CHURCH JR., EARNIE MITCHELL, Ph.D. Curation-Based Network Marketing: Strategies for Network Growth and Electronic Word-of-Mouth Diffusion. (2013) Directed by Dr. Lakshmi Iyer and Dr. Xia Zhao. 189 pp.

In the last couple of years, a new aspect of online social networking has emerged, in which the strength of social network connections is based not on social ties but mutually shared interests. This dissertation studies these *curation-based* online social networks (CBN) and their suitability for the diffusion of electronic word-of-mouth information (eWOM). Within CBN, users do not rely on profiles full of personal information to identify network “friends”. Rather, CBN users curate collections of digital content that becomes their digital self-expression within the network. This digital content can then be viewed, commented on, and shared across the pages of other CBN users. As the dissertation will show, this process of digital content curation, a relatively new online practice that centers around the collection and sharing of rich digital media, builds CBN, and presents exciting opportunities for the study of eWOM. The dissertation presents three studies around digital content curation, CBN, and eWOM diffusion.

Study 1 examines individual level antecedents of digital content curation behavior. In this study, we use theory from sociology and behavioral psychology to develop a model of user intentions towards digital content curation behavior. We find that digital content curation is comprised of a mixture of social and utilitarian motivations, and that the management and organization of digital content is a major reason that people spend time on CBN.

Study 2 examines the way that digital content curation behaviors grow CBN. We study a sample of 1800 CBN users to determine the way that their digital content curation behaviors attract and retain interested CBN followers. We find that the most successful CBN users are those that can generate an eWOM response around their content

collections. Additionally, we find that textual eWOM plays a very limited role in attracting followers in the CBN environment. Finally, Study 3 examines eWOM diffusion by analyzing data on the structure and diffusion of digital content through real-world CBN network structures. This descriptive analysis of eWOM in CBN presents details on the way that CBN data is structured, and the methods and techniques that can be used to collect and analyze real-world eWOM collected from a CBN site. The study uses the UCINET network visualization software package to examine the networks of thirty companies operating CBN pages. Using a unique data set specifically compiled for this study, we are able to visualize the diffusion of curated digital content through the networks of these companies, and show how companies can identify their most influential followers as targets for further eWOM and traditional marketing efforts.

Together, the three dissertation studies offer a holistic view of content curation behavior and curation-based online social networking and has the potential to fill the gap in the literature on information diffusion and online marketing. We make substantial contributions to the areas of sociology, economics, and marketing, and offer one of the first treatments of the role of digital content curation in online social networks.

CURATION-BASED NETWORK MARKETING: STRATEGIES FOR NETWORK GROWTH AND ELECTRONIC WORD-OF-MOUTH DIFFUSION

by

Earnie Mitchell Church Jr.

A Dissertation Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2013

Approved by

____
Committee Co-Chair

____
Committee Co-Chair

To Natalie, Daniel and Evelyn; the next generation.

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Co-Chair ______________________________
Lakshmi Iyer

Committee Co-Chair ______________________________
Xia Zhao

Committee Members ______________________________
Dora Gicheva

____
Jing Deng

____
Date of Acceptance by Committee

____
Date of Final Oral Examination

ACKNOWLEDGMENTS

I would like to express my appreciation for my two dissertation co-chairs, Drs. Lakshmi Iyer and Xia Zhao, for their patience and willingness to impart their research expertise and guidance along each step of this process. I would also like to thank the other two members of my committee, Drs. Jing Deng and Dora Gicheva, each of whom contributed invaluable advice without which this project would not have been possible. Finally, I would like to thank the extremely patient and helpful user communities of StackExchange and Scrapy-Users, for answering what must have been an overwhelming number of questions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Overview of Dissertation

The rise in popularity of online social networks represents one of the major trends in computing over the last several years (Parameswaran and Whinston, 2007). Online social networks are an extension of internet messaging and email technologies that allow individuals to communicate with and keep track of large numbers of people. Originally, the most popular of these networks (eg. MySpace, Facebook) were designed to allow for communication between friends and people who shared some manner of social relationship (Ellison et al., 2007b). As online social technology has diffused in both popularity and prominence, it has been applied to a number of different settings. To accommodate the specific requirements of each application, online social network technologies have been adapted to the needs of many niche markets and user segments. For example, businesses have adopted social networking technologies to aid in knowledge transfer and information sharing (Wasko and Faraj, 2005). Companies like LinkedIn have developed social networking sites designed to allow business executives to meet and network.

In many of these applications, the general aspects of the social network technology remain the same. Social network sites allow people to post messages, locate other individuals and enter into “friend” or “following” relationships. Social networks allow users to keep track of their friends easily, and share pictures and other media with those other network users. Typically, social network sites allow for selective disclosure of information, so that more information is shared with individuals who are not inside the person's friend network.

What changes from one social network site to the next is not the fundamental actions that make up social network user behavior, but rather the context used to connect one user to another. The context is what determines the theme of the entire network. For example, Facebook creates networks of friends and peers, whereas LinkedIn creates networks of business acquaintances and coworkers. This dissertation studies a new online social network context, in which the criteria for link formation is based not on social or business connections, but rather mutually shared interest. Within these networks, users do not rely on profiles full of personal information to identify network connections. Rather, users curate collections of digital content that become their digital self-expression within the network. This dissertation studies these online social networks, which we dub "curation-based".

Understanding the "curation-based" online social network context has a number of important implications, and these form the primary motivations for this dissertation. First, nearly all online social network sites earn revenue through advertising. Advertising activities in these sites consists of traditional marketing functions (eg. advertising) as well as advertising designed to be spread from one user to another via electronic word-of-mouth (eWOM). eWOM in particular is seen as a major area of revenue growth for online social network sites.

As an example of the revenue growth potential of eWOM marketing, consider that as Facebook prepared for its IPO earlier this year, many analysts estimated that the company could achieve an initial valuation of over $100 billion (Miller, 2012) based largely on its potential to capitalize on word-of-mouth and other types of social advertising (Miller, 2012). However, despite the intense amount of media attention leading up to Facebook's huge IPO, the company has failed to realize investor expectations [1]. Almost immediately after going public, Facebook's market capitalization, along with the value of many other social media companies, began to sharply decline.

There may be good reason to investors and industry practitioners to be wary of investing in eWOM advertising. Effective eWOM is based on identifying people that influence your commercial tastes and purchases (Susarla et al., 2012a; Trusov et al., 2009). However, thus far online social networks have proven ineffective in doing just this. It increasingly appears that the marketing interest around online social networks may be built on a faulty assumption: that your friends and social connections drive, or even share, your commercial tastes and interests.

The unique context of curation-based networks makes them worthy of in-depth and thorough investigation. Within curation-based online social networks (CBN), people are not linked together using any traditional means common to many of today's modern social network sites. They are not necessarily linked to friends, coworkers, or acquaintances. Rather, people within these sites *curate* digital content, and then form networks around this content, which servers as an abstract representation of themselves, their interests, and their affinities within the online social network. Curation-based networks exist and grow though a process of *digital content curation*, which involves the careful collection, maintenance and management of a set of digital assets (Yakel, 2007). In curation-based networks, it is this content that becomes the primary means of user self-expression. CBN are therefore networks of *what you like*, not *who you are*.

[1] http://www.technologyreview.com/news/427972/the-facebook-fallacy/

Because CBN are comprised of content related to user's interests, they often are full of pictures and media related to products and services. Many interests and hobbies cannot be easily separated from the commercial products that accompany them. For example, a chef needs knives and ingredients to cook, just as a painter needs brushes and paints to create art. Because these networks are literally built from content pertaining to commercial products and services, they provide a natural context for eWOM advertising. Such a context would set these online social networks apart from all the other iterations of this technology, and may provide exactly the kind of eWOM platform that marketers have envisioned since the earliest mention of eWOM. We study these CBN and the digital content curation process that they facilitate. The dissertation examines the marketing potential of these networks. We are interested in addressing three primary questions:

(1) What are the antecedents that drive people to engage in digital content curation?

(2) What is the impact of the electronic word-of-mouth communication created by digital content curation within a CBN?

(3) Finally, how can companies better understand the impact and spread of CBN content in order to be successful with CBN marketing?

Interestingly, despite the important implications of CBN, and the digital content curation process that they make possible, the IS literature has so far ignored the implications of CBN. This dissertation represents a first attempt to bring CBN to the forefront of the IS discipline, and promote the unique and potentially wide-reaching implications of this new and exciting technology. In the next sections, we discuss digital content curation in greater detail, before presenting some examples of CBN that incorporate digital content curation as an integral part of their user experience.

1.2 Defining Digital Content Curation

At the center of CBN is the process of digital content curation. Yakel (2007) defines digital content collection as the collection and maintenance of a carefully chosen set of digital assets. Today, many websites, including some online social networks, allow for digital content curation to some degree. For example, twitter allows users to embed pictures into their tweets, and users of sites like tumblr ofter build images and videos into their blog posts. This trend is expected to continue as rich media sharing continues to become a major part of online digital communication (Counts and Fellheimer, 2004; Sumi et al., 2008). Additionally, the rise of mobile communications technologies has made it easier to tweet a picture than type a large block of text (Counts and Fellheimer, 2004; Van Audenhove et al., 2007).

Scholarly work in the area of digital content curation originated in the early part of this century. It was originally conceived in response to the rapidly changing and fragile nature of the early internet. To date, most digital content curation has been undertaken by institutions and libraries as a means of preserving digital knowledge for future generations (Dallas, 2007). Guy et al. (2009), in writing on the spirit of this endeavor, put it thusly:

> The Web runs at risk. Our generation has witnessed a revolution in human communications on a trajectory similar to that of the origins of the written word and language itself. Early Web pages have an historical importance comparable with prehistoric cave paintings or proto-historic pressed clay ciphers. They are just as fragile…

Technological innovations have made it easy to collect and distribute digital content, so that today not only organizations and libraries, but also individuals maintain collections of digital assets. Specifically, recent technological advances in the area of rich media sharing have made the distribution of digital content across online social network sites easy and inexpensive (Guy et al., 2009). It is exactly this functionality that we use to define a CBN *as an online social network that allows for the collection and management of digital content as a major part of the network's core functionality.*

While digital content curation is interesting and worthy of exploration in its own right, we are primarily motivated to examine digital content curation as a new form of electronic word-of-mouth communication. eWOM is defined as positive or negative statements about a product, media personality, or company, made widely available via the internet (Thorson and Rodgers, 2006). This definition includes a large amount of electronic communication. Twitter feeds, Facebook posts, online customer reviews, as well as images and YouTube videos can all be considered forms of eWOM. Because eWOM is relatively new, and new opportunities for the study of eWOM are appearing at a rapid pace, multiple definitions and terms have been used in the extant research (Cheung and Thadani, 2012; Vilpponen et al., 2006). Viral marketing, online feedback mechanisms, online customer reviews, word-of-mouth advertising, online referral systems, peer endorsement systems, and word-of-mouse can all be considered as subsets of the larger category of eWOM (Vilpponen et al., 2006).

eWOM has been tied to a number of desireable marketing outcomes. eWOM can lead to increased sales (Chevalier and Mayzlin, 2003). Customers acquired through word-of-mouth means are typically more loyal and often more profitable than other customers (Trusov et al., 2009). Positive eWOM can help YouTube channels acquire and keep new members (Susarla et al., 2012b) and influence people's buying behavior online (Garg et al., 2011). Despite the positive benefits of eWOM, few strategies for ensuring eWOM success exist (Trusov et al., 2009). Pinterest shows record amounts of eWOM referral business, while at the same time many other online social networks like Facebook, MySpace and LinkedIn struggle to find ways to monetize eWOM communication. This has created a real challenge for online social network sites, who see eWOM advertising as a major source of future revenue. There is some evidence, however, that the difficulties around eWOM within some online social networks may not be related to the nature of eWOM, but instead to the actual way that peer-based online social networks, or networks of friends, family, and co-workers, are structured (Agarwal et al., 2009).

Effective WOM advertising is based on identifying people that influence your commercial tastes and purchases (Susarla et al., 2012a; Trusov et al., 2009). A recent paper by (Agarwal et al., 2009) examined whether it was possible to identify user interests within the peer-based network Facebook by looking at the users in the network around them. What they found was that social peers are not particularly useful for predicting interests, in that they do not necessarily share interests in any predictable way. Lewis et al. (2012) made a similar finding in their study of Facebook, namely that social peers often carry little weight in terms of influencing what someone will buy. Additional studies have reported the same. Friends are not the deciding factor in what tv shows we watch (Susarla et al., 2012a), or the music we listen to (Garg et al., 2011).

Digital content curation provides network users with digital content that they use as the basis for link formation. It is the digital content itself, not the people in the network, that becomes the context and topic of network activities and user communication. CBN are therefore networks of *what you like*, not *who you are*. Figure 1 shows an example of how curation-based networks are structurally different from peer-based. In a traditional peer-based network, the individual user forms the context for link formation. Thus, close nodes are those that share a large number of social connections.

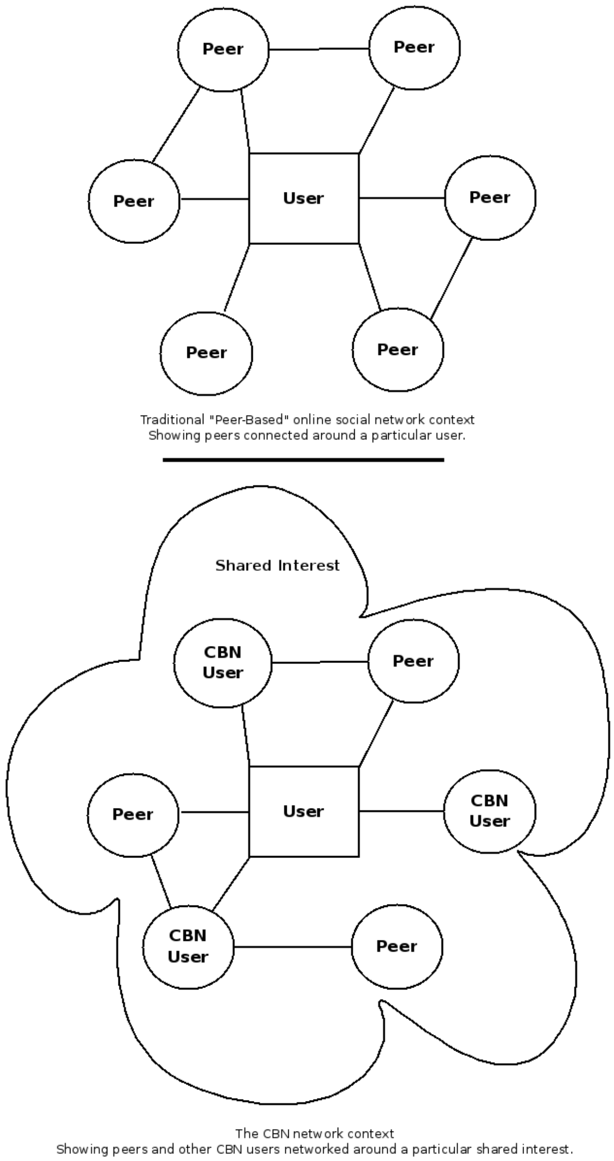

Figure 1. Curation-based and “peer-to-peer” social network contexts.

Within a curation-based network, close network nodes are people that share a large number of connections with people with similar *interests*. Additionally, curation-based networks can have network nodes that have no social connection to a particular user, but are only present in the network because of a shared interest. This has potentially profound implications for the spread of eWOM in these networks, and may account for the observed success of Pinterest and other CBN in driving eWOM referrals.

In the next sections, we present some examples of CBN, and discuss the differences between traditional online social networks and networks that allow for and incorporate digital content curation technologies. Next, we provide a bit of background on the history of online social networks to emphasize the gradual trend towards the incorporation of rich media content into these environments. This is done to show that digital content curation is not a completely new phenomenon, but rather the obvious result of a progressive trend towards rich media sharing exhibited by online social network sites over the last ten years.

1.3 Examples of Content Curation in Social Networks

Here we present some examples of existing CBN in order to show the unique nature of data presentation within these sites. The examples come from Pinterest, Facebook's CBN offering "PinView", and from Gentlemint, another site similar in style to Pinterest. The Figures presented here represent the home pages of typical site users. As can be seen from the Figures, these pages feature little in the way of textual information, or information pertaining to the identity or personal life of the user. In fact, even profile pictures, which have been studied so much in the social network literature (Ellison et al., 2006) are absent. Instead, the pages are dominated by digital content that has been collected by the user in this space by means of digital content curation.

