validation dataset. After transforming back the coefficient to original scales, the model suggests that each 1000 points increase in an individual’s social status (higher rank in the community) in non-purchasing social media settings means that the influence of his/her UGC is approximately 1.03 times higher than before in terms of total number of views.
Furthermore, each increase of 1000 reviews in an individual’s engagement level in non-purchasing social media settings means that the influence of the individual’s UGC is approximately 1.15 times higher in terms of total number of views.

Figure 19. Comparison of Model Performance for H1.2
4.2 Results of Hypothesis 2.1 to 2.3

4.2.1 Hypothesis 2.1

To test H 2.1, one linear regression model was fit to the log transformed data collected from Runnersworld.com. Namely, content co-creation (measured by the number of replies for the posting) was used as a predictor to model the influence of the UGC (measured by the total number of views for the posting). Diagnostics were performed and influential observations and outliers were removed based on the same criteria used in the previous section. The model estimates are summarized and shown in Table 7. The result suggests that content co-creation was significantly correlated with the log transformed influence of the UGC.

<table>
<thead>
<tr>
<th>Dependent Variable: Influence of the UGC</th>
<th>Main Dataset (n=1615)</th>
<th>p-value</th>
<th>Validation Dataset (n=558)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Replies:</td>
<td>0.078</td>
<td>&lt;.001</td>
<td>0.084</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Intercept:</td>
<td>6.63</td>
<td>&lt;.001</td>
<td>6.34</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Regression lines and residual plots are shown in Figure 20. Surprisingly, after log transformation, the histogram of the residuals shows a bimodal distribution, suggesting a possible uncontrolled class variable. Namely, the relationship between the number of views and the number of replies differ between two categories existing in the postings. Therefore,
two regression lines might be fitted individually for each group. Nevertheless, the joint
density estimation and heat map of the model residuals (Shown in 3D in Figure 21) suggest
that the effect of the uncontrolled class variable is primarily on intercept, but not on the linear
coefficients of the linear relationship between number of replies and number of views. The
Mean predictive values and mean actual values were plotted for the main model and the
validation model and the plot is shown in Figure 22. The plot indicates that model
performances are consistent for both the main and the validation datasets.
Figure 20. Regression Lines and Residual Plots for H2.1
Figure 21. Joint Density Estimation and Heat Maps
4.2.2 Hypothesis 2.2

To test H2.2, three linear models were built for modeling the log transformed influence of UGC: 1) a regression model including all terms (measured by Term-document Matrix after SVD) in the posting title as predictors, 2) a regression model including only brand names (measured by Term-document Matrix for only brand name keywords) in the posting title as predictors, and 3) a regression model including only product characteristics (measured by Term-document Matrix for only product characteristics keywords) in the posting title as predictors. The results are summarized and shown in Table 8. The model R-squares suggests that all three models did not fit the data well. Even if all terms in the posting title were included in the model using bag-of-words reorientation, they could only explain less than 10
percent of the total variance of the influence of UGC. The total variation (SSE) explained by brand names and product characteristics in the posting title together contributes about 35 percent of the total variation explained by the model using all terms in the posting title.
### Table 8. Summary of Results of Log-linear Models for H2.2

**Dependent Variable: Influence of UGC**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>18</td>
<td>140.822058</td>
<td>7.823448</td>
<td>7.83</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>1666</td>
<td>1664.675597</td>
<td>0.999205</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected</td>
<td></td>
<td>1684</td>
<td>1805.497654</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Model Fit Statistics**

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Adj R-Sq</th>
<th>AIC</th>
<th>BIC</th>
<th>SBC</th>
<th>C(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.078</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.068</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>16.08</td>
<td>16.08</td>
<td>13.56</td>
<td>0.0002</td>
</tr>
<tr>
<td>Error</td>
<td>1019</td>
<td>1208.422</td>
<td>1.18589</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected</td>
<td></td>
<td>1020</td>
<td>1224.50772</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Model Fit Statistics**