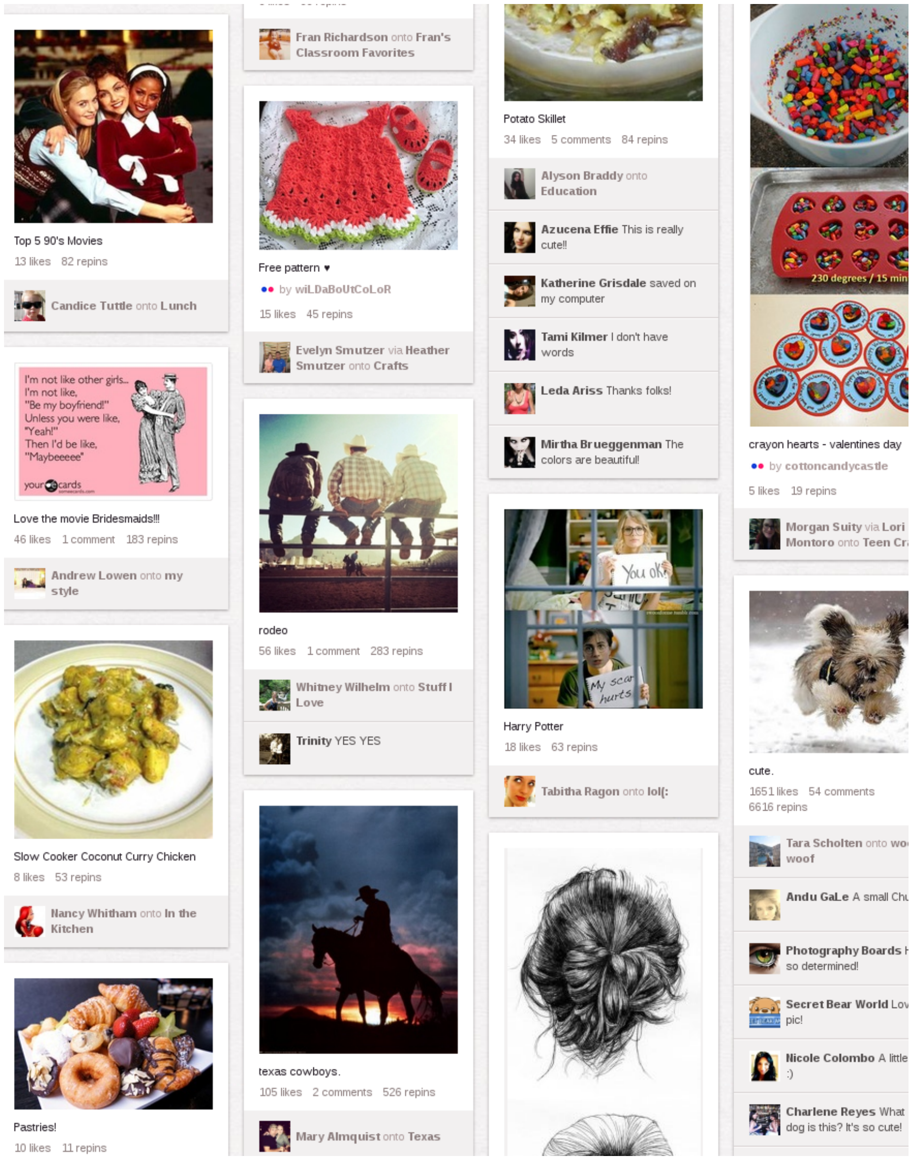

Figure 2. Pinterest's content-based social networking format

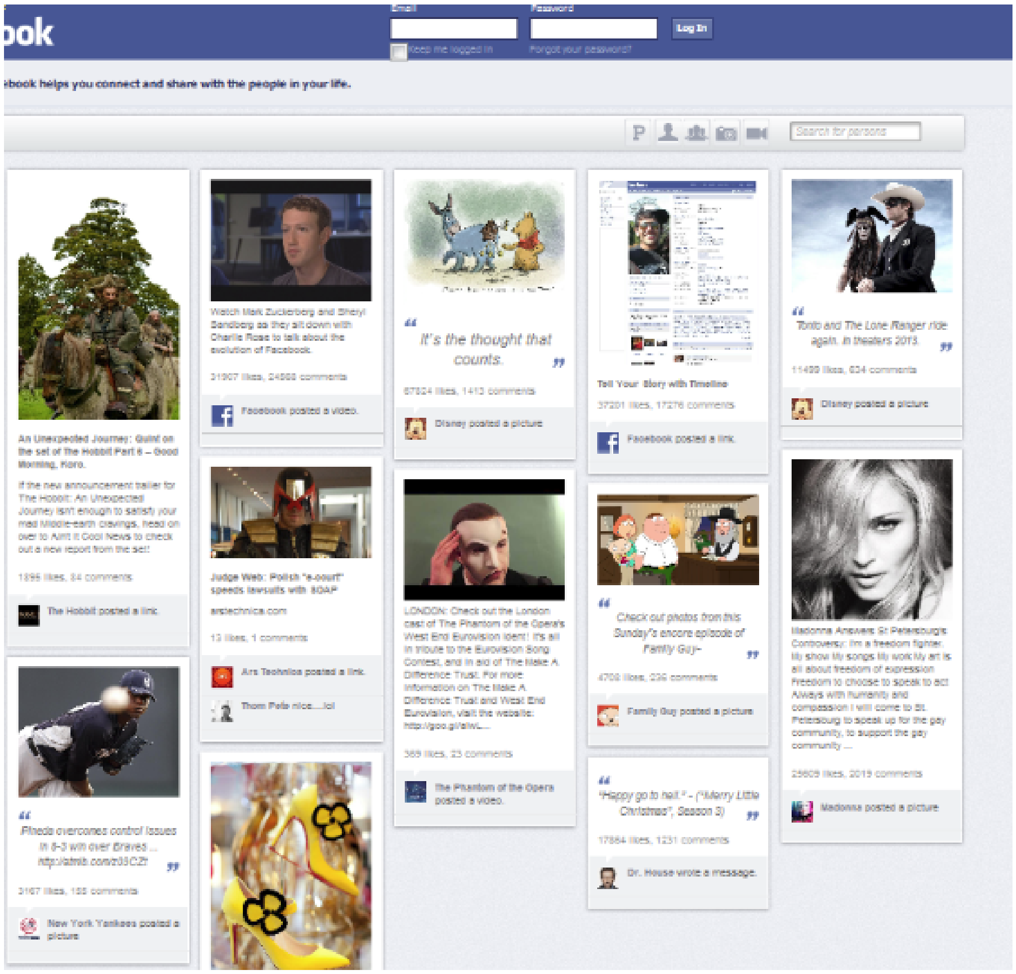

Figure 3. The recently implemented “Pinview” Facebook format.

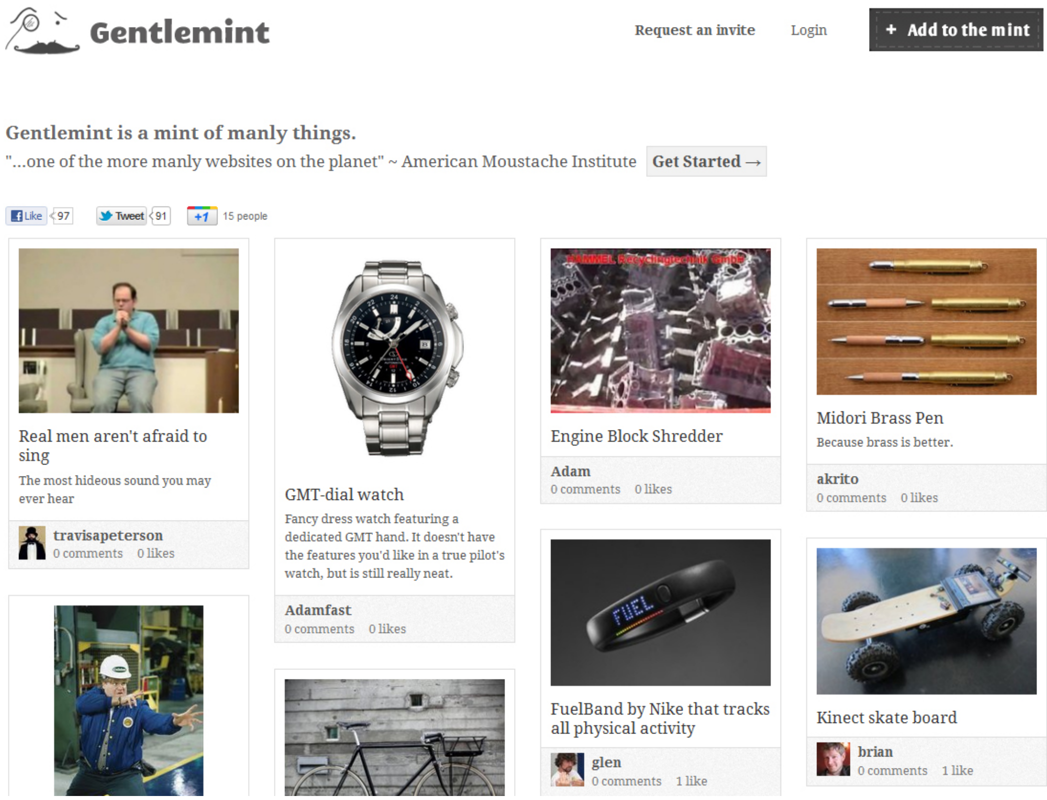

Figure 4. A similar presentation format from Gentlemint.

Curation-based networks represent a new context in online social networking, yet they are the result of a systematic progression of small changes in two areas:

(1) Online social networking has embraced improvements in information technology and networking to allow for ever-richer modes of communication

(2) Users of online social networks are trending towards smaller networks that allow for targeted communication around selected topics, and moving away from large networks that may be unwieldy or present uncertain privacy implications.

Some historical background on online social networks explains these trends. Since their inception, online social networks have seen steady growth in popularity. Today, online social networking represents one of the major uses of internet technology (Ellison et al., 2007b). Such rapid growth rarely comes without challenges and opportunities (Nault and Vandenbosch, 2000). The popularity and business potential of online social networks has resulted in many companies attempting to enter the industry CITE. These companies bring with them new technologies and core competencies. One area in which many online social networks have sought to create competitive advantage is through the use of rich media (eg. photos, videos, music, etc) (Choudhury et al., 2009; Ellison et al., 2007b). As a result, users of sites like Facebook, Twitter, and MySpace are now able to share rich media including images and videos (Safko, 2010). These richer media formats have become a large part of online social network usage, with companies working to develop applications and even complete social network sites designed to promote these forms of media as primary means of communication (Evans and McKee, 2010). For example, the recent example of Instagram, an online social network site for sharing pictures was acquired by Facebook for $2 billion, when the company was less that two years old [2].

The trend towards ever-richer forms of media has resulted in many online social networks now mediating at least some degree of content curation behavior. For example, Facebook allows users to save photos and videos within their user pages. The popular microblogging network Twitter limits “tweets” to 150 characters in length, but many people use third party applications to add photos and videos to their tweets. Youtube is built around uploading and sharing videos. Often, these videos drive community discussion within the network, so that the rich media content becomes the driver of social interaction (Susarla et al., 2012a,b).

On some social networks, textual discourse has been supplanted to a large degree by communication via digital content. This dissertation focuses on one of these networks, Pinterest (Carr, 2012). Pinterest has been around for about two years, but only recently has the site come to prominence. While growing quietly since its inception, the popularity of the site has since exploded, leading to Pinterest earning the distinction of becoming the fastest web site to reach 10 million unique users [3]. A recent report found that Pinterest ranked third among social network sites in terms of number of US visitors, behind only Facebook and Twitter [4]. Of the technologies discussed thus far, Pinterest exhibits the highest degree of content curation. Within Pinterest, users do not create profiles full of personal information. Rather, Pinterest pages consist of a set of digital content that

[2] urlhttp://www.forbes.com/sites/forbesleadershipforum/2012/04/25/will-instagram-be-a-billion-dollar-loss-for-facebook/

[3] from `http://tech.fortune.cnn.com/2012/03/22/`

[4] `http://www.cnn.com/2012/04/06/tech/social-media/pinterest-third-social-network/index.html`

becomes the digital self-expression of the user within the network. This digital content can then be viewed, commented on, or even copied onto the pages of other users (Carr, 2012).

Another online social networking trend over the last few years has been towards the desire for more tightly controlled networks (Binder et al., 2012). In this sense, tighter control refers to the ability to direct communications to different subsets of the larger, complete network. Their are several reasons for this trend. First, online social networking has seen a fair share of concern over issues of privacy and controlled information sharing (Debatin et al., 2009). Online social networks have a tendency to grow over time, leading to network structures that some users may find unwieldy or too open . Personal information that a user wants to share with friends may be inappropriate for grandparents, for example. Additionally, some aspects of a person's life may be irrelevant or inappropriate for portions of the person's social network. In order to accommodate these aspects of usage, online social networking sites have implemented features designed to allow for creating network *subsets*. When Google rolled out its social network offering, Google+, the ability to create circles, or network subsets that are distinct and separate from a user's larger social network, was a major selling point of the technology. Different social networking platforms allow for network subsets to varying degrees. Facebook networks have always been built based on social connections. However, when network connections are based on social connections, network paths will always cross into various aspects of social life, sometimes resulting in unwanted information sharing.

Unlike traditional online social networks, curation-based social networks are built around digital content, and network connections in these systems are formed when users share an interest in the same or similar content collections. Whether these users happen to also share some social connection is irrelevant to the process of network formation.

Because curation-based social network profiles are not built on personal characterists, content curation represents the only real way of forming network connections. In traditional online social networks, users create extensive profiles that list hobbies, interests, and affinities together with numerous personal and demographic information related to their age, gender, work history, etc. Curation-based social network users do not maintain profiles, but instead build collections of content to represent themselves within the network community.

Curation-based networking is new enough that little is understood about its potential or the impact that the technology may have on information technology and society at large. Although academic writing on the topic is scant, several potential issues concerning the process of content curation have been identified. On account of these issues, greater understanding of the nature of curation networks has several significant implications for both practitioners and academics.

Dellarocas et al. (2013), in an upcoming paper, identifies the changing relationship of content producers and content aggregators in today's *link economy* (Jarvis, 2008). Content producers are companies that produce digital content for sale or distribution online. The New York times would be an example of a content producer. Content aggregators pull this content from all over the Internet and display it in one place (eg. boingboing.com). These content producers and aggregators all compete for attention and visitors in the link economy, which refers to the web of hyperlink connections that bind together the internet (Jarvis, 2008). These connections are valuable, in that without them users would have little chance of navigating the highly unstructured and nebulous world of cyberspace. The entire search engine industry is built on maintaining an index of online content and pointing users in the direction they want to go (Osnos, 2009).

As Dellarocas et al. (2013) discuss, there is an ongoing debate around property rights and revenue generation in these types of interactions. Content producers populate the internet with huge amounts of interesting and original content, and without them the internet would be a pretty boring place. Content aggregators also provide a valuable service, and likely drive more traffic to the sites of content producers than they could ever obtain on their own (Chowdhury and Landoni, 2006; Jarvis, 2008). Typically, advertising revenues often go to websites that are able to generate the most traffic. Each person an aggregator sends to a producer's site also sees ads in both places. For this reason, this is an area where content aggregators often do extremely well (Van Audenhove et al., 2007). For example, the blog boingboing.com generates more traffic than the New York Times by aggregating content from the New York Times and other similar news sites[5].

Today, there are relatively few content aggregators compared to the large number of companies producing online content. Curation-based networks may fundamentally change this, resulting in profound repercussions for the internet advertising industry. Because curation-based networks put the tools for content collection and aggregation into the hands of every single network user, they have the potential to exponentially increase the aggregator population.

The situation is similar to the industry-shaping impact felt by Napster back in 1999. Before Napster, all the tools were available for people with access to the internet to share music and files. However, these tools were difficult to use and finding people with the files that you wanted was difficult (McCourt and Burkart, 2003). Napster did not make it possibly to share files, it made that filesharing incredibly easy. As a result, illegal filesharing changed from an annoyance to an industry-redefining phenomenon (Hong,

[5] `http://www.impactbnd.com/content-curation-inbound-marketing-imperative/`

2013; McCourt and Burkart, 2003). Curation-based networks provide a set of tools that make content aggregation easy. Many of these systems have tools that let users import content from other websites into their collections with the click of a browser button. Once the content is in the network, it can be viewed and distributed to any number of users, all without ever leaving the network system. Many of these sites, including Pinterest, do preserve the origin of the content, and provide links to the originating site, though visiting these outside sites is never required.

Curation-based networking has another interesting implication. As users aggregate content into these networks, they are bringing together content spread all over the web and centering it within relatively few domains. Over time, as users come to rely on their collections, and the content they can find within these networks, the curation-based networks become hubs for all subsequent internet travel. This could potentially have the effect of making the internet a very fragile system. If these hubs were compromised through denial-of-service attacks or other means, users that come to rely on them for internet navigation would be effectively shut off from internet content. The content itself may sit safely on the site of the original content producer, but knowledge of how to get to it would be lost (Seneca, 2009).

All of this research helps to make the case for why the practice of digital content curation is worthy of study. However, to date digital content curation, especially digital content curation outside of museum and academic settings, has received very little attention (Guy et al., 2009; Yakel, 2007). As a result, some basic questions about the nature of digital content curation within online social networks have yet to be answered. The next section presents an overview of three studies that comprise the bulk of this dissertation. Together, these studies are meant to provide a holistic overview of curation-based network technologies, its challenges, and implications. We also seek to lay a groundwork for future research into this new and exciting area of online social network technology.

1.4 Three Studies Around CBN and eWOM

The dissertation makes contributions to the existing online social network research, as well as current research in the area of eWOM. Three studies are presented that examine digital content curation, its antecedents and impacts, under a number of different settings and conditions 5. Study one starts by examining the antecedents of digital content curation. Because the phenomenon of digital content curation is so new, extant research that studies digital content curation is rare Guy et al. (2009), and research that looks at digital content curation in the context of an online social network is essentially nonexistent. Study one makes a first attempt at addressing this gap in research. The study incorporates research and theory from the areas of personal information management (Jones et al., 2001, 2002) and sociology to present digital content curation as an activity that 1) provides utility to the curator, and 2) has social implications through its association with online social networks. The study presents the results of a survey-based examination of active CBN users. Notable findings include support for the dual nature of digital content curation as a social and utilitarian process, as well as strong support for the role of content organization and management in driving digital content curation behaviors and CBN usage.

The ultimate goal of this first study is to explain the individual antecedents that drive digital content curation behaviors. Understanding these antecedents is important, given that CBN are fundamentally based around digital content, and thus they cannot exist without people engaging in digital content curation. For this reason, the first study begins with the behavioral antecedents of content curation.

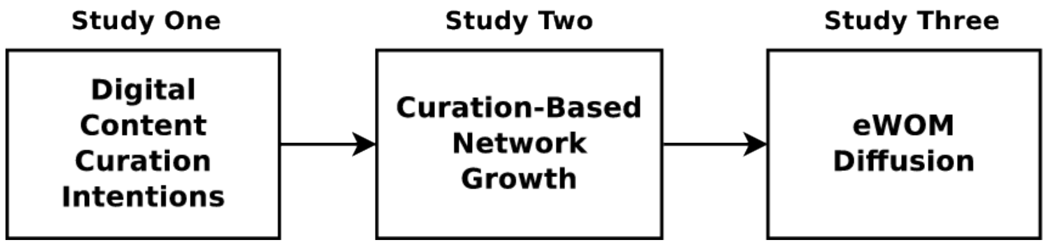

Figure 5. Research framework

After establishing the antecedents of digital content curation, the study makes a logical shift towards the examination of actual digital content curation behavior. The second study examines the way that specific digital content curation actions attract followers in a CBN environment. Here, we use theory from marketing to divide digital content curation activities into those actions that generate traditional media (advertising) and *socially-earned media*. Socially-earned media, which researchers have only recently begun to examine, consists of any blog posts, tweets or other user-generated online social network media about a company, individual or brand (Stephen and Galak, 2012). We examine the impact of traditional and socially-earned media within a real-world CBN to determine their impact a CBN user's follower network. Data for this study comes from *Pinterest*, one of the largest and most successful CBN today. We collected data for this study over a period of ten weeks using a number of webcrawling applications developed using the Scrapy[6] open-source framework. Scrapy provides a collection of python classes

[6] http://www.Scrapy.org

that allow for web crawling and html scraping, using modern text selection techniques such as XML path language (Xpath) (Clark et al., 1999).

The study makes a number of important contributions to research and practice in the growing area of eWOM research. The study provides some surprising evidence that the format of socially-earned media may be extremely important in a CBN context. We find that textual eWOM plays little to no role, while the sharing of rich media digital content made possible by CBN is strongly associated with follower growth. This finding is very surprising and should make the dissertation study of great interest to researchers and academics, especially given the established effectiveness of textual comments as electronic word-of-mouth (eWOM) in other research settings (Chevalier and Mayzlin, 2003; Mudambi and Schuff, 2010). Our research extends and contributes to this by revealing that textual eWOM may be sensitive to context. We are able to conclude that the CBN context, because it allows for such robust digital media sharing, naturally suppresses the role of textual communication to some degree.

Finally, the third dissertation study expands on the findings of study two, by studying the diffusion of digital content through CBN. 30 different companies within the Pinterest CBN are studied over a period of ten weeks. During this time, data was collected on the content curation activities of all of the companies in the sample, as well as every one of their CBN followers. Additionally, through the use of more advanced web-crawling techniques, data was collected on the interconnections between the followers of each company. In this way we are able to endogenate the network as a function of the observed digital content curation activities.

The study makes numerous contributions. Most notably, it addresses the problem of interest similarity in information diffusion. Past research in information systems has studied diffusion of information through online social networks (eg. Susarla et al. (2012b); Trusov et al. (2009)). These studies invariably point to the important role that individual interests and preferences play in determining the way that information diffuses across a network. However, because information on user interests is rare, these studies merely control for interest similarity without directly measuring it. The data collection for this final dissertation study does not suffer from these limitations. We develop a measure of interest similarity and study the impact of similarity on information diffusion within CBN.

Together, the three dissertation studies are meant to provide a holistic picture of the role of digital content curation in CBN. In the next chapters, we set the stage for the studies by providing a thorough discussion of the extant literature around eWOM in the IS literature. This is done to ground the studies of the dissertation in this established body of work and more clearly define the way that each study makes a unique and valuable contribution to this important area of research. Subsequent chapters then present the three studies in detail. Each study is presented with its own introduction, which identifies the relevant research gap and questions that the study will answer. Each study also has sections on the methods and data collection used for the study, as well as a discussion of the major findings of each study.

1.5 Research Domain

Perhaps the most well-known CBN today is Pinterest[7]. Pinterest has been around for about two years, but only recently has the site come to prominence. While growing quietly since its inception, the popularity of the site has since exploded, leading to Pinterest earning the distinction of becoming the fastest web site to reach 10 million

[7] www.pinterest.com

unique users [8]. A recent report found that Pinterest ranked third among social network sites in terms of number of US visitors, behind only Facebook and Twitter [9].

Within Pinterest, users do not create profiles full of personal information. In fact, users can only share their name, location, and a brief 200 character personal description. Rather, Pinterest pages consist of a set of digital content that becomes the digital self-expression of the user within the network. This digital content can then be viewed, commented on, or even copied onto the pages of other users (Carr, 2012). Pinterest and sites like it give users numerous tools with which to collect content form around the Internet. Most notably, the site provides a browser applet that lets users "pin" content from any page they visit. The applet preserves the name of the person who pinned the site, as well as the site from which the content originates. This allows every pin within Pinterest to point back to the original site, and provides Pinterest with a means of referral revenue.

Many social network sites earn revenue by pointing users to content on other sites. Pinterest partners with a company called skimlinks [10] to turn the links on their pages into *affiliate links*. These links preserve Pinterest's name when users go to another site. What this means is that if a person clicks on a product they see in Pinterest, and then goes and buys on Amazon, then Pinterest gets some revenue.

Pinterest has been extremely successful at doing this. One recent study found that Pinterest was third of all websites, behind only Google and Facebook, in terms of driving referral traffic [11]. This is amazing considering that Pinterest has only a small fraction of the user base of either Google or Facebook, yet it makes sense when one sees the sheer

[8] from `http://tech.fortune.cnn.com/2012/03/22/` Accessed on April 9, 2012
[9] `http://www.cnn.com/2012/04/06/tech/social-media/pinterest-third-social-network/index.html`
[10] `http://skimlinks.com`
[11] `http://blog.shareaholic.com/2012/03/pinterest-referral-traffic-2/`

amount of products and services that the Pinterest format is able to display at once. Figure 6 shows a sample of a sample of a Pinterest user's virtual pinboard to which they have "Pinned Interests".

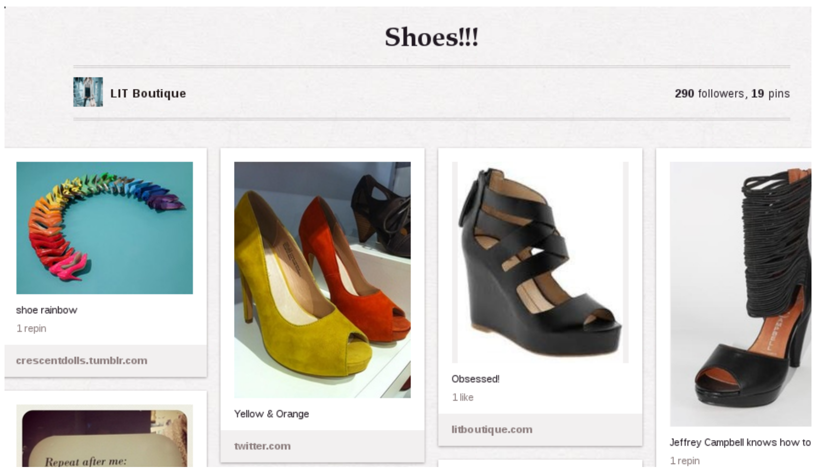

Figure 6. Pinterest users "pin interests" onto virtual Pinboards.

The trend towards ever-richer forms of media has resulted in many online social networks now mediating at least some degree of content curation behavior. Sites like Pinterest simply represent the latest incarnation of content-based social networking. Of the technologies discussed thus far, Pinterest exhibits the highest degree of content curation. Within Pinterest, users do not create profiles full of personal information. Rather, Pinterest pages consist of a set of digital content that becomes the digital self-expression of the user within the network. This digital content can then be viewed, commented on, or even copied onto the pages of other users (Carr, 2012). In this way, Pinterest mediates an active and robust transfer of eWOM marketing information. It provides a context and level of eWOM diffusion on a scale that did not previously exist within online social networks.

For these reasons, the dissertation, while dealing with a number of topics and methodologies across the three proposed studies, is still fundamentally grounded within the larger category of research around eWOM. We incorporate data from the Pinterest CBN in each of the three dissertation studies. Study 1 observes a population of current Pinterest users to identify the antecedents of eWOM generation, study 2 uses a large panel of data collected from the Pinterest CBN to analyze the the impact of eWOM in a CBN enivornment, while study 3 incorporates network-level data on the networks of 30 companies within the Pinterest CBN to examine the diffusion of eWOM through CBN structures.

Before presenting these studies in detail, we provide an overview of the extant eWOM literature relevant to our discussion. Understanding and explaining the reasons that CBN and digital content curation are well-suited for eWOM diffusion, and as a result excellent avenues for online social network revenue generation, is the overarching motivation behind the entire dissertation project. This literature review is the subject of the next dissertation chapter.

CHAPTER II

EWOM IN THE IS LITERATURE

We will eventually argue that CBN and digital content curation create not only new avenues, but also a new type of eWOM, in which the primary means of information transmission takes place as digital content, rather than textual communication between network users. Before discussing digital content curation in detail, however, we begin by developing a comprehensive understanding of past research around eWOM, its challenges and opportunities. In the remainder of this opening chapter, we discuss the extant literature around online social networks and electronic word-of-mouth communication. We then discuss an overview of some of the literature around digital content curation. The goal of this opening chapter is to build a bridge between the concepts of digital content curation and the information diffusion made possible by electronic word-of-mouth.

Research in and around eWOM has continued apace within the IS field in in recent years. Changing technologies have created fruitful avenues for eWOM research, and researchers, quick to capitalize on these opportunities, have greatly expanded the domain of eWOM research. While eWOM is fully deserving of the attention it has received, the introduction of new methodologies and theories from areas such as economics, sociology and marketing have created a fragmented body of work with few concrete definitions of themes. As a result, eWOM research in the IS literature today has significant breadth but little depth. The purpose of the following sections is to analyze and assimilate a large amount of the best work from the IS literature in and around eWOM. We seek a definition of eWOM that is both concise enough to provide a definite boundary around the eWOM research domain, yet flexible enough to allow for generalizations in the face of changing and expanding eWOM technologies.

Early research on word-of-mouth dates back over fifty years (Arndt, 1968; Dichter, 1966). The majority of early research focused on understanding the nature of, and potential impact, of word-of-mouth. WOM was seen as a relatively inexpensive means to reach a large number of customers, with the WOM channel conceptualized as simply another channel for companies to use for marketing purposes. A major reason for the increased interest in *electronic* word-of-mouth arose together with the dyadic communication made possible by Web 2.0 technologies. These technologies, for example blogs, social networking sites, online opinion forums etc, allow user to create and share their thoughts and opinions with large numbers of people. Today, many Web 2.0 features are standard on nearly all major websites (Parameswaran and Whinston, 2007). Examples include the online customer review systems seen on e-Commerce sites like Amazon.com, comment threads and discussions on social network sites on Facebook, and sites that allow for rich media sharing such as Pinterest, Flickr and YouTube.

Web 2.0 technologies have expanded the scope of WOM research, which was once almost solely the area of marketers and advertisers. Today, eWOM research draws in scholars from any number of fields and incorporates a multitude of different theories and methodologies. As a result, while the academic interest around eWOM is valuable and necessary, the pace of eWOM research growth has led to body of work that is fragmented and lacks few consistent definitions and themes (Cheung and Thadani, 2012). Additionally, rapidly changing technologies have produced a research community that places emphasis on exploring eWOM in new settings, rather than rigorously arguing for the definition of boundaries on the eWOM research domain.

Electronic Word-of-mouth refers to statements, either positive or negative, made about a company media personality or brand made over the Internet. More simply, word-of-mouth happens when people talk to one another online. For decades, organizations have spent considerable effort and resources on developing a carefully cultivated marketing message that presents their firm or brand to the public (Mudambi and Schuff, 2010). Misner (1994) calls it the most-effective, yet least understood form of marketing. This is because eWOM runs parallel and somewhat outside of the formal communication actions of organizations. As a result, it can be influenced by the organization, but the level of control that a firm can exert over eWOM is far less than traditional media channels like television, radio and internet banner advertising (Trusov et al., 2009).

For its difficulties, eWOM has some excellent advantages. First, its reach is unmatched (Dellarocas, 2003). eWOM can spread through a population at tremendous speed (Bikhchandani et al., 1992). eWOM is also highly motivating, as people are often influenced to a great extent by what their peers think, say and do (Stephen and Galak, 2012). Finally, eWOM is usually thought to be cheap compared to traditional advertising. These characteristics combine to make the eWOM channel highly desirable.

There is another reason that organizations must care about eWOM. Unlike formal marketing campaigns, word-of-mouth conversations exist independently of organizational action. People are always free to talk and share their opinions. More importantly, people are always free to *complain*. Online social networks have provided customers with ample opportunity to discuss products and services. As Dellarocas (2006) points out, companies must embrace the eWOM discussion around their products and brands, as negative word of mouth can significantly impact a company that does not have control, or at least an understanding of, the nature of discussion within their brand communities.

2.1 What is Electronic Word-of-Mouth

As previously stated, eWOM refers to any positive or negative statements about brands, companies and people spread via the internet Thorson and Rodgers (2006). Naturally, this definition includes a large amount of electronic communication. Twitter feeds, Facebook posts, online customer reviews, as well as images and YouTube videos can all be considered forms of eWOM. Because eWOM is relatively new, and new opportunities for the study of eWOM are appearing at a rapid pace, multiple definitions and terms have been used in the extant research (Cheung and Thadani, 2012; Vilpponen et al., 2006). Viral marketing, online feedback mechanisms, online customer reviews, word-of-mouth advertising, online referral systems, peer endorsement systems, and word-of-mouse can all be considered as subsets of the larger category of eWOM (Vilpponen et al., 2006).

eWOM information can be logically divided based on several criteria. eWOM can be differentiated based on whether it is actively *solicited*, as in the case of customer feedback forums (Chevalier and Mayzlin, 2003; Mudambi and Schuff, 2010) and systems for collecting online customer reviews. eWOM information can also be divided based on whether it is under the control of one particular company. Figure 7 shows our conceptualization of eWOM systems in terms of active eWOM solicitation and control. The upper right quadrant contains eWOM that is actively solicited from users, and fall under the control of a particular company or brand. eWOM in this category includes feedback forums or comment sections on a company's website.

Companies can, and do, strategically manipulate eWOM in these environments to their advantage (Dellarocas, 2006). As a result, the company ultimately has the control over the discussion. Undesirable comments and posts can always be deleted, and representatives from the company can step in to manipulate the discussion at any time (Dellarocas, 2006). These factors make the eWOM message within these systems is easier to control, yet for precisely this reason, eWOM within these systems is not as credible as other independent sources.