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Adj R-Sq</th>
<th>AIC</th>
<th>BIC</th>
<th>SBC</th>
<th>C(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0131</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0122</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>33.475615</td>
<td>8.368904</td>
<td>7.14</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>1016</td>
<td>1191.032105</td>
<td>1.172276</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected</td>
<td></td>
<td>1020</td>
<td>1224.50772</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Model Fit Statistics**

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Adj R-Sq</th>
<th>AIC</th>
<th>BIC</th>
<th>SBC</th>
<th>C(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0273</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0235</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Model Fit Statistics**

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Adj R-Sq</th>
<th>AIC</th>
<th>BIC</th>
<th>SBC</th>
<th>C(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.2.3 Hypothesis 2.3

To test H2.3, three log-linear models were built for modeling the log transformed content co-creation: 1) a regression model including all terms (measured by Term-document Matrix after SVD) in the posting body as predictors, 2) a regression model including only brand names (measured by Term-document Matrix for only brand name keywords) in the posting body as predictors, and 3) a regression model including only product characteristics (measured by Term-document Matrix for only product characteristics keywords) in the posting body as predictors. The results of the three regression models are summarized in Table 9. Unlike regression models for the influence of UGC, all the three models for content co-creation show good fit to the data. Model R-square in the first model suggests that a significant proportion (73 percent) of the variance in the content co-creation was explained by including all terms in the posting body. Brand names ($R^2=0.38$) alone and product characteristics ($R^2=0.45$) alone could also explain a large proportion of the variance in the content co-creation. Combining model sum of squares (SSE) of brand names and product characteristics in the posting body together, they contribute about 76 percent of the total variation that can be explained by the model including all terms in the posting body.
Table 9. Summary of Results of Log-linear Models for H2.2

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>12</td>
<td>1082.0848</td>
<td>90.173733</td>
<td>394.38</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>1688</td>
<td>385.960053</td>
<td>0.228649</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>1700</td>
<td>1468.044853</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Model Fit Statistics**

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Adj R-Sq</th>
<th>BIC</th>
<th>-2494.8984</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>-2496.99</td>
<td>BIC</td>
<td>-2494.8984</td>
</tr>
<tr>
<td>SBC</td>
<td>-2426.28</td>
<td>C(p)</td>
<td>20.201</td>
</tr>
</tbody>
</table>

**Product Characteristics Keywords in the Posting Body**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>18</td>
<td>452.051354</td>
<td>25.113964</td>
<td>54.42</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>1172</td>
<td>540.829914</td>
<td>0.461459</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>1190</td>
<td>992.881268</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Model Fit Statistics**

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Adj R-Sq</th>
<th>BIC</th>
<th>-899.52</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>-902.228</td>
<td>BIC</td>
<td>-899.52</td>
</tr>
<tr>
<td>SBC</td>
<td>-805.659</td>
<td>C(p)</td>
<td>16.2537</td>
</tr>
</tbody>
</table>

**Brand Name Keywords in the Posting Body**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>15</td>
<td>378.823114</td>
<td>25.254874</td>
<td>48.33</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>1175</td>
<td>614.058154</td>
<td>0.522603</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>1190</td>
<td>992.881268</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Model Fit Statistics**

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Adj R-Sq</th>
<th>BIC</th>
<th>-754.7785</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>-756.989</td>
<td>BIC</td>
<td>-754.7785</td>
</tr>
<tr>
<td>SBC</td>
<td>-675.668</td>
<td>C(p)</td>
<td>24.2271</td>
</tr>
</tbody>
</table>
4.2.4 Brand Names and Product Characteristics Clusters

In the previous step, the results of the log-linear models indicate that content co-creation could be well explained by the inclusion of brand names and product characteristics in the posting body. Next, the effects of the clustered brand names and product characteristics on content co-creation were explored further. Cluster analysis was performed on the Term-Document Matrix containing only the 37 brand name keywords in posting body. Figure 23 shows the clustering result. Each small node indicates individual brand names and each large node indicates the cluster membership. Lines connecting them suggest the existence of correlation.
After clustering, the resulting 9 brand cluster variables were then used as predictors in a regression model with log transformed content co-creation as dependent variable. The results of the model are shown in Table 10. Among others, cluster 1 (β=0.29) and cluster 4 (β=0.19) were estimated to be the most influential brand clusters. Compared with regression model using all individual brand names (R-square =0.38), the result suggests a satisfactory performance (R-square =0.34). As shown in Figure 24, the model performances are consistent in main dataset and validation dataset.
Table 10. Results of Log-linear Models using Brand Name Clusters