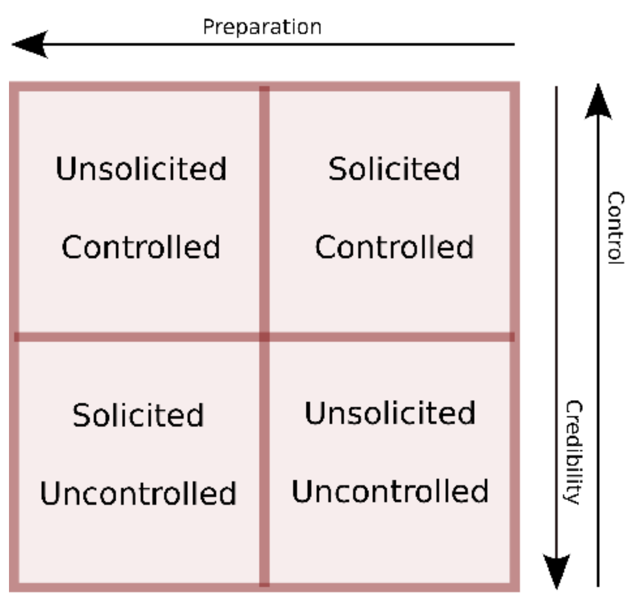

Figure 7. Grid showing the four classifications of eWOM systems.

An example of solicited, yet uncontrolled eWOM is seen on customer review systems used by companies like Amazon.com and Yelp. Unlike eWOM on a company's website, companies like Amazon and Yelp solicit reviews about a large number of products, services, and businesses A customer review system like Yelp.com takes the control away from the company being reviewed. Occasionally these sites will allow a product manufacturer or representative to add their own comments to reviews, in an attempt to defend themselves or encourage accurate discussion. By and large, however, the eWOM discussions on these sites are out of the company's control.

Many *viral marketing* campaigns also promote solicited, yet uncontrolled eWOM. Viral marketing is defined as a marketing strategy in which a company uses the internet to encourage consumers to spread a message about a product or service to other consumers (Bampo et al., 2008). As Vilpponen et al. (2006) point out, viral marketing is only as valuable as the number of users it attracts. Viral marketing is solicited eWOM in that it is typically initiated by a company with a goal of attracting positive attention for the company or brand. As a result, it has specific objectives and goals (Bampo et al., 2008).

Realizing these goals and objectives is not a trivial challenge, however, and Viral marketing can often be difficult to control. Such was the case of Chevrolet, who in 2006 decided to initiate a viral marketing campaign by encouraging users to create their own commercials for the Chevy Tahoe (Bosman, 2006; Donaton, 2006). Chevy solicited these commercials by providing a robust website for compiling Chevy resources into official looking video commercials. However, because users could use the provided resources with little restriction, and then upload their creations to YouTube and other sites, the viral marketing campaign quickly went out of Chevy's control. This was potentially a problem, as many ads mocked the Tahoe as an oversized environment-destroying tank.

To the credit of Chevrolet and their parent company General Motors, they did not attempt to remove or block negative ads. It may be the case, however, that this experience made General Motors more wary of eWOM marketing efforts in the future. The company recently made news for pulling its ads from Facebook in the lead-up to the social network site's large initial public offering [1]. At the time, GM cited the ads as being ineffective and impossible to evaluate.

Figure 8. A user-made Chevy Tahoe ad goes a little off-message.

The upper left and lower right quadrants of Figure 7 contain unsolicited forms of eWOM. This category includes many online social network sites like Facebook and Twitter, which are not specifically designed to capture eWOM about companies products and people. Nevertheless, these systems often serve as the vehicle for the transmission of large amounts of eWOM around products companies and brands. The actions of companies and media personalities often cross over into these systems, and sometimes form important topics of conversation among the systems' users.

This concludes our discussion the types of eWOM. This section is meant to show the differences between types of eWOM, and the practical ways that eWOM systems

[1] http://www.forbes.com/sites/joannmuller/2012/05/15/gm-says-facebook-ads-dont-work-pulls-10-million-account/

introduce an element of stratification into the eWOM landscape. In the next section, we move away from a discussion of the nature of eWOM, and instead discuss some of the practical realities of eWOM data collection that influence existing research efforts.

2.2 Literature Review Methods

A seminal article in eWOM literature is Dellarocas (2003)'s 2003 article in Management Science. Our initial review of eWOM literature in IS considered articles citing this paper. We also looked at articles that referenced other important eWOM papers (eg. (Chevalier and Mayzlin, 2003)) and searched for eWOM articles using a number of different keywords in online databases and Google Scholar. Table 1 shows the distribution of journals for the papers in our sample.

Journal	Number of Articles
Information & Management	3
Information Systems Research	15
Journal of Consumer Research	3
Journal of Management Information Systems	13
Journal of Marketing	3
Management Science	11
MIS Quarterly	3

Table 1. Distribution of eWOM articles by journal

Consistent with the method used by (Roberts et al., 2012), we divided articles up into three categories:

(1) *Referenced but not central*: These articles cite Dellarocas (2003) but are not explicitly about eWOM. They do not make significant contributions to eWOM

literature and do not rely much on existing eWOM work to build their theoretical arguments.

(2) *Provides Theoretical Support*: These articles use eWOM to build their theoretical arguments.

(3) *Used in a hypothesis or research model*: These articles actually build hypotheses around eWOM, and look to add a contribution to the extant literature.

(4) *Forms the theoretical base for the article*.

2.3 Data-Driven Assumptions: Issues of Homogeneity

eWOM systems collect and diffuse very different types of information, and involve different types of actors and community populations (Cheung and Thadani, 2012). Additionally, the many types of information that have been classified as eWOM in the extant research have created several assumptions about eWOM that may lead to problems related to measurement or incomplete explanations. For this reason, it is important to understand the common assumptions that appear in much of the extant eWOM literature.

First, a great many studies of eWOM communities assume that populations are what we call *behaviorally homogeneous*. Behaviorally homogeneity implies that the antecedents that drive eWOM impact or diffusion are unrelated to behavioral or psychological differences between individuals, but rather related to the eWOM information itself. This assumption can be considered data-driven, in that it often occurs in studies that take eWOM data as the unit of analysis. For example, in Mudambi and Schuff (2010) study of online customer reviews on Amazon.com, the authors study characteristics of the eWOM information (the reviews themselves) to identify which reviews are likely to be flagged as helpful by the Amazon community. Studies like this are quite common in the extant eWOM literature, yet many of them fail to ask "helpful to *whom*". By assuming away individual differences between members of the population, it is impossible to ask whether certain populations may utilize eWOM information in different ways.

This is problematic, especially in light of research from psychology and information that shows that people tend to use online information in different ways, based on their own problem solving preferences. For example, the theory of regulatory orientation developed by Higgins et al. (1982) and others identifies two groups of people. According to this research, while some people may benefit from the open-ended messages offered by certain online customer reviews, other groups may desire more structure or summary.

Based on our literature analysis, the assumption of behavioral homogeneity occurs in eWOM literature for two primary reasons. First, studies that focus on eWOM information ignore the latent behavioral effects that dictate how that information is received or passed on. This is interesting, because there are numerous studies of eWOM systems that unfortunately offer us little in addressing this question. For example, Facebook has its own huge stream of behavioral research (Boyd, 2009; Ellison et al., 2007b; Utz and Kramer, 2009). Many studies have specifically examined behavioral antecedents of Facebook usage from a number of social and psychological perspectives. Unfortunately, these studies almost unanimously take a generic view of usage as a dependent variable.

General notions of online social network usage cannot realistically contribute much to the study of eWOM, in that it is impossible to conclude that *intentions to use Facebook* necessarily equal *intentions to use Facebook to spread eWOM*. In some ways this is an interesting irony, as it is exactly Facebook's potential for eWOM marketing that is at the root of its current valuation woes and investor confusion about the potential of the company. To address this problem, eWOM researchers need to begin to *unpack* the general usage construct and identify exactly what social and behavioral antecedents drive a person's need to either acquire or pass on eWOM information in both Facebook and other eWOM systems.

Another major data-driven assumption in eWOM research is *informational homogeneity.* Studies that assume informational homogeneity assert that the spread or impact of eWOM has little to do with what comprises the eWOM information. In other words, we study the spread of the eWOM, while ignoring the content of the eWOM message. This assumption exists in both research that studies eWOM spread, or diffusion, and studies that look at the impact of eWOM information (Cheung and Thadani, 2012). In studies of eWOM impact, there is typically some conception of positive or negative eWOM. For example, if it is a study of online feedback or reputation, a boolean value denoting a positive or negative feedback is often available. This conceptualization of eWOM as simply positive or negative unfortunately strips away much of the value of eWOM as a written narrative.

There is usually some emotional aspect of eWOM information that can be easily identified by a human audience (Kim and Gupta, 2012). These concepts do not fit easily into such mathematical definitions as positive and negative. Unfortunately, opening the black box on eWOM content is very difficult. As an example, consider the ongoing initiative around summarizing online customer review content (Hu and Liu, 2004; Hu et al., 2006). There is an ongoing debate about how best to accomplish this objective, with researchers asking tough questions about what exactly is valuable in a particular review. Is it the objective information about a product, i.e. its goodness or badness, or rather is it the emotional connection that the review engenders between the reader and the author (Hu et al., 2006; Jeong and Jang, 2011)? The former might lend itself to summarization. The latter is much more difficult, and in fact we are only aware of one published study that attempts to address the issue of narrative context and connection in online customer reviews (Kim and Gupta, 2012).

Studies of eWOM spread also exhibit informational homogeneity. In these studies, diffusion is tied to structural characteristics of an eWOM network (i.e. its connectedness or the presence of structural holes) (Susarla et al., 2012a). Attention is rarely given to the actual message that is diffused. This causes two problems. First, it may make it difficult to identify messages that have economic or marketing value from those that are merely part of everyday public discourse (Mayer, 2009). For example, Twitter maintains track of daily hot topics. Very often, these are hugely popular, but offer little connection to a product, brand, or even a media personality. While we believe they are still eWOM, they are of little actionable value. Without knowing the nature of eWOM diffused through a network, how can be distinguish the fluff from the good stuff?

In studies of eWOM spread, informational homogeneity also often extends to the network members themselves. This creates a double black box for many studies in this area. In these studies, we know the shape and structural nature of the network, but know little or nothing about the information passing through the network, or the personal characteristics of the individuals that occupy network nodes. In recent years, several studies have attempted to address this issue, with interesting results (Agarwal et al., 2009; Lewis et al., 2012). These studies have found consistently that information symmetry is an important aspect of information diffusion. Information symmetry essentially means that close network nodes share some similarity around the information that is to be diffused. When such similarity is lacking, diffusion is less likely to occur (Lewis et al., 2012).

This has significant implications for eWOM marketing initiatives. Companies are typically interested in diffusing certain types of information. In particular, information related to the commercial products or services that the company sells. In order for a network to offer much in the way of eWOM diffusion potential, close network nodes need to share an interest in this type of information. In other words, if a company wants to diffuse information about a particular interest, say for example tennis, then it needs to get its information into a network of Tennis players.

Current online social networks may not be able to guarantee such a network. For example, a typical user of a site like Facebook may share connections with some tennis players, but their network will be populated by many others who do not share this interest. That is because the context of Facebook is one of social connections, not shared interests. More recent social networks like Pinterest have a context that is interest-based, and *may* offer better opportunities for diffusing this type of information. To avoid assuming information homogeneity, however, researchers studying diffusion in any network context should make an effort to collect data on eWOM information in addition to the usual network structure data.

2.4 Conceptualization

In this section, we examine the different ways in which eWOM has been conceptualized in the existing IS literature.

In the majority of existing eWOM research, eWOM is characterized as information. This view is consistent with the commonly accepted definition of eWOM as a "positive or negative *statement* made about a brand, company or media personality". For many types of eWOM research, this type of conceptualization works well. For example, studies of online customer reviews and product recommendations undoubtedly involve statements that can be passed from one person to the other (Mudambi and Schuff, 2010).

However, for other studies of eWOM, the concept of eWOM as a statement seems less appropriate. For example, studies of viral marketing focus on the proliferation of eWOM as a virus. Ultimately, these studies measure individual states (i.e., infected with the virus, or not). In such a context, it seems that the conceptualization of eWOM as information is less clear. This distinction is perhaps seen best in studies of eWOM as interest diffusion through social networks. These studies view eWOM not as a particular statement about a company, but rather some characteristic of a network node.

For example, if a user in a social network has an interest in an Apple iphone, and through their own influence is able to pass this interest on to one of their close network nodes, then can we say eWOM has occured? Obviously the sharing of interests has strong commercial implications for the company in question, in this case Apple, but it is not clear what exactly makes up the eWOM information. We propose that the existing definition of eWOM may not be adequate to address this type of interest diffusion.

Additionally, another aspect of eWOM has emerged recently, in the form of rich media sharing. Here again, the actual information being diffused may not take the form of a statement, but rather an image or rich media. Obviously, images can be powerful motivators of human behavior and opinion (CITE). Pictures can also make statements, though the shape and nature of these statements may be more open to more complicated personal interpretation than textual statements.

Consider the example of Pinterest. Pinterest is quickly becoming a leading driver of referrals in the internet community. A recent study found that Pinterest drives more referral traffic than YouTube, LinkedIn and Google+ combined. What do we call these referrals? They are not textual in the sense that one Pinterest user does not tell another to visit a particular page. Rather, they are link driven referrals made possible by *content curation*. Through the process of content curation, Pinterest users create webs of image links that allow other users to traverse sites that interest them. These links are very often devoid of any sort of textual eWOM information, yet they are eWOM nonetheless. A random sample of 2,000 Pinterest users found that far less than 1% of their webpages contained textual comments of any kind.

2.5 Measurement and Level of Analysis

eWOM studies can be loosely placed into four main categories in terms of level of analysis. These categories, owing to a number of factors, in turn drive most of the measurement and data collection that takes place within the studies. The three levels are: 1) Individual, 2) Organization, 3) Information and 4) Network. Studies that examine eWOM from an individual level can be empirical or analytical. Empirical studies of individual eWOM diffusion typically examine the antecedents that lead to either positive or negative word-of-mouth. In much of this research, eWOM diffusion is considered to be either good or bad for the company, and studies use survey data collected from individuals to measure how likely the spread of the eWOM message would be.

Many analytical studies also examine the behavior of individuals operating within eWOM environments. These studies focus on the rational behaviors that lead to information sharing within the network. Very often, analytical studies make use of game theory or other analytical modeling techniques to show the way that individuals would share information.

Organizational eWOM focuses on the impact to the organization of eWOM data or diffusion. A good example of an organizational-minded eWOM study is Dellarocas (2006). In this study, the authors discuss the benefits and circumstances that drive strategic manipulation of online opinion forums. The focus of the study is on whether or not it is beneficial to take an active role in manipulating the eWOM discussion within these types of information systems.

Informational studies of eWOM deal with the information itself. These studies actually examine the spread of a particular piece of content or information. Before electronic word-of-mouth, such studies were particularly difficult. Milgram (1967)'s seminal work is a notable exception. However, in the internet age it is easier to trace a message as it travels from one person to another. Many types of eWOM studies, especially studies of viral marketing, actually study the diffusion of a particular piece of information or innovation.

Finally, there are numerous eWOM studies that examine the eWOM network. These studies take two forms. First, they may be individual-based network studies that look at the network position of a particular node and measure its potential for eWOM diffusion. Additionally, some studies examine the network as a whole. Owing to the fact that network-level data is rare and hard to come by, many studies focus on the characteristics of individual nodes. However, some studies examine ego networks.

Another form of individual level eWOM collects data from individual nodes within a larger eWOM network. Within this type of study, eWOM can also be measured as the spread of a message

2.6 Network Space (How has it been used?)

Two major themes dominate much of the existing literature around eWOM. These are eWOM *spread* and eWOM *impact*.

1. eWOM spread refers to the study of eWOM diffusion within networks or communities. Much of the research on eWOM spread draws upon the work of Rogers (Rogers, 1983) around the diffusion of innovations. Because a good amount of eWOM research involves analytical models, many studies adopt and extend the Bass model (Bass, 1969). The Bass model adapts many of the concepts from the diffusion of innovations theory to an analytical context. It provides a prescription for the mathematical treatment of many of the ideas contained in the theory of diffusion of innovations.

Network spread research can be both analytical or empirical in nature. Analytical studies use mathematical modeling techniques to show optimal network structures and characteristics that maximize the diffusion of eWOM. Empirical studies rely on data collected from individuals around eWOM activities. Much of the empirical eWOM research uses data collected on networks and network structure. This research maps out network structures and attempts to model and predict the spread of eWOM through these structures. Some studies of eWOM spread do not consider network structure, but rather focus on the intentions to share information between individuals. ‘

2. The other major theme in eWOM research focuses on eWOM impact. These studies are not so interested in the diffusion of eWOM, but rather the impact to the individual or organization that eWOM diffusion entails. The majority of studies of eWOM are empirical in nature. This line of work has yielded some interesting findings for a number of different areas. For example, studies of eWOM and product sales (Chevalier and

Mayzlin, 2003). eWOM has also been shown to improve customer loyalty and trust, as well as improve a customer's opinion of a product (Mudambi and Schuff, 2010).