<table>
<thead>
<tr>
<th>Brand Name Clusters in Posting Body</th>
<th>Model Fit Statistics</th>
<th>Analysis of Maximum Likelihood Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R-Square 0.341</td>
<td>Parameter</td>
</tr>
<tr>
<td></td>
<td>Adj R-Sq 0.3383</td>
<td>Intercept</td>
</tr>
<tr>
<td></td>
<td>BIC -701.431</td>
<td>Clus1</td>
</tr>
<tr>
<td></td>
<td>C(p) -670.936</td>
<td>Clus3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clus4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clus6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clus7</td>
</tr>
</tbody>
</table>

Dependent Variable: Content Co-creation of UGC
Next, a total number of 22 product characteristics keywords obtained in posting body from the data were clustered. The resulting dataset contains 5 clusters as shown in Figure 25. According to the meaning of the keywords in each cluster, these clusters could be labeled as the following product characteristics of running shoes: comfort (cluster 1), material (cluster 2), performance (cluster 3), price (cluster 4), and fit issue (cluster 5).
These product characteristic clusters were then used as predictors for a regression model with log transformed content co-creation as dependent variable. As a result, all the 5 clusters showed significant influence on the number of replies. Among others, comfort (cluster 1, $\beta=0.23$) and fit issues (cluster 5, $\beta=0.25$) were estimated to be more influential than other product characteristics ($\beta<0.2$). The result of the regression model is illustrated in Table 11. As shown in Figure 26, the model performances are consistent between the training dataset and the validation dataset.
Table 11. Results of Log-linear Models using Product Characteristic Clusters

<table>
<thead>
<tr>
<th>Dependent Variable: Content Co-creation of UGC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Fit Statistics</td>
</tr>
<tr>
<td>R-Square 0.4388  Adj R-Sq 0.4364</td>
</tr>
<tr>
<td>AIC -892.672  BIC -890.6109</td>
</tr>
<tr>
<td>SBC -862.176  C(p) 6</td>
</tr>
<tr>
<td>Analysis of Maximum Likelihood Estimates</td>
</tr>
<tr>
<td>Parameter  DF  Standard Estimate Error  t Value  Pr &gt;</td>
</tr>
<tr>
<td>Intercept  1  1.5729  0.0199  79.16  &lt;.0001</td>
</tr>
<tr>
<td>Clus1      1  0.2326  0.0231  10.08  &lt;.0001</td>
</tr>
<tr>
<td>Clus2      1  0.1313  0.0213  6.17   &lt;.0001</td>
</tr>
<tr>
<td>Clus3      1  0.1662  0.0224  7.43   &lt;.0001</td>
</tr>
<tr>
<td>Clus4      1  0.1302  0.0206  6.31   &lt;.0001</td>
</tr>
<tr>
<td>Clus5      1  0.2541  0.0228  11.14  &lt;.0001</td>
</tr>
</tbody>
</table>

Figure 26. Comparison of Model Performance: Product Characteristic Clusters
4. 3 Results of Hypothesis 3

4.3.1 Magnitude Change of Attitude

Testing of H3 was performed based on the customized brand marketing communication data on two brands (Mizuno and Saucony) collected from the running shoes forum hosted by Rnnersworld.com. The attitudinal change in the participants’ posting behaviors before and after interacting with brand representatives was shown in Figure 27 for both Mizuno and Saucony. Specifically, the two graphs show histograms and approximated distributions of paired difference in the participants’ average sentiment between before and after interacting with brand representatives toward Mizuno and Saucony, respectively. Figure 28 shows the each participant’s change in average sentiment (represented by the regression lines) toward Mizuno and Saucony. The average of the regression lines (in red) for all participants is also shown in the graph, indicating the overall change in average sentiment.
Figure 27. Distributions of Paired Difference in Average Sentiment (After-Before)
The attitudinal change in participants’ posting behaviors before and after interacting with brand representatives was tested using paired t-test. In addition, Wilcoxon signed rank test was also performed as an alternative of paired t-test. The results are summarized and shown in Table 12. Both tests suggest a non-significant difference between the average of sentiment before and after the interaction. The non-significant results are also consistent between Mizuno and Saucony.
Table 12. Paired T-test & Wilcoxon Signed Rank Test for Paired Difference in Average Sentiment

| After - Before | Mizuno | | Saucony | |
|---------------|--------|------------------|--------|
| **Test** | **Statistic** | **p - Value** | **Test** | **Statistic** | **p - Value** |
| Student's t | $t = 1.053004$ | $Pr > |t| = 0.2996$ | Student's t | $t = 0.234689$ | $Pr > |t| = 0.8164$ |
| Signed Rank | $S = 34.5$ | $Pr >= |S| = 0.5949$ | Signed Rank | $S = 13$ | $Pr >= |S| = 0.7017$ |

Finally, the linear relationship between the total number of sentences related to the brand and the attitude change toward the brand was tested. The results suggest a non-significant linear relationship for Mizuno posting thread ($t = 0.06$, $p$-value $= 0.95$) and Saucony posting thread ($t = 1.34$, $p$-value $= 0.19$) in terms of change in the average sentiment. The non-significant linear relationship are consistent for Mizuno posting thread ($t = 0.1$, $p$-value $= 0.92$) and Saucony posting thread ($t = 0.4$, $p$-value $= 0.69$) in terms of change in the proportion of non-negative sentences.

4.3.2 Proportional Change of Non-negative Attitude

In addition to the change of average sentiment, a different approach to measure the change was adopted also using the proportion of non-negative sentences in each participant’s posting
Difference in the proportion of non-negative sentences between before and after interacting with brand representatives was also tested using paired t-test and Wilcoxon signed rank test. Figure 29 shows the histograms and approximated distributions of paired difference in proportion of non-negative sentences between before and after interacting with brand representatives toward Mizuno and Saucony, respectively. Figure 30 shows each participant’s change in the proportion of non-negative sentences (represented by the regression lines) toward Mizuno and Saucony. The average of the regression lines (in red) for all participants is also shown in the graph, indicating the overall change in the proportion of non-negative sentences.

Figure 29. Distributions of Paired Difference in Proportion of Non-Negative Sentences (After-Before)
The results of t-test and signed rank test are summarized in Table 13. Results of both tests suggest a significant difference in the proportion of non-negative sentences between the before and after interacting with brand representatives. The significant results are consistent for both Mizuno and Saucony.

Figure 30. Each Participant’s Change in Proportion of Non-Negative Sentences
Table 13. Paired T-test & Wilcoxon Signed Rank Test for Paired Difference in the Proportion of Non-Negative Sentences

<table>
<thead>
<tr>
<th>After -Before</th>
<th>Mizuno</th>
<th></th>
<th>Saucony</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>Statistic</td>
<td>p-Value</td>
<td>Test</td>
<td>Statistic</td>
</tr>
<tr>
<td>Student's</td>
<td>$t= 2.27247$</td>
<td>Pr $&gt;</td>
<td>t</td>
<td>=0.0293$</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>$S=71$</td>
<td>Pr $&gt;=</td>
<td>S</td>
<td>=0.0395$</td>
</tr>
</tbody>
</table>
Chapter 5 General Discussion