Over the last several decades, eWOM research within these themes has produced several key findings. First, early studies of eWOM impact worked to tie eWOM to variables of organizational interest, with some success. For example, Chevalier and Mayzlin (2003)'s study was one of the first to tie eWOM to sales cuss within the organization. Other studies have since expanded on this to identify the specific ways in which eWOM impacts customer behavior favorably for the organization. This work has successfully attributed eWOM to improved trust, opinions of products and brands, and greater customer satisfaction.

Another major finding concerns the type of products and situations for which eWOM is most effective. Mudambi and Schuff (2010) was one of the first to consider the role of eWOM in product searchers for *experience* and *search* goods. Experience goods are goods for which a product valuation cannot be created prior to using or experiencing the product. These types of products are typically tied to personal taste rather than some objective measure of product quality. Books, movies and music would all be examples of experience goods.

Because of the nature of experience goods, consumers shopping for them tend to rely more on the opinions and experiences of each other, as recounted through online customer reviews (A type of eWOM). Mudambi and Schuff (2010) showed that online review systems like those seen on Amazon.com are more valuable to customers shopping for experience goods.

Studies of eWOM spread have also made several major findings. One of these which has come about fairly recently concerns the importance of *network context* in the spread of eWOM. Network context refers to the criteria used in link formation within the network. For example, Facebook users connect based on social relationships, while LinkedIn users connect based on work relationships. Studies of online word-of-mouth have shown that network context directly impacts the similarity of close network nodes around a particular piece of information. This is important, because for information diffusion to occur, network nodes must share some similarity around the information to be diffused (Rogers, 1983).

2.7 Guidelines and Conclusion

As previously stated, one major concern with much eWOM research concerns the difficult nature of the data collection. This is particularly true with research on eWOM spread. Because of the invisible nature of eWOM networks, certain assumptions often have to be taken related to visualizing any network. For example, studies of online social networks are often dealing with networks that are too large for any reasonable data collection. The primary means of collecting social network data involves web-scraping spiders that crawl network links. A network of millions of users provides too great a challenge in terms of network load and computing time.Additionally, as many online social networks allow users the ability to make some or all of their profile information private, it creates gaps of invisibility within the network.

To help with these problems, researchers may want to focus on *ego networks*. Ego network analysis is a useful subset of social network analysis that is appropriate anytime the entire network cannot be seen (Prell, 2011). Unlike traditional methods of social network analysis that study the entire network, Ego network analysis studies a subset of the network surrounding a particular network node of interest (the ego) and all the nodes connected to the ego (the alters). Ego network analysis allows for the calculation of traditional network analysis measures when the whole network cannot be located or seen.

Our goal is this dissertation chapter has been a thorough and robust treatment of the extant IS and marketing literature around eWOM communication. Based on our literature review, we have identified several trends and areas in the IS literature that demand attention. First, extant studies have identified that interests, tastes and preferences play an important role in the diffusion of marketing information (Lewis et al., 2012). As we have said, each online social network offers a context of link formation. Interestingly, many of the largest online social networks, for example Facebook and LinkedIn, do not possess a context which lends itself well to interest-based link formation. This dissertation has the potential to fill a gap in the extant literature by focusing on the CBN context, in which people form network links based on shared interests in digital content. As a result, we expect that many of the findings of the dissertation will contribute significantly to our understanding of interest-based information diffusion.

Another major gap in research concerns the nature of eWOM data collection. As our literature review has discovered, much eWOM literature is forced to adopt several important assumptions. For example, studies of eWOM spread typically are not privy to exactly what information is being diffused, and are therefore forced to assume informational homogeneity. This dissertation, as we will discuss, uses novel data collection methods that allow us to develop a taxonomy of eWOM information, and discuss the impact and spread of eWOM of various forms.

Finally, we discovered in the course of our literature review that many studies of online social network eWOM consider the actual network structure as exogenous (Jackson, 2005; Mayer, 2009). This is done because of practical realities related to the collection of network-level data. This dissertation aims to fill a gap in this literature by studying CBN in a manner that allows the network to be assumed as endogenous. Our data collection methods allow us to view network formation in real-time, as a function of observed CBN user behaviors.

In conclusion, now that we have established our research domain and the literature base that forms the backbone of this dissertation project, we can move into a discussion of the individual dissertation studies. Recall that we will first examine the antecedents of digital content curation using a survey-based methodology and a sample of CBN user participants. We then move to an econometric analysis of the impact of digital content curation in the CBN environment that incorporates a large amount of panel data on the digital content curation actions of a sample of CBN users. Finally, we study the spread of eWOM information through CBN network structures, using some of the latest methods of social network analysis (SNA).

CHAPTER III

STUDY 1: WHAT YOU LIKE, NOT WHO YOU ARE

> It used to be that we were all just consumers - or most of us were, anyway. We'd watch TV or read a book or listen to the music on the radio that was selected by others for us. But lately. . . tools like Tumblr and Pinterest are making it easier for users to create collections of interesting content and to share them with anyone else in the world. [1].

3.1 Research Gap and Motivation

This dissertation chapter is focused on the individual antecedents of digital content curation within online social networks. Our eventual goal is the development of a theoretical model that identifies both utilitarian and interpersonal motivations for why people collect digital content. We are interested in addressing the research question: *what are the antecedents that drive people to engage in the curation of digital content?*. Our motivations for the study are three-fold.

First, owing to recent technological advances around digital content curation, this phenomenon has received only limited attention in the information systems literature. The research that does exist around digital curation has been limited to libraries and university settings (Guy et al., 2009; Jarvis, 2008). Business applications of digital content curation, specifically those related eWOM, have not been explored. As one of the first studies to actively study curation-based networks and digital content curation, this study fills this gap in the literature. Specifically, we examine the antecedents that lead to digital content curation behavior.

[1] `http://paidcontent.org/2012/03/13/`

The IS has a well-established history of studying the antecedents of adoption of new technologies (Davis et al., 1989; Venkatesh et al., 2003). This is especially true in areas where there is some uncertainty about the kinds of factors that might motivate individuals. The CBN environment exhibits some interesting characteristics that put it in this category. Unlike traditional online social networks, in which primarily social motivations for usage make both theoretical and practical sense, CBN seem to exhibit some different aspects of usage. Indeed, past IS research has had some success mapping the antecedents of online social network usage to such social theories as social capital theory (Putnam, 1995) and social exchange theory (Church and Salam, 2010). However, it is difficult to see how such theories could account for the content management functionality of a CBN. We fill this gap in research by developing a theory that integrates utilitarian functionality of CBN into some extant notions of CBN usage.

Second, the IS community has called for investigations into the nature of influence of tastes and preferences among peers in online social networks (Susarla et al., 2012a). We have previously discussed some previous literature which shows that traditional online social networks, which create networks of peers and social relations, may not necessarily create networks where tastes and interests diffuse easily, and pose challenges for eWOM marketing (Garg et al., 2011; Lewis et al., 2012; Susarla et al., 2012b). Digital content curation creates curation-based collections of content that are used for the formation of links within CBN. We argue that such a network context is inherently suited for the diffusion of eWOM, as evidenced by the huge eWOM referral numbers generated by these networks.

Finally, over the last decade, interest in the study of eWOM has grown substantially. Studies in marketing and information systems literature have tied eWOM to a number of different marketing outcomes, including increased sales (Chevalier and Mayzlin, 2003), customer satisfaction (Godes and Mayzlin, 2004). While much of this literature has focused on eWOM as textual discourse in the form of online customer reviews, youtube comments etc. (Mudambi and Schuff, 2010), technologies for the distribution of rich media have advanced to the point that images now spread as fast or faster than words. This will only continue, as a side effect of the data-entry requirements of mobile technology make it easier to send a picture than a long message. This study is one of the first to identify and explore this extention of eWOM.

This study aims to enhance our understanding of digital content curation and its implications for the development of research around CBN and online marketing. To this end, we develop a model that examines the antecedents of user intentions towards digital content curation behavior within the Pinterest CBN. We contribute to extant literature by incorporating theory from the areas of sociology and personal information management to present digital content curation as a socially observable process of digital content management. Additionally, the study makes valuable first steps towards understanding curation-based networks, and sets the stage for future research into this new online social network context.

3.2 Theoretical Model

To understand CBN, we first have to understand the content curation process by which they are formed. In the next sections, we present a model of the antecedents of digital content curation behavior. Digital content curation differs from typical online social network usage in that it involves a socially observable process of online content management. The major appeal of the model that we develop is that it attempts to explain digital content curation as a social process with practical benefits in the areas of content discovery and organization.

To address the practical benefits of content management, we incorporate theory from the area of personal information management. Personal information management theory explores the activities that people use to acquire, organize and collect the information items that they require to complete everyday tasks and objectives (Jones et al., 2001, 2002). Research on personal information management seeks to understand how people find things that interest them online, and, once found, how do these things *stay found*. In our model, we analyze several types of personal information management in detail, going from the process of finding new content, and then through the organization of this content so that it can be found again efficiently and effectively at a later date.

The social side of digital content curation is modelled based on the work of Goffman (2002) around the theory of self-presentation to explain the way that people use digital content for social impression management (Giacalone and Rosenfeld, 1989). Under self-presentation theory, people are actors performing different roles for different types of social interaction. Key to this theory is the notion of object acquisition. In order to create the desired impression, people seek out and acquire objects, or for our study, digital content (Kim et al., 2012).

The extant IS literature contains examples of research that uses both personal information management (Jones et al., 2001) and self-presentation theory (Kim et al., 2012). Through the combination of these diverse research streams, the study makes novel contributions to both the literature on online social networks and digital content curation. These contributions will be of interest to anyone looking to understand the phenomenon of digital content curation within online social networks. This includes not only companies that want to capitalize on digital content curation's potential to diffuse eWOM for marketing purposes, but also the numerous companies and individuals that produce digital content for internet sites. Digital content curation represents a fundamental shift in the way that information spreads online, and as such it has significant implications for the future of all aspects of electronic commerce.

This study is interested in studying content curation, not in the context of a museum or a closed business or organizational environment. Rather, we study the process of social content curation through online networks. Because digital content curation in this context takes place within a social community, in full view of many other people, our model necessarily include both social and utilitarian antecedents. Whenever an action is observable to others, the nature of that action is changed in fundamental ways (Latane, 1981). Consider the example of playing a musical instrument. Musicians typically have several goals related to playing their instrument. They often give public performances, a conspicuous activity done for the benefit of an audience. During these performances, musicians may adjust their playing or introduce flare and flourishes designed to elicit an audience response (Becker, 1951). When the audience is pleased, they may provide the musician with the conspicuous rewards of money or applause. However, the actual time that a musician spends performing is a relatively small percentage of the total time that the musician spends with his or her instrument (Ericsson et al., 1993). Most of the musician's time is spent practicing in private. Practicing a musical instrument provides intrinsic benefits to the player. For example, studies have shown that practicing a musical instrument can increase cognitive ability and mood (Schellenberg, 2006). This benefit derives from the relationship between the musician and the object of his attention (the instrument) alone.

Researchers in the social sciences have long appreciated the way that socialization impacts the use of technology. This impact is especially apparent in the extant online social network research, where past work has incorporated elements from many major sociological theories (eg. social capital theory (Ellison et al., 2007a), social exchange theory (Blau, 1964; Roloff, 1981), social norms (Fetscherin and Lattemann, 2008) etc.).

Much of this research is interested in the way that people act within a community, in order to elicit a desired response from other community members. Social exchange theory (Blau, 1964), for example, involves a series of mutually-beneficial social transactions, all of which depend on community members observing and reacting to the behavior of the individual (Gatignon et al., 1985). Social capital theory studies the conspicuous benefits that exist within relationships between people (Putnam, 1995). This study contributes to, and extends, this body of work. For while researchers have rightly focused on social theories usage within online social networks, they have largely overlooked the fact that these technologies are being used all the time for personal information management (Shklovski, 2012; Young, 2011). Personal information management involves the activities that people use to acquire, organize and collect the information items that they require to complete everyday tasks and objectives (Jones et al., 2001, 2002). Digital content curation fits well within the personal information framework. However, as we have argued, digtal content curation in the context of social networks is observable, and thus likely subject to some degree of social impact (Latane, 1981). This study combines theories from both sociology and personal information management to approach the interesting phenomenon of digital content curation in a novel way, and in so doing makes a valuable contribution to the extant online social network and personal information management literature. Digital content curation has powerful potential for personal content management, yet this potential is always tempered by the fact that any content added to an online social network is not expressly private. This is a good thing, as numerous studies have shown that users of online social networks are often willing to share information in order to meet and interact with others. In fact, numerous studies have identified that users of online social networks are willing to disclose large amounts of this type of information (Brandtzæg et al., 2010; Christofides et al., 2009; Wasko and Faraj, 2005), provided they something in return (Wasko and Faraj, 2005). Information is not freely given, but rather traded for information about others (Rafaeli et al., 2004).

3.2.1 Serendipitous Information discovery

As users swap, share, and trade digital content within an online social network, they contribute to the network's total knowledge resources, and improve Information discovery within the network. In this study, information discovery refers to activities related to locating a specific digital object based on networked information resources (Lynch, 1995). The nature of the activities that lead to information discovery can be varied, and result in different types of discovery. For example, information retrieval tools like web search engines are capable of returning a huge number of varied results. However, the actual activities that facilitate information retrieval are highly structured (Dumais et al., 2003). These systems return results based on some short query from the user. In many information discovery problems, the actual nature of the desired query may be unknown of so unstructured as to be impossible to articulate. People often approach information discovery with some vague notion of what they're looking for ("I'll know it when I see it"). While the actual answer or final goal of the information search is unknown, there are some basic criteria that are understood and that guide the information discovery process.

In this context, successful information discovery depends on information systems that allow for multiple pathways and circuitous journeys (Huwe, 1999). These elements give a system the flexibility to allow the user to follow strands of thought wherever they might leave, and in so doing discover serendipitous connections between information resources (Foster and Ford, 2003; Zhang and Watts, 2008).

Seredipity is defined by Webster as “The Faculty of making happy, unexpected discoveries by accident”. To quote Lawrence Block: “Serendipity: Look for something, find something else, and realize that what you’ve found is more suited to your needs than what you thought you were looking for.”. The value and importance of serendipity in scientific advancement and human life should not be underestimated. History and Science are both littered with famous examples of serendipity. Vulcanized Rubber was invented when Charles Goodyear accidentally left a piece of rubber mixed with sulfur on a hotplate overlong. Penicillin was famously discovered through serendipitous chance.

So it is that we see that serendipity can be valuable. Yet can it be used or harnessed in the cognitive sense (Foster and Ford, 2003)? As Huwe (1999, p. 2) says, “…searching the (library) shelves for a serendipitous connection may seem quaint, but it remains powerful”. The challenge for the digital library is to preserve this opportunity in cyberspace. The actual process of serendipity is closely related to the joining of cleverly related pieces of information (André et al., 2009). Modern information networks have the ability to connect nearly limitless information resources in a number of ways. Researchers have sought to identify ways of organizing these networks in ways that maximize the potential for serendipitous events (Dumais et al., 2003; Foster and Ford, 2003). There have been some notable successes. For example, the targeted search and query requirements of online search engines would seem to stymie the potential for serendipitous information discovery. However, other features of these technologies, such as their capacity for personalization and social content rankings, the use of these technologies actually increases the possibility of serendipity (André et al., 2009). Fan et al. (2012). In general, however, building serendipitous information discovery systems has proved challenging for researchers. Björneborn (2008) identify several elements of an ideal serendipitous system. These are presented in Table 2.

While a human reader can typically identify interesting information, designing a system that can determine what lies at the heart of this interest can be hard to quantify. Thus, while people can talk about serendipitous moments, actually identifying what triggered the serendipitous event, or designing a system that can promote this type of information discovery, is very challenging (Björneborn, 2008).

Serendipity dimension	**Explanation**
Unhampered access	Unhampered direct access to information resources
Diversity	Rich and dense variety of topics, genres, resources, activities, sections
Display	Curiosity-teasing mediation of information resources
Contrasts	Eye-catching differentiation including quiet zones and display zones
Pointers	Distinct signage, maps, and markers
Imperfection	Imperfect “cracks” and “loopholes” in interfaces
Cross contacts	Contact surfaces across different topics, genres, resources, activities, sections
Multi-reachability	Many different access routes across interfaces
Explorability	Interface invites users to move, explore and browse
Stopability	Interface invites users to stop, “touch” and assess found materials

Table 2. Serendipity dimensions - from Björneborn (2008)

Web 2.0 technologies may offer a promising solution to this design problem (Passant et al., 2008). These systems allow serendipitous connections to form organically, as a function of the process of social information creation (Chang and Quiroga, 2010; Passant et al., 2008). Rich serendipitous information discovery may therefore be found in information systems that allow for socially constructed information connections, rather than cleverly programmed information retrieval algorithms.

Socially constructed information connections give people access to unexpected and often useful information (Borgatti and Cross, 2003). In a seminal work, (Granovetter, 1995) observes that people often learn valuable information through interpersonal connections. This was a major motivation behind the creation of the online social network LinkedIn [2]. Relational information discovery has also been studied around a number of IT related problems, including IT outsourcing (Higgins and Kram, 2001) and open-source applications development (Grewal et al., 2006; Schwab and Miner, 2008).