5.1 The Relationship between Individuals’ Social media Behaviors and the influence of their UGC

In the first study, two relationships proposed in H1 were tested in both purchasing and non-purchasing social media settings: 1) the relationship between the overall influence of an individual’s UGC and the individual’s engagement level on social media, and 2) the relationship between the overall influence of an individual’s UGC and the individual’s social status. In both purchasing social media context and non-purchasing social media context, the finding of this study shows that the overall influence of an individual’s UGC was significantly associated with the individual’s social engagement and social status in a log-linear fashion. In purchasing social media context, this result was supported by fitting the model on validation dataset. However, despite the statistical significance, the influence of engagement level and social status on the overall influence of an individual’s UGC is much smaller in non-purchasing settings. In non-purchasing social media context, even though similar estimates were given by fitting the model on validation dataset, model fit on the validation dataset was less satisfactory. Statistically, this lack of fit was due to the relatively small coefficient means and large coefficient variations. Namely, the variation of the model is larger in non-purchasing social media context than in purchasing social media context. Consequently, H 1.1 and H1.2 were supported by the research findings for both purchasing social media setting and non-purchasing social media setting. However, some practical issues are worth of further discussion.
The influence of social media behavioral characteristics could be explained by the source credibility and relationship ties in online social networks (Chu & Kim, 2011). Because of the community users’ high product involvement, such influence could also be explained by peripheral cues described in the Elaboration likelihood model (ELM), which suggests that consumers with low product involvement and are less capable of processing product knowledge tend to be influenced by some visible indicators such as rankings (Cho, 1999). Members in the purchasing and non-purchasing settings might be quite different in terms of their product involvement. For example, individuals on purchasing setting seek for high quality and credible reviews to decide purchase, while members for non-purchasing virtual communities are more or less random with diversified motivations in reading a posting (Hennig-Thurau, Gwinner, & Gremler, 2002). In other words, browsing product related information on non-purchasing social media is less goal-oriented than seeking information from others’ reviews in assisting in a purchase decision. Therefore, the credibility of the source may be less important in non-purchasing setting than in the purchasing setting. Additionally, unlike the consumer reviewers on Amazon.com, community members in Runnersworld.com could either provide opinions or seek opinions (Chu & Kim, 2011). Namely, the social media behavior in non-purchasing social media contexts represents more social interactions than credibility in providing product or brand knowledge. Therefore, the relationship between social media behavioral characteristics and social influence in virtual community is less obvious.

Given the research findings for H1, social status and social media engagement level provided in users’ profiles could be used to indicate the overall influence of an individual’s UGC. In
product review page on e-commerce websites or other purchasing social media contexts, one important managerial implication of the findings is that social CRM decision could be based on the prediction of the individuals’ social influence. Marketing practitioners need to recognize those influential reviewers via their social status and engagement level and implement marketing strategies specifically customized for them. Yet, in non-purchasing social media context, a community member’s community social status and engagement levels may not be sufficient for evaluating how much impact his or her postings could have on the brand knowledge diffusion.

5.2 The Relationship between UGC Contents and the influence of the UGC

In the second study, the objective was to identify the relationship between the content of UGC and the influence of UGC. The analysis in this study examined the influence of both original postings and replies contributed by other community members. Specifically, the following three relationships were tested: 1) the contents of the posting title and the number of views received by a posting; 2) the number of replies and the number of views received by a posting; and 3) the contents of the posting body and the number of replies the posting received.

Surprisingly, although brand names and product characteristics used in posting titles could significantly affect the posting’s total number of views, the results indicate that only small proportion of the influence of UGC could be explained. Even if all terms used in the posting titles were included, the model still can only explain the influence of UGC weakly. This
finding suggests that individuals’ viewing decision may not largely depend on the key terms used in the posting title. Other factors that were not measured in this study, such as underlying meaning and logic of how the postings are displayed, may also significantly affect the viewing decision.

Despite the weak explanatory power of contents in posting titles in explaining the influence of UGC, a significant log-linear relationship was found between the number of replies in UGC and the total number of views a posting received. That is, content co-creation from other members had a significant influence on the influence a UGC may have. Further analysis suggested that the high number of replies were affected strongly by the contents of the posting body, which is not visible until the viewer entered the posting page (after clicking on the posting link). Among all the terms used in the posting body, brand names and product characteristics were shown to be largely influential. That is, in product virtual community members’ replying behaviors was significantly triggered by the contents of the original posting.

The impact of brand names and product characteristics in the posting body on the number of replies was further explored. By clustering these keywords, nine clusters of brand names and five clusters of product characteristics were generated and used as predictors to model the number of replies for a UGC. Overall, these clusters show significant impact on the number of replies. Specifically, brand names in the most influential cluster, including ASICS, Brooks, Saucony, Mizuno and New Balance, are consistent with the top running shoe brands identified by Running USA’s National Runner Survey (2013). Despite having less overall
market share, these “professional running shoes” brands have higher influence on community members’ content co-creation than other clusters containing brands that dominate the sportswear market (e.g. Nike and Adidas). That is, active social media users may tend to share more opinions about specialized brands in a product category rather than overall famous brand names. So domain specific and professional brand images could be more welcomed for community members with higher engagement level. In terms of product characteristics related to running shoes, the following 5 categories were identified: fit, comfort, materials, price and fit issues. Out of these characteristics, relatively, fit and size issues had more significant influence on other members’ reply behavior.

In summary, H2.1, H2.2 and H 2.3 were supported by the findings in this study. The results suggest that the contents in the posting title had a significant influence on other members’ replying behavior, which in turn, significantly influenced more other members’ viewing of the original posting. However, while the contents in the posting title had a statistically significant influence on the number of views the posting received, the influence was relatively much weaker compared to the influence of number of replies.

5.3 The influence of Customized Brand Marketing Communication on Individual’s Attitude toward the Brand

In the third study, the objective was to investigate the influence of customized brand marketing communication on consumers’ attitude toward the brand, with the expectation of a positive influence. Paired t-test was performed to test the change of consumers’ attitude
toward the brand before and after the interaction with the brand representative via posting a question, viewing and replying to the messages posted by the brand representatives.

The findings indicate that while the magnitude of an individual’s averaged sentiment was not significantly increased, the proportion of negative sentiment in a member’s postings was significantly reduced. As previous suggested, since eWOM behaviors are primarily motivated for expressing negative concerns regarding a product, community members are less willing to post strong positive opinions in social media (J. Lee et al., 2008). However, high satisfaction doesn’t necessarily lead to large increase in positive brand sentiment in the postings (Hennig-Thurau, Gwinner, & Gremler, 2002). Therefore, for unsatisfied consumers, customized brand marketing communication may help to change their attitude from negative to non-negative (either neutral or positive). However, for consumers who already have positive attitude (satisfied consumers), even the customized brand marketing communication may help to improve the satisfaction level, this change will not be manifested via their expressed sentiment in their postings. This may explain why the magnitude of sentiment is significantly increased while the proportion of non-negative sentiments significantly increases after the interaction. Therefore, it is fair to conclude to a major function of companies’ participation in member hosted virtual communities is to convert those unhappy potential consumers to positive attitude holders, hence to avoid the further effect of the negative eWOM of those unhappy consumers.

Overall, these findings provide great evidence on the helpfulness of social media marketing communication in reducing community members’ negative feelings by honestly discussing
product related problems. In social CRM program, the key role of brand representatives in a product forum could be effectively helping unsatisfied customers instead of directly promoting the brand. Strategically, marketing communication on social media could be used as complementary strategy in addition to other marketing programs aiming at establishing brand image. Consequently, Hypothesis 3 was partially supported by the findings.
Chapter 6 Conclusions and Limitations

6.1 Theoretical and Managerial Contribution

In general, this present study explores an important strategic question of how to leverage online consumers’ social influence into brand marketing. Based on the Brand Value Chain model, a conceptual strategic model for brand marketing was proposed by integrating marketing communication, customer relationship building, brand knowledge diffusion, Social CRM and CBBE. The conceptual integration of these marketing fields could provide great theoretical guidance to companies in their systematical implementation of brand marketing campaign from various media channels. Meanwhile, the relationship management concentrating on social influence would also help companies to efficiently recognize the importance of the voice of consumers and develop consumer centered marketing communications and market programs.