Despite the large amount of interest in social information discovery, there are few methods for measuring the value of information discovery relationships (Borgatti and Cross, 2003; Cross and Borgatti, 2004). One such in this area is that of Borgatti and Cross (2003). In this study, the authors conclude that social information discovery happens most when people value what a person knows, are able to gain access to that person, and know that information can be sought without incurring a high cost (Borgatti and Cross, 2003).

Social networks based on digital content curations are well-positioned to create instances of serendipitous information discovery. While programming a system to deliver serendipitous connections would be difficult, they grow organically as a happy result of social content curation. People look for these connections, and desire serendipity in the

[2] http://mixergy.com/konstantin-guericke-linkedin-interview/

knowledge discovery process (Foster and Ford, 2003; Zhang and Watts, 2008). As a result, systems that promote serendipitous information discovery may be more likely to see frequent use. Frrm this we derive the following hypotheses.

H1: Serendipitous information discovery will be positively related to Content Curation Intentions.

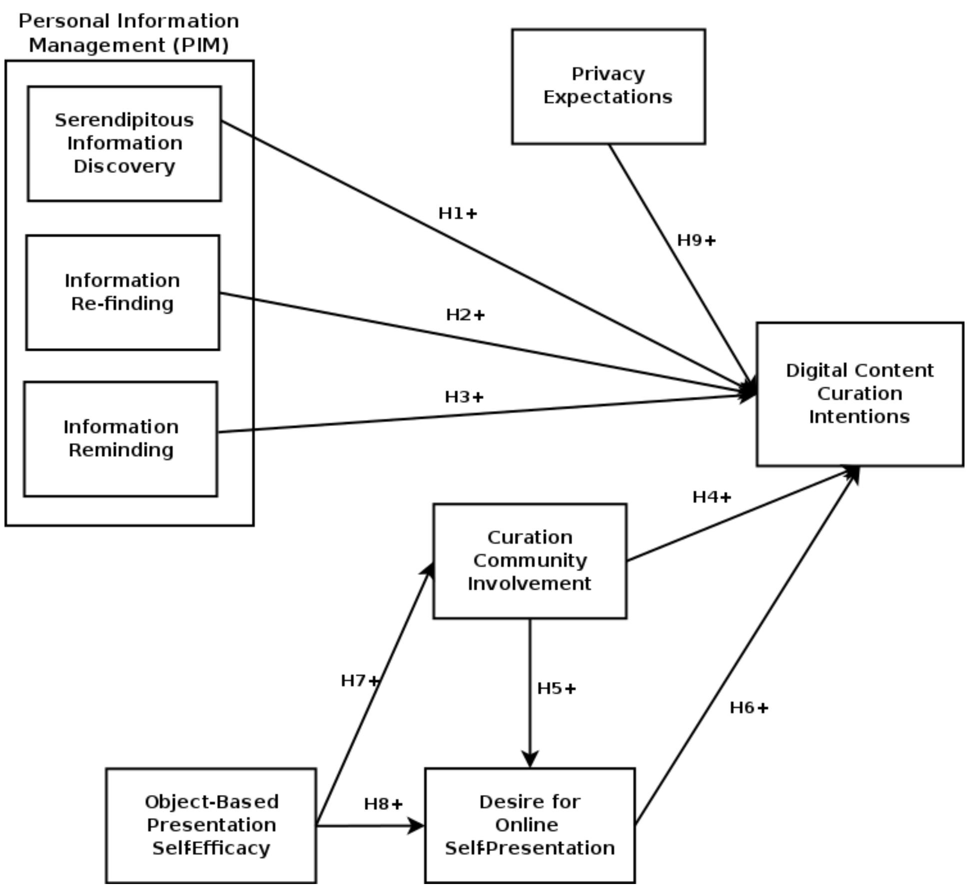

Figure 9. Theoretical model of digital content curation.

3.2.2 Information Re-Finding and Reminding

Knowledge work often involves reusing and integrating familiar information (Dumais et al., 2003). Personal information management theory states that people organize information so that they can find it again ("re-finding") (Barreau and Nardi, 1995). Information that is organized effectively is never in the way, but neither is it so far out of reach that it cannot be retrieved quickly and efficiency. Past research has also shown that people also organize to *remind* (Malone, 1983). Much of the material on a common office desk, or computer desktop, for example, is meant to remind the person of some task that needs to be completed (Barreau and Nardi, 1995). While finding involves intentionally searching for something, reminding happens automatically. Malone (1983, p. 106) discusses the distinction between finding and reminding; "Even though people sometimes intentionally look at their desktops to find what else needs to be done, a primary reason for placing tasks on the desktop in the first place is so that intentional search does not have to be relied on."

As the role of computers in human life has increased, researchers have begun to explore the role of technology in re-finding and reminding. This research involves the development of strategies, techniques, and systems that individuals can use to make sure that they have access to an adequate amount of high quality information suited to the task at hand (Jones et al., 2001). Understanding and improving the process of online bookmarking has been a major area of interest in this area (Jones et al., 2002). Every web browser has methods for creating bookmarks of web addresses. A typical hyperlink is fairly nondescript. The current system for web addressing has no formal indexing system and no consistent naming conventions (Abrams et al., 1998). This creates a situation in which trying to remember the actual addresses of specific web pages is just simply not practical.

Adding bookmarks helps users to manage website addresses that they would like to visit again. However, as (Jones et al., 2001) points out, many bookmarks are forgotten or buried by the introduction of new bookmarks. One reason for this may be that bookmark systems still store web addresses as text. A textual title for a web page may be easier to remember than the actual address, but because of the level of abstraction inherent in written language, remembering and making sense of text is more difficult for human beings that other, rich forms of expression (Lim and Benbasat, 2002).

Another difficulty concerns the fact that users are forced to develop categories and textual descriptions of bookmarks. This creates problems because the cognitive effort required to develop these classifications is quite high (Malone, 1983). Often, many pieces of information do not fit neatly into any established category (Marshall and Shipman III, 1997). The recent development and popularity of applications that allow for graphically enhanced bookmarking points to the growing desire for technological solutions to these problems. Notable examples of these technologies include Instapaper [3], which allows users to save and archive text copies of websites, and the popular iPhone application Flipbook [4], which saves websites and converts them into a format similar to magazine pages for storage and sharing.

Adding rich media, including videos and photos, to the bookmarking process allows for easier mental model development (Jones et al., 2001). This helps with reminding and finding tasks, as pieces of information are now tied to visual representations. Digital content curation offers a means for this type of data organization.

---

[3] `www.instapaper.com`
[4] `www.flipboard.com`

When users add content to a digital connection, the link to the site that originated the content is always preserved. Thus, the collections of digital content become a graphical bookmarks leading back to the locations of interesting content (Cockburn et al., 1999). On Pinterest, this appears to be especially common with recipes and crafts. Often a user may post a picture of food or the completed craft. When users actually click on the picture, they are taken back to the original site where they can view the recipe or the instructions necessary to create the item pictured.

H2: Perceptions of information re-finding utility will be positively associated with digital content curation intentions.

H3: Perceptions of Reminding Utility will be positively associated with digital content curation intentions.

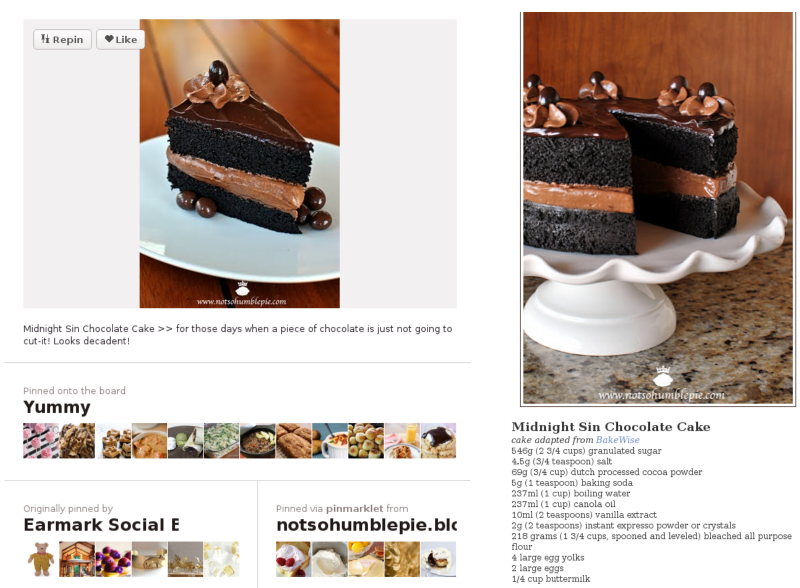

Figure 10. A tasty Pinterest Pin.

3.2.3 Virtual Community Involvement

Virtual community involvement here is defined as the level of commitment and care that a user has for the health and continued life of the digital content curation community (Shang et al., 2006). In studies of online communities, strong ties within a virtual network community are powerful drivers of sustained online social network usage (Ellison et al., 2011).

When people feel invested in a particular virtual community, they care more about the fate of the community, and this manifests itself in a number of ways. For one, people may care more about the way they are perceived within communities that they value (Kim et al., 2012). In this way, involvement may impact a person's desire for online self-presentation.

Virtual community involvement may also produce an indirect effect on the desire for self-presentation, by making the self-presentation process easier and more effective. People who are involved within virtual communities at a high level typically have more channels for information discovery within these communities. Within an CBN, more information discovery channels means that a user is in a better position to uncover interesting content that may be useful for self-presentation. At the same time, a user that exhibits high community involvement may have a better hold on the excepted norms and values within the community (Baumeister and Leary, 1995). Users can leverage this knowledge of community values to more easily craft an appealing self-presentation. For these reasons we propose the following hypothesis.

H4:Virtual community involvement is positively associated with the desire for online self-presentation.

Virtual community involvement, while effecting the desire for self-presentation, may also exert a direct effect on digital content curation intentions. For the current study, we may think of the attention and effort paid to content collections as a representative measure of individual effort within the virtual community. Past studies have shown (DiMaggio et al., 2001; Kavanaugh and Patterson, 2001) that involved community members work hard and exert significant effort to ensure the continued life and vibrancy of the communities that they care about. In knowledge communities like Wikipedia, users invest significant time and effort to make articles as interesting, accurate, and helpful as possible (Ransbotham and Kane, 2011). Within commerce sites like Amazon, users write detailed product reviews and spend time editing, commenting on, and voting for the most helpful reviews (Mudambi and Schuff, 2010). All of these actions take significant effort on the part of the user, yet this effort is essential in creating virtual communities to which people want to belong. We therefore posit that users who report high perceptions of virtual community involvement will exhibit higher intentions to curate digital content.

H5:Curation Community Involvement is positively associated with digital content curation intentions.

3.2.4 Desire for Self-presentation

As humans, we wear a variety of different hats and roles as we interact with different sections of our social and professional lives. People must maintain an appearance for work, while acting and looking another way when at home. A person behaves differently as a spouse than they would as a parent. According to the theory of self-presentation (Goffman, 2002),these different parts of life are akin to actors' roles, in which we as the performers are constantly adapting our behavior for a changing audience, and always looking to present ourselves in a positive light (Giacalone and Rosenfeld, 1989; Goffman, 2002).

In his seminal work, *The Presentation of Self in Everyday Life*, Goffman (2002) uses the analogy of the theatre to describe interpersonal interaction. Goffman conceptualizes people as actors, seeking to influence and control the way that they are perceived by others. In order to achieve this sense of control, people carefully tailor all aspects of their lives toward a desired effect. This process continues when people take their lives online (Ellison et al., 2006; Schau and Gilly, 2003). For example, Ellison et al. (2006) studied self-presentation practices on online dating sites. They found that people used methodical and conscious strategies to carefully monitor the way they were presented to others through the website. Self-presentation has also been examined in the context of the creation and maintenance of personal web pages (Dominick, 1999; Schau and Gilly, 2003).

Much communication online takes place using rich media, symbols, and digital assets (King and Xia, 1997; Robert and Dennis, 2005). These things supplement the otherwise limited form of textual expression (Kock, 2007). As Schau and Gilly (2003) discuss,

> . . . projecting a social presence at a distance has existed in other forms, through letter writing and telephonic communication, and, arguably, since early man's cave drawings, but *telepresence* in computer-mediated environments is uniquely rich. It may include visual, textual, audio, animated, and even haptic sensation. . . this ability to present multiple selves simultaneously is almost inimitable in real life (emphasis added).

According to Goffman (2002) The "social actor" has the ability to influence all aspects of the performance, from the stage to the props. When the "stage" is the internet, users may seek out digital objects that possess some shared value within the group of interest

(Hinz et al., 2010). For example, when designing pages to represent them in the digital domain, people consistently employ strategies that make use of object and brand associations (Schau and Gilly, 2003). In fact, people go to great lengths to acquire these associations, even spending real money to acquire virtual objects (Kim et al., 2012). In their study of digital item sales within online social networks, Kim et al. (2012) found that the desire for self-presentation was a major reason that people purchased digital items.

When the social setting, or format of communication is especially abstract, the shared meaning of digital objects become even more important. For example, the social networks considered by Kim et al. (2012), Cyworld and Habbo, are somewhat different from the model seen on sites like Facebook and Myspace. These sites both use an alter-ego based avatar system designed to create a level of abstraction between the personal identity of the user and the online persona. This abstraction strips away the characteristics of the individual (eg. the person's own language, clothes, etc.), and replaces these real world objects with digital alternatives. For this reason, self-presentation by means of these objects becomes very important (Jung et al., 2007; Kim et al., 2012; Shin, 2008). Personal web pages also present self-presentation challenges. Because the potential audience is essentially the entire internet, it becomes impossible to know the characteristics of the people viewing the page, and thus exactly what self-presentation strategies to employ (Dominick, 1999). Here again, the shared meaning of objects and brand associations becomes an important part of a self-presentation strategy (Dominick, 1999; Ellison et al., 2006; Papacharissi, 2002).

This study proposes that the characteristics of CBN will make object-based self-presentation an important aspect of CBN usage. Like Cyworld and Habbo, CBN communication takes place largely via the use of digital objects, and thus CBN introduce a level of abstraction between the user's personal identity and their network representation (Kim et al., 2012; Schau and Gilly, 2003). This level of abstraction creates a situation in which the collection of, and interaction with, digital content forms the primary way that an individual engages in self-presentation. For this reason, we propose that the desire for self-presentation will motivate individuals not only to collect digital content, but also to engage with the content of others. Thus we propose the following two hypotheses:

H5: Curation community involvement will be positively related to the desire for online self-presentation.

H6: The desire for self-presentation will be positively related to digital content curation intentions.

3.2.5 Object-Based presentation self-efficacy

The desire for self-presentation may be based on a person's level of confidence that they can actually achieve the desired effect (Döring, 2002). Bandura (1977) identifies efficacy beliefs as an important determinant of human behavior and a predictor of outcome performance. Since the seminal work by Bandura, self-efficacy has grown to become one of the dominant concepts within studies of human behavior (Bandura, 1989). The main idea behind self-efficacy is that when people feel that they are good at a task, their actual performance increases (Bandura, 1977).

Efficacy beliefs influence the choices that people make and the social situations that they engage in (Bandura, 2001). These beliefs are task specific (Compeau et al., 1995; Venkatesh et al., 2003), implying that people do not maintain universal self-confidence, but rather exhibit varying efficacy beliefs based on the task at hand. For example, studies have shown that self-efficacy with technology plays a strong role in technology usage, adoption and performance (Hasan, 2006; Hsu and Chiu, 2004; Venkatesh et al., 2003). When a person maintains high efficacy beliefs around a particular task, they are drawn to those behaviors, based on a feeling that they will perform well (Kim et al., 2012). Thus, self-efficacy around self-presentation online should be associated with the desire for self-presentation (Kim et al., 2012).

Object-based presentation self-efficacy refers specifically to efficacy beliefs around the task of self-presentation by means of digital objects. While there is no research that specifically examines object-based online self-efficacy, there are studies that look at self-presentation in object-related tasks. Kim et al. (2012) examines self-presentation in the virtual worlds Habbo and Cyworld. The use of digital objects in these environments is similar in spirit to the conceptualization used here.

Online self-efficacy may be related to a person's familiarity or experience with an online computing task (Compeau et al., 1995; Doyle et al., 2005) As previously mentioned, all electronic discourse involves the use of symbols and abstraction to some extent (Kock, 2007; Schau and Gilly, 2003). This is certainly true within the CBN, where users express themselves largely by means of digital content. The CBN format is likely unfamiliar for some users, and it is easy to conceive of a person who is an effective face-to-face communicator, yet would struggle to express themselves in pictures and symbols. Early studies studies of internet technology examined much the same thing. They found that self-efficacy with these technologies and with textual communication in general were important in predicting people's usage of these technologies (Hsu and Chiu, 2004; Murphy et al., 1989).

While many people today have now used email technologies for years, new technologies and platforms have taken their place (eg. social networks, twitter). Once again, studies of these new technologies have examined the role of efficacy beliefs related to the technology (Gangadharbatla, 2008), and once again these beliefs have been shown to be closely related to usage of the technology. This study proposes that, as these technologies come to rely more on objects and symbols, and less on textual communication, It may be that efficacy beliefs related specifically to object-based self-presentation will become most important for predicting usage (Haferkamp et al., 2012).

For example, object-based self-presentation self-efficacy may explain the gender gap that exist on major social networks. According to studies, 54% of women make use of social networks,compared to 34% of men [5]. On the leading CBN Pinterest, the gender gap is even greater. Pinterest's audience has been measured as comprised of between 68 percent and 83 percent women [6].

That a gender divide exists related to internet technologies is well documented (Haferkamp et al., 2012; Jackson et al., 2001; Ono and Zavodny, 2003). Not only do women use the internet more, they also use it differently from men. For example, women are more likely to write emails while men use the internet for general information search (Jackson et al., 2001). Certainly, gender could explain the population makeup of an CBN like Pinterest. Women are typically more interested in self-presentation (Haferkamp et al., 2012), and because the content within CBN networks is created by users, a network populated largely by women may simply offer content that is not interesting to male users. However, many other CBN exist, some of which are marketed specifically to men. The

[5] http://www.thedrum.co.uk/news/2012/05/29/

[6] http://www.salon.com/2012/05/02/pinterests_gender_trouble/

most notable of these is Gentlemint [7], a “mint for all things manly”. Yet none of these male alternatives have garnered anything close to the attention of Pinterest.

Despite the attention given to gender in internet and online social network studies, detailed explanations of gender-related differences in self-presentation settings are still lacking (Haferkamp et al., 2012). This study proposes that self-efficacy may offer a satisfying explanation for gender-related differences in CBN usage. Efficacy beliefs are intimately tied to the amount of practice and experience that an individual has in the domain of a particular task (Doyle et al., 2005; McPherson and McCormick, 2006). Thus, efficacy beliefs may be higher for people who have past experience in object-based self-presentation tasks.

It remains to define exactly what type of task involve object-based self-presentation. (Csikszentmihalyi and Rochberg-Halton, 1981) discusses the importance of symbols and objects in the home. According to Csikszentmihalyi and Rochberg-Halton (1981) their is perhaps no more significant object in human life that is so intimately tied to self-presentation as the home. Within the home, women traditionally hold several important object-based roles. The first of these is “photo-curator”. Many households maintain collections of pictures of both close and extended family (Csikszentmihalyi and Rochberg-Halton, 1981). In managing these collections, careful attention is given to the way that these pictures fit together and tell a story. Often, it may not be the “best” picture of a person that is chosen for the collection, but rather the one that fits the desired historical period or exemplifies a significant aspect of family life. The goal of the photo-curation task is to manage a presentation of the family to create a desired impression upon anyone who visits the home.

[7] http://www.gentlemint.com

It appears that women are overwhelmingly the photo curators within the home (Durrant et al., 2009a,b). As Durrant et al. (2009b) states:

> ...the mother of the nuclear family continues to assume the roles of ”family photographer“ and ”family chronicler“...we’ve noted in addition how she coordinates the display of printed photos throughout the home environs on behalf of the household-at-large, we refer to this coordinating activity as *home curation* and see that it functions to unify a presentation of the family group, or household, throughout the home.

The other task that we consider here is furniture arrangement. The arrangement of furniture has been shown to be a powerful form of nonverbal communication (Giacalone and Rosenfeld, 1989; Mehrabian and Diamond, 1971; Sommer and Ross, 1958). People arrange furniture to create a desired impression (Giacalone and Rosenfeld, 1989). For example, a doctor might arrange a desk and chairs to create an impression of authority, or separation between the patient and the physician (Miles and Leathers, 1984). People do the same thing at home (Miller, 2001). Purchasing new furniture involves a significant energy and economic investment, with visits to multiple shops and consultations with friends and, often, fashion experts (Miller, 2001). Moving furniture within the home is thus a key part of building a “domestic narrative” (Giddens, 1991) and integral to creating a collective presentation of the home and the family.

Here again, women take the primary role in furniture selection and placement (Pheterson et al., 1971). Additionally, women are much more likely to select a piece of furniture for its self-presentation value. Men are much more likely to select furniture that is comfortable, while women are likely to select furniture that looks good is a space, or creates a desired self-impression (Garvey, 2001).

In summary then, while there is no research that explicitly studies object-based self-efficacy, there is a well-documented gender gap related to internet and CBN usage, with women using these technologies significantly more than men (Haferkamp et al., 2012). Interestingly, women are also much more active in two types of activities within the home that bear a close resemblance to object-based self-presentation. Because, as we propose here, CBN are platforms for online self-presentation, it may be that their experience with these closely related tasks gives women higher self-efficacy in this domain. Based on this, we propose the following hypothesis.

H7:Object-Based presentation self-efficacy is positively associated with desire for online self-presentation.

H8:Object-Based presentation self-efficacy is positively associated with virtual community involvement.

3.2.6 Privacy Expectations

The final antecedent to digital content curation that we consider is user privacy expectations. All online social networks trade in personal, and often private, information to varying degrees. It is well-known that social network sites like Facebook sell personal information directly to marketing organizations for the purpose of advertising and product marketing (Dwyer et al., 2007). Advertising revenues, and by extension the future success of the online social network, all depend on collecting and trading huge amounts of personal information that can then be used by other companies for marketing purposes (Utz and Kramer, 2009).

Privacy has played such an important role in the study of online social networks that numerous studies have been devoted purely to the study of privacy implications in these networks (Son and Kim, 2008; Utz and Kramer, 2009). We believe that digital content curation creates some separation between *who a person is* and *what they like*, and that this separation should change the impact of privacy expectations on the content curation intentions of users within these networks. However, because ultimately any digital content collections within a curation-based network are visible to some degree, privacy expectations likely play some role in determining user content curation intentions.

We do expect, however, that privacy will not be *the* driving factor in determining content curation intentions, on account of the distance between their content and their personal selves. Curation, by definition, is not just a process for preserving, but for preserving with intention to display. Thus, digital content curation can be thought of as a process for displaying *art*, while traditional online social networks display the *artists*. For this reason, we expect to see a significant effect of privacy implications on digital content curation, though the impact of this effect should not outweigh the other social and utilitarian uses of digital content curation considered in the study. Thus, as users feel that their curated content is *less* private and personal, they should curate more content. Based on this we propose the following hypothesis:

H9: Lower expectations of privacy will be positively associated with digital content curation intentions.

3.3 Methods

3.3.1 Research Instrument Development

The measurement scales for all the research constructs in this study were either taken from existing scales or adapted slightly for this research. For items that were adapted from existing scales, care was taken to select items that had been previously validated using established and rigorous means. All items were measured using 7-point Likert type items anchored between strongly disagree and strongly agree.

The measurement items for desire for self-presentation and virtual community involvement were adapted from Kim et al. (2012). The measurement items for serendipitous information discovery were adapted from McCay-Peet and Toms (2011) and Björneborn (2008), in which the main purpose of both these studies was the specific development and validation of an instrument for serendipitous information search and discovery. The measures for the personal information management constructs of information reminding and re-finding were adapted from Liaw and Huang (2003), which itself was an extension of an early seminal work in this area by Malone (1983). The work of Liaw and Huang (2003) focused on the testing and measurement of the constructs identified in Malone's earlier work.

The measurement items for object-based self-efficacy were adapted from extant literature, with some adjustment to account for the object-oriented nature of the construct. Finally, the questions used to measure digital content curation intentions were adapted from Kim et al. (2012).

According to Straub (1989), instrument validation should take place before any other statistical investigation. As a first step towards instrument validation, we conducted a pilot test of the instrument using a convenience population of graduate students, all of whom had experience using the Pinterest website. Pilot participants were asked to address any inconsistencies or confusing questions, as well as provide feedback on the nature or wording of questions. This feedback was incorporated, and the instrument was then made available for large scale distribution.

Construct	Definition	Source
Desire for self-presentation	An individual's desire to exert influence or gain benefits through presenting themselves in a favorable way	(Kim et al., 2012; Goffman, 1959, Leary, 1996)
Object-Based Presentation Self-Efficacy	Perceived efficacy believes around the use of objects for self-presentation	(Doyle et al. 2005; Compeau and Higgins, 1995; Döring 2002)
Virtual Community Involvement	The perceived level of commitment and care that a user has for the health and continued life of a virtual community.	(Shang et al. 2006; Ellison et al. 2011)
Serendipitous Information Discovery	Activities related to locating interesting and unexpected digital content within networked information resources	(André et al. 2009; Foster and Ford, 2003)
Information Re-Finding	Perceptions of CBN system usefulness for the "re-finding" of information	(Liaw and Huang, 2003)
Information Reminding	Perceptions of CBN system usefulness for "reminding" the user of information	(Liaw and Huang, 2003;Malone, 1983)
Privacy Expectations	User perceptions that the digital content they add to the network is private	(Agarwal and Rodhain, 2002)
Digital Content Curation Intentions	Perceived user intentions to curate digital content within the CBN	(Kim et al. 2012)

Table 3. Construct definitions and literature sources.

3.3.2 Data Collection

Data collection for this study used an online, web-based questionnaire administered through Qualtrics [8]. Past research has shown that online questionnaires are appropriate for capturing data from target populations of internet users (Son and Kim, 2008). Participants for the online survey were recruited in a number of ways. Because the study was

[8] www.qualtrics.com

interested in users of Pinterest, which has only been around for two years, we knew that identifying Pinterest users would be a challenge. For this reason, respondent solicitation efforts began with a large population of internet users. These participants were asked to take the survey only if they were in fact Pinterest users. If not, then they were asked to find a person familiar to them that could take the survey instead. To further increase survey participation, the web survey link was placed on numerous individuals' Facebook and Pinterest pages.

Ultimately the survey was taken by 488 participants. The first question of the survey was used to pre-screen the survey participants to determine their level of Pinterest usage. Since the survey link proliferation was out of our control, we predicted that many non-users would encounter the survey. In order to preserve the integrity of the data and insure that our sample was indicative of true Pinterest users, we purposefully screened out those users that had never used Pinterest prior to receiving the survey. This was done using display logic commands in Qualtrics. The first question of the survey asked participants "How often do you use Pinterest?". Users that answered never were immediately taken to the end of the survey, and no further responses were recorded. All other users were allowed to continue with the remainder of the survey. Our prediction proved to be accurate, as 171 of the survey participants had never actually encountered Pinterest before. After excluding these individuals, we arrived at a working sample size of $n = 317$.

Of these 317 participants, 24% were male, and 76% were female. 64% were college students, and 35% were working either full or part-time. 47% reported computer use of more than 4 hours per day, with 46% reporting that more 3 hours a day are spent on the Internet. A full 72% of responded reported that they use social network sites every day.

Because of the nature of the data collection, it is not possible to determine what the true response rate was, since we have no way of knowing how many people actually were shown a link to the survey. However, for the people who did not select “never” to the question of using Pinterest, 100% of survey participants went ahead and completed the rest of the survey.

Consistent with Armstrong (1971) we checked for nonresponse bias by comparing early responders with the late responders. No significant mean differences were found between the two groups on either the demographic questions or the variables of interest, indicating that nonresponse bias was not a problem for this data set.

3.4 Analysis and Results

3.4.1 Measurement Model

Using Straub (1989) as a guide, our first step upon completing the full data collection involved the examination of various measures of content validity. For an instrument to exhibit *content validity*, it is expected that the instrument consists of questions that are representative of the entire domain of research around the construct. Of all the types of validity, content validity is likely the most difficult to attain. This is due to the fact that many constructs ultimately have a near-infinite possible domain. For this reason, the best, or at least the most accepted, means of achieving content validity is to perform a thorough review of the extant literature. The instrument used in this study, while not necessarily representing. a complete and thoroughly exhausting review of every potential resource available, was put together in a systematic way, and all the items used for the instrument have been used and validated in at least one other study.

Construct validity determines whether measures are consistently measuring what we would expect. If a measure for a particular construct exhibit high correlations with other items measuring the same construct, and at the same time low correlations with items measuring a different construct, this is a good indication that the constructs measured by the instrument are in fact valid and real. To assess construct validity, we conducted a confirmatory factor analysis (CFA) in STATA, the results of which are detailed in Table 4. During CFA, measures that correlate within the same construct are said to exhibit convergent validity, while low correlations across constructs are an indicator of divergent validity (Straub, 1989).

The 9-factor measurement model was used to assess the quality of the constructs. Consistent with the CFA method, each item was set to load only on its pre-determined factor, and the factors were all allowed to correlate. The model in STATA converged to an acceptable level. Table 4 shows the means and standard deviations for each factor, along with the cronbach alpha and composite reliability measures, average variance extracted (AVE) and factor correlations. Each individual factor loading is displayed beside the item in Table 4.

Different fit indices all indicate a reasonable fit of the model to the data. Each of the following fit indices for the model (χ^2 (744) = 2578.580) were above their recommended thresholds. $\frac{\chi^2}{df} = 3.46$, root mean square error of approximation [RMSEA] = 0.094, standardized root mean square residual [SRMR] = 0.0960, normed fit index [NFI] = 0.947, comparative fit index [CFI] = 0.962, and goodness-of-fit index [GFI] = 0.682.

CFA methods are most effective when paired with measures of *reliability*. Reliability measures can detect methods of systematic variable within the way that participants answer the questions related to a particular construct. A very common method for assessing reliability is Cronbach's alpha (Cronbach, 1951).

The practical purpose of cronbach's alpha is to assess how well respondents answers group together around a particular construct. For an instrument to have high reliability, as measured by Cronbach's alpha, respondents must answer all items related to a particular construct in roughly the same way. For example, if a likert-type scale consists of 4 items, and participants consistently answer "strongly agree" for the first three, and "strongly disagree" for the fourth, there may be a reliability problem around this fourth item. While cronbach's alpha is consistently reported as a measure of reliability, it is not in and of itself sufficient to assess homogeneity among items (Sijtsma, 2009). We therefore also considered several other measures to further assess instrument reliability.

The instrument was next checked for convergent and discriminant validity. These two types of validity assess two questions. First, do things go together that ,should. Second, do things that should be separate actually stay separate. Convergent reliability is meant to ensure that measures that should be related in fact are in reality related (Campbell and Fiske, 1959). One of the main purposes for a CFA is the establishment of convergent reliability. It was assessed here by comparing each individual standardized factor loading with the recommended cutoff value of 0.60. A loading above 0.60 implies that it is reasonable to assume that this item does in fact belong with the other items used to measure a particular construct. The lowest observed standardized factor loading was 0.58 for an item used to measure serendipitous information discovery. While this loading is beneath the established threshold of 0.60, because the construct did not show problems in other areas of the measurement model, we decided to keep this item in the structural analysis.

Discriminant validity is meant to evaluate whether items that should be unrelated are in fact unrelated. Discriminant validity was assessed by comparing the square root of the average variance extracted (AVE) with the inter-construct correlations. AVE captures a measure of the amount of error that remains in the items after imposing a certain factor structure upon the measures. Generally, an AVE less than 0.5 is adequate, indicating that the imposed factor structure has captured more than half of the total error (Fornell and Larcker, 1981). As shown in Table 4, the square root of the AVE was greater than any of the correlations between the constructs.

						Correlations							
Variable	**Mean**	**SD**	**Alpha**	**CR**	**AVE**	**1**	**2**	**3**	**4**	**6**	**7**	**8**	**9**
1. DSP	4.49	1.67	0.92	0.98	0.77	**0.88**							
2. OBSE	5.08	1.58	0.87	0.98	0.71	0.61	**0.84**						
3. ICDC	5.67	1.72	0.97	0.99	0.92	0.39	0.57	**0.96**					
4. SEREN	5.71	1.40	0.91	0.98	0.58	0.34	0.43	0.51	**0.76**				
5. REFIN	5.72	1.40	0.92	0.97	0.79	0.35	0.50	0.54	0.62	**0.89**			
6. REMIN	5.41	1.56	0.93	0.98	0.81	0.37	0.45	0.45	0.57	0.72	**0.90**		
7. VCI	4.39	1.69	0.91	0.98	0.75	0.47	0.46	0.45	0.52	0.59	0.82	**0.87**	
8. PRIV	5.37	1.61	0.89	0.96	0.63	0.30	0.40	0.45	0.49	0.50	0.49	0.44	**0.79**

Notes:
1. Diagonal elements show the square root of AVE.
1. DSP = Desire for self-presentation; OBSE = Object-based Self-efficacy, ICDC = Intentions to curate digital content, VCI = Virtual Community Involvement

Table 4. Construct reliability measures and cross correlations.