To test the hypotheses proposed in the conceptual model, observational data, from both purchasing and non-purchasing social media settings, was collected and three empirical studies were performed. These hypotheses tested in this research addressed the following three relationships: 1) the relationship between individual’s social media behavioral characteristics and the influence of their UGC, 2) the relationship between content of a UGC and the influence of the UGC, and 3) the effectiveness of customized brand marketing communication on influencing individuals’ attitude toward the brand. Consequently, H1 and H2 were supported by the findings of the empirical studies and H3 was partially supported.
This study merits several theoretical contributions to the field of brand marketing. First, this research expands the scope of brand marketing by integrating several marketing concepts into one united brand marketing framework with concentration on the role of social CRM. Despite the preliminary discussions in existing literatures about incorporating consumers’ social value into CRM strategy (e.g. Baird & Parasnis, 2011b and Kumar & Reinartz, 2012), the proposed model in this study would fill the gap due to the lack of integration with brand marketing. Second, using observational data, instead of information solicited from consumers, this research provides empirical support to existing literature on eWOM behavior and social CRM. For instance, some studies (e.g. Doh & Hwang, 2009) attribute consumer viewers’ responses solely to the verbal and semantic feature of the UGC. Yet, many argue that content co-creation may also contribute to consumers’ participation in marketing programs (e.g. Hargittai & Walejko, 2008). Findings in this study would provide empirical support to these arguments and suggest their viewing behaviors depends more on the amount of collaboration in the content rather than the interestingness of content itself. Third, this study also contributes several pieces of new knowledge on behavioral characteristics of eWOM in purchasing and non-purchasing social media settings. For example, this study suggests a positive relationship between the overall influence of UGC and individuals’ social media behavioral characteristics, recognizing that the strength of the relationship depends on different social media settings. The usefulness of social media marketing and marketing intervention on social media has been largely discussed, but has yet to be justified with empirical evidence (Stead et al., 2007). So another piece of new knowledge contributed
from this study is the finding on the function of social media marketing communication in reducing consumers’ negative feelings toward their brands.

Methodologically, a set of latest information and data analysis techniques were identified and employed in this study, including automated data collection using Web Scraping, text mining, and sentiment analysis. This study also provides an alternative approach to measure the impact of social media marketing, in addition to traditional approaches such as attitudinal scale measurements used in survey. Overcoming the drawbacks associated with survey, automated data collection and analysis has much potential in the social media age as abundant behavioral data is available and information is updating faster than ever. The text mining techniques adopted in this study proved to be working and can be applied for other brand marketing researches, such as perceived brand poisoning. The perception of the positions of brands in a market can be seen as a social network, where each of the brands has certain possibility to be compared with each other by consumers, and the co-mentions of brand names can be used as a measure of brand similarity (Netzer, Feldman, Goldenberg, & Fresko, 2012). The exploration of perceived brand poisoning is beyond the scope of this study and some examples of brand poisoning map based on the sample data are provided in the appendix. In general, while being exploratory, this study provides great inspiration for future studies dealing with large unstructured textual data, particularly from social media.

Managerially, by incorporating relationship building processes and integrating eWOM management into brand marketing for different marketing channels, the proposed strategic model provides a useful framework for business practitioners. From the perspective of
strategy development in brand marketing, the present study argues the importance and benefit for companies to include consumers’ social influence in the marketing strategy. When designing the brand loyalty programs, for instance, more weight would be given to the individual’s predicted social influence in the distribution of marketing incentives (e.g. customized coupon offers, promotions and benefits to loyalty card holders) in addition to the existing customer value measurements, such as Recency, Frequency and Monetary (RFM) method (Kumar & Reinartz, 2012).

From the perspective of social CRM implementation, the findings of this study also provide great implications for brand marketing practitioners in their understanding and serving of their target consumers at the age of social media. First, the individual’s social behavioral information portrayed in their profiles on the social media would be helpful for companies to target opinion leaders in purchasing social media setting. Second, in non-purchasing setting, it can be very inefficient for companies to target individuals only according to the profiles shown on the forum. One important factor to consider is an individual’s capability to generate content co-creation activities. Domain specific popular brands and personalized product experience may suggest higher content co-creation from active users. Lastly, this study suggests that marketing communication on social media could be an effective way to reduce negative impact on the brand but not so effective in promoting brand. Therefore, marketing communication on social media could be used as complementary brand marketing stagey in addition to other traditional marketing strategies. In summary, to increase brand marketing performance, the proposed brand marketing model could be a useful strategic
roadmap. The implications suggested by the empirical findings would provide useful
guidance in strategy implementation

6.2 Limitations and Suggestions for Future Study

While the findings of this study provide meaningful contributions to both the literature and practitioners for brand marketing, few limitations exist in this study. The first one is the limited sources where the sample data were collected from due to technological difficulties and time limit. Random sampling from a variety of social media websites would increase the representativeness of the sample, which could be done in future studies. The second limitation comes from the drawbacks of numerical representation of human language. While bag-of-words representation can be used to analyze a large collection of consumer reviews or postings efficiently and quickly, the true meanings behind those words could be far beyond the capability of this methodology. For instance, grammar and context are not captured using keywords and sentiment. So, numerically generalized data from the texts could be only considered as a simplified version. Thirdly, the sentiment change due to marketing communication on social media was tested based on convenient samples with limited size, which could be a limitation in many social sciences studies. Also, there could be reasons other than the marketing communication that causes the change in the members’ attitude. Therefore, the effect of company-generated posting threads on consumers’ brand sentiment is relatively less robust. Therefore, future studies on marketing communication on social media and consumers’ brand attitude need to validate the results with larger sample sizes. In
conclusion, the present study aims at establishing a strategic brand marketing model and providing empirical support, but the research design is exploratory in nature.
References

YOgPqdb5M_3L4RT8s


http://doi.org/10.1108/10878571111176600


http://doi.org/10.1108/10878571111161507


http://doi.org/10.2307/2555626


Appendices
Appendix A - Example Data Analysis Flow Chart in SAS Enterprise Miner
Appendix B - Examples of Brand Position Map Obtained from the Data
Appendix C- Python Code Example: Web Scraping Data

# Obtain data from RW
import urllib
import re

# get level 1 and level 2 data into file
data=open('file.txt','w')
for i in range(1,71):
    url= "http://community.runnersworld.com/forum/shoes?page="+str(i)
    htmlfile=urllib.urlopen(url)
    htmltext=htmlfile.read()

    reply=re.findall('(?:<div class="smallfont
forumTopicStats">\n\s*\n(?:replies|reply)',htmltext)
    view=re.findall('(?:·\n\s*\n(?:views</div>\n\s*)',htmltext)
    title=re.findall('(?:<div class="medbold forumTopicSubject">\n\s*<a href="http://community.runnersworld.com/topic/)(.*)(?:"\n\s*>.*</a>)',htmltext)
    ID=re.findall('(?:<ahref=")(http://community.runnersworld.com/profile/.*)(?:"\n\s*>)',htmltext)
    time=re.findall('(?:<td class="datetimeColumn">\n\s*\n</td>)',htmltext)

    #user profile data
    user_rank=[]
    user_activity=[]
    content=[]
    for k in range(len(ID)):
        url_user=str(ID[k])
        hfile_user=urllib.urlopen(url_user)
        htext_user=hfile_user.read()
        rank=re.findall('(?:<div
id="communityRankPointsSection">\n\s*#\n</div>)',htext_user)
        act=re.findall('(?:<div id="activityPointsPointsSection"\nclass="divCell">\n\s*\n</div>)',htext_user)
        user_rank.append(rank)
        user_activity.append(act)

    #level 2 content data
    url_l2="http://community.runnersworld.com/topic/"+str(title[k])
    htmlfile_l2=urllib.urlopen(url_l2)
    htmltext_l2=htmlfile_l2.read()
    postings="".join(re.findall('(?:forumTopicMessageBody">\n\n?.*\n</body>)',htmltext_l2))
    content.append(cleanhtml(postings))
    print "User profile"+str(k)+" obtained"
for j in range(len(reply)):
    data.write(str(time[j]))
data.write("|")
data.write(str(ID[j]))
data.write("|")
data.write(str(reply[j]))
data.write("|")
data.write(str(view[j]))
data.write("|")
data.write(str(title[j]))
data.write("|")
data.write(str(content[j]))
data.write("|")
data.write(str(user_rank[j]))
data.write("|")
data.write(str(user_activity[j]))
data.write('|
')
print "Number"+str(j)+" obs wrote into file"

    print "topic page "+str(i)+"/70 complete"
data.close()
print "scraping process complete"

#############################################################