3.4.2 Structural Model

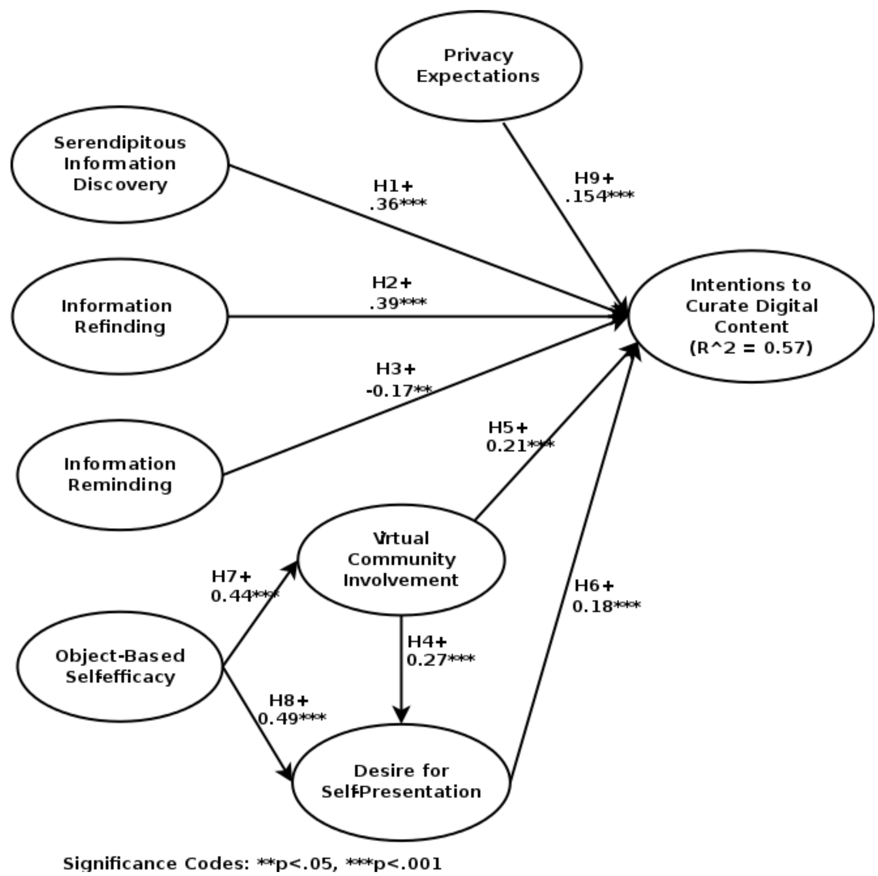

Figure 11. Structural model estimation results.

A structural model was produced using the SEM package in STATA. Desire for self-presentation and virtual community involvement were modeled as second order constructs, taking object-based self-efficacy as an antecedent. Additionally, a relationship was modeled between virtual community-involvement and the desire for self-presentation. Each of these second-order constructs was linked to intentions to curate digital content. The three personal information management constructs of serendipitous information discovery, re-finding and reminding, were also linked to intentions to curate digital content. All exogenous constructs were allowed to covary.

The overall fit statistics for this model indicated good fit to the data. Fit indices for the model (χ^2 (305) = 1001) were as follows: χ^2/df=3.83, SRMR=0.071,RMSEA=0.72, CFI=0.95, TLI=0.94.

Figure 11 reports the path coefficients and t-statistics, together with their significance values and R^2 values for each of the relationships in the model. Based on the structural analysis, all of the relationships in the model were significant. However, the direction of the relationship from remind to intentions to curate was negative, contrary to our expectations. The R^2 value for intentions to curate digital content was 0.57. Table 5 summarizes the results of the structural analysis.

Hypothesis	**Coef.**	**Z value**	**Supported?**
H1:	0.36	6.29^{***}	YES
H2:	0.39	4.96^{***}	YES
H3:	-0.17	-2.60^{**}	No
H4:	0.27	6.25^{***}	YES
H5:	0.21	4.89^{***}	YES
H6:	0.18	4.21^{***}	YES
H7:	0.44	10.70^{***}	YES
H8:	0.49	12.23^{***}	YES

Table 5. Results of hypothesis testing for study 1.

3.5 Discussion

This paper offers several important findings about the new and interesting phenomena of digital content curation validated with a survey-based analysis of over 400 actualcuration-based network users. In our theoretical development, we argue that intentions to curate digital content is best explained as a dual process that takes into account elements of digital content management and the display and endorsement of digital content in a social setting. We find support for the role of both of these areas in determining usage intentions. This finding provides some evidence that individuals curate content not only because they find the process of collecting and managing digital content interesting and useful, but because they want to show off their collections to other network users. Specifically, users curate content because they want to use the content to present a desired impression of themselves within the social network. They also collect content because they feel a connection and sense of involvement with the online social network.

In an interesting caveat to this finding, we found a significant coefficient for our construct of information reminding, though the direction of the coefficient was contrary to our hypothesis. While our analysis showed no indication of measurement error in evaluating this construct, it may be that the nature of the construct itself makes it hard to capture on a survey instrument. For example, asking someone how often they are reminded by a system may be difficult, since reminding is by definition an involuntary process and one that happens with little awareness on the part of the user. This provides an interesting opportunity for future research, as information reminding is no doubt an important part of personal information management, and one that could benefit from greater exposition and further treatment in the literature.

Second, we find that, while privacy implications have a role to play in digital content curation, they are a secondary concern for many users. Of all the constructs in the model, expectations of privacy had the lowest impact (i.e. the lowest β coefficient of any in the model). We interpret this finding as evidence of the separation between the curator and the content within these networks. Since the digital content is more an expression of what users like, rather than who they are, it may be natural to assume a certain distance from the content on the part of the users, which may account for the observed lower effect of privacy expectations.

It remains to discuss the specific literature contributions of the study. The paper finds support that the desire for self-presentation and virtual community involvement are instrumental in inspiring people to collect digital content, thus contributing to the stream of research around the acquisition of digital items online (Kim et al., 2012). This finding offers some interesting theoretical implications. Recently, several papers have been published that explore the novel and unique motivations that lead people to form online self-expressions different from their offline identities (Hinz et al., 2010; Kim et al., 2012).

These studies have consistently shown that not only is the desire for identity formation and self-presentation strong enough to encourage usage, but that people will pay real money to acquire goods with virtual self-presentation value. In our study, we also find support that the desire for self-presentation plays a large role in determining user motivations, and that users truly exhibit effort to go out and acquire the tools, in this case digital content, with which to forge their desired self-presentations. This finding is particularly interesting in light of the fact that traditional online social networks continue to struggle with pay-to-play pricing models, while platforms for digital self-presentation continue to be profitable. The question of how users value their true identity compared to their self-created representations, and how these identities factor into user purchase decisions, is certainly worthy of attention by the academic community.

The study also expands and contributes to extant literature on online social networks. Overwhelmingly, the majority of research into online social networks considers usage as a function of social benefits. Studies that incorporate elements from social capital theory (Ellison et al., 2007a), social exchange theory (Thambusamy et al., 2010), friendship theory and other socially grounded theories may fail to capture the utilitarian side of digital content curation. This poses a significant theoretical problem as our statistical analysis demonstrates that content management utility plays a large role in determining overall usage intentions.

The study makes a novel contribution to the literature concerning the role of privacy in online social networks. Whereas privacy plays an almost constant role in most studies of online social networks (Utz and Kramer, 2009), and rarely a week goes by without change and consternation around the privacy policies of these technologies, we find that privacy implications, while important, play a much less significant role within networks for digital content curation. The study raises some interesting questions about the role of digital content for self-expression and the separation between *content* and *self* in online social networks. This may prove to be a fruitful area for future research, as many recent papers and theories in the IS literature (eg. (Son and Kim, 2008)) are focused on the protection and privacy of personal information, and make little mention of rich media privacy.

In fact, in much of the extant literature, photos and personal information are considered as synonymous (Liu et al., 2011). A paper by Ahern et al. (2007) is one of the few that does not do this, by considering the privacy level of different photo categories. However, because content curation networks revolve around a user's interest and not necessarily their personal information, many of their photos may not feature prominently, or even relate to them directly. This potentially open up a new area around interest-based privacy, or privacy around the things one owns and cares about, rather than privacy exclusively about the individual.

Finally, the study makes a contribution to the literature around eWOM. Recent papers studying eWOm in social networks (Garg et al., 2011; Susarla et al., 2012b) have shown that eWOM in a textual form is a powerful means of influencing behavior, tastes, and interests. This is an exciting area of research and one that online companies care about immensely. To date, the extant eWOM research has focused almost exclusively on textual communication as the means by which eWOM diffuses. In this study, we extend the boundaries of eWOM literature into the area of digital content curation. This has significant potential as a future research stream, for as digital content curation technologies become more pervasive, their capacity to spread profitable eWOM marketing information will increase exponentially. Additionally, because the rich media shared through digital content curation is so data rich, and preserves hyperlinks back to the originating websites, a piece of content has the ability to open up a new referral channel every time it passes from one digital collection to another.

CHAPTER IV

STUDY 2: PIN IT TO WIN IT

4.1 Research Gap and Motivation

Social media platforms have diffused to the point of ubiquity in modern society (Parameswaran and Whinston, 2007). It is not uncommon today to see celebrities, companies and brands, even the Pope [1] using blogs, Twitter and Facebook. Not surprisingly, standing out in these crowded platforms has become big business (Mangold and Faulds, 2009). Those lucky few individuals that are able to attract large numbers of interested followers have the potential to become celebrities overnight. Companies also stand to benefit, for social computing platforms have the power to spread marketing information cheaply and instantaneously to incredible numbers of people (Dellarocas, 2003). The past few years have seen some truly startling examples of the monetary value attached to social media reputation.

For example, consider the story of *Instagram*[2] Instagram makes an application for sharing editing and sharing digital images. It works on mobile devices, and integrates easily with social network sites like Facebook and Myspace. Instagram benefited mightily from its online reputation when it was purchased by Facebook for $1 billion while still less than two years old. By way of comparison, the *New York Times*, at 160 years old, is valued at just under $1 billion [3]. the time of purchase, Instagram had very little in the way of a tangible business model, and small revenues. What it did have was a great reputation

[1] http://www.cnn.com/2012/12/12/world/europe/vatican-pope-twitter
[2] http://www.instagram.com
[3] http://thenextweb.com/insider/2012/04/09/

and a large base of loyal users. When Facebook purchased the company, CNN reported that they had just bought a company with "lots of buzz but no business model"[4].

Interested followers in online social media environments represent a valuable asset for just about any organization, regardless of industry or occupation. When cholera broke out in Haiti following the 2010 earthquake, social media buzz around the spread of the disease was able to track the outbreak two weeks ahead of official reports (Chunara et al., 2012). Yet, despite the very practical value of social media buzz, its antecedents, and the factors that allowing one company to become a billion-dollar Instagram, while others toil away in obscurity, are still poorly understood (Susarla et al., 2012a).

It is not surprising that attracting followers represents a challenge. Social media technologies change rapidly, with each new iteration offering new features and implementing the latest technologies. One recent advancement involves the ability of online social networks to incorporate and organize digital content from other online sources through a process known as *digital content curation* (Yakel, 2007). This has created new aspects of online social networking, in which most of the communication between network users exists through a process collecting, managing and sharing digital content.

While past research has examined the social implications of online social networks Ellison et al. (2007b), the implications of digital content curation within these networks have gone largely unexplored. We study these "curation-based" online social networks (CBN), and explore the impact of the digital content sharing promoted by the process of digital content curation. Our goal is to examine how users of CBNs leverage digital content to generate "buzz", as measured by their number of CBN followers. Specifically,

[4] `http://money.cnn.com/2012/04/09/technology/`

we examine the research question: *How do CBN members use digital content curation to attract followers?*

The study has three primary motivations. The first is to understand the marketing implications of curation-based networks, which are literally built from rich-media digital content. As CBN users engage in digital content curation actions within these CBN, they are generating, modifying, and spreading digital content at at amazing rate. This makes CBN extremely important sources of online referrals for companies that produce content for distribution online. For example Pinterest[5], the fastest growing and most popular curation-based network, generates more referral traffic than social network sites LinkedIn, Google+ and YouTube combined [6]. Pinterest ranks third among all websites, and sits behind only Google Websearch and Facebook, in terms of driving online referral traffic, despite having a user-base only 1% the size of Facebook [7]. However, Pinterest and other CBN are relatively new technologies, and as a result this phenomena remains largely unexplored.

The second motivation for the study stems from a desire to contribute to a stream of research that examines the role of *socially-earned* media and marketing outcomes. The rise of Web 2.0 technologies in general, and online social networks in particular, has introduced some fundamental changes into the marketing practices of many organizations (Trusov et al., 2009). Successful online marketing initiatives today typically involve the use of traditional media (eg. advertisements) that in turn engender *socially earned media* (Stephen and Galak, 2012). Socially earned media consists blog posts, online customer reviews, Tweets, and any other word-of-mouth communication about a particular person,

[5] www.pinterest.com
[6] http://techcrunch.com/2012/03/08/
[7] http://blog.shareaholic.com/2012/03/

company, or brand (Dellarocas, 2003). Recent research efforts have begun to identify the important role that socially earned media plays in generating buzz and obtaining customers. Still, while traditional media has a long history of association with desirable marketing outcomes, the impact of socially earned media is not well-understood (Stephen and Galak, 2012).

Third, digital content curation potentially represents a major shift in the nature of media delivery and consumption. Much of the content that populates the internet is produced by companies interested in attracting consumers to their websites for the purposes of generating advertising revenue (Jarvis, 2008). These content producers desperately depend on a steady stream of visitors. Yet today, a huge portion of online revenue is generated by *content aggregators* like Google and Yahoo, who direct visitors to content producers' sites, earning revenue along the way.

In the first half of 2012, Google earned $20.8 billion in advertising revenue, more than all print media advertising in the United States combined [8]. As the popularity of CBNs continues to grow, the internet will be populated by literally millions of content aggregators sharing and distributing digital content in virtual spaces not owned by content producers, and in which content producers have no power to show advertising and generate revenue. Content producers must learn to utilize these technologies and take ownership of their content within the CBN space. The current study will be of interest for this group, in that it delivers tangible strategies that actual content producers are actively using today to attract followers in a real-world CBN. In sum, the objective of this study is to examine the way that individuals within CBN use digital content curation actions to collect and manage digital media for the purpose of maximizing their follower base. We

[8] http://www.slate.com/blogs/future_tense/2012/11/12/

theorize that this process depends on a marketing mix of three key elements. 1) Advertising that incorporates the right kind of traditional media advertising content in the correct amounts. 2) generating a socially-earned media response to your advertising efforts, and 3) engaging with other network users in reciprocal digital content exchanges.

To test this theory, we empirically analyze several econometric models comprised of panel data collected from a leading CBN. We collect and analyze data on all of the media and actions of over 1800 Pinterest users over the course of 8-weeks. The results of this study should be interesting to academics who want to better understand the generation and spread of digital content through an online social network. Practitioners will benefit from gaining a practical understanding of what works, and what doesn't work, when designing a strategy for buzz generation within an online social network.

In the next sections, we provide details on the domain of study for this work. We then discuss the extant research on the impact of media on marketing outcomes. Next, we discuss the data collection and our variables of interest. We present an econometric analysis of panel data collected from the Pinterest CBN. The results are discussed, together with some limitations of the study and directions for extending the research.

4.2 Media Within Curation-Based Networks

Curation-based networks are online social networks in which people form network links around shared interest in digital content. Users of these networks append digtial content to "virtual pinboards" by means of a browser applet. This applet allows users to collect content (usually pictures) from websites all over the Internet, copy the content to their boards within the CBN. Users visiting these boards can then view the images as one stream, with new images added as the user scrolls down the page.

CBN boards have some striking differences from the online social network pages seen on sites like Facebook, MySpace, or Google+. First, content on these boards is typically representative of things that are of interest to network users, without being *directly about* them. Boards contain very little personal information, and many users do not use the CBN to display pictures of themselves, rather CBN boards represent a "best of the web" for each user.

Many companies also are members of the CBN and maintain CBN boards. Companies are free to add digital content from their company websites, and interact and network with other CBN users, users that they hope to turn into followers, and ultimately into paying customers. "Better Homes and Gardens" [9], an online home improvement and lifestyle magazine, manages a number of boards within the Pinterest CBN that cumulatively have more than 300,000 followers. The company freely distributes large amounts of content 12
through the Pinterest network in an effort to drive traffic to its company-owned websites where it can sell products and generate advertising revenues.

In order to attract followers, users like Better Homes and Gardens employ a number of different *digital content curation* actions. Digital content curation refers to the selection and maintenance of a carefully chosen set of digital assets (Yakel, 2007). CBN provide a number of different tools for digital content curation, which allow for a robust number of content collection and management actions (table X). Taken together, these digital content curation actions make up the user's marketing mix within the CBN.

[9] `http://www.pinterest.com/bhg` or online at `http://www.bhg.com`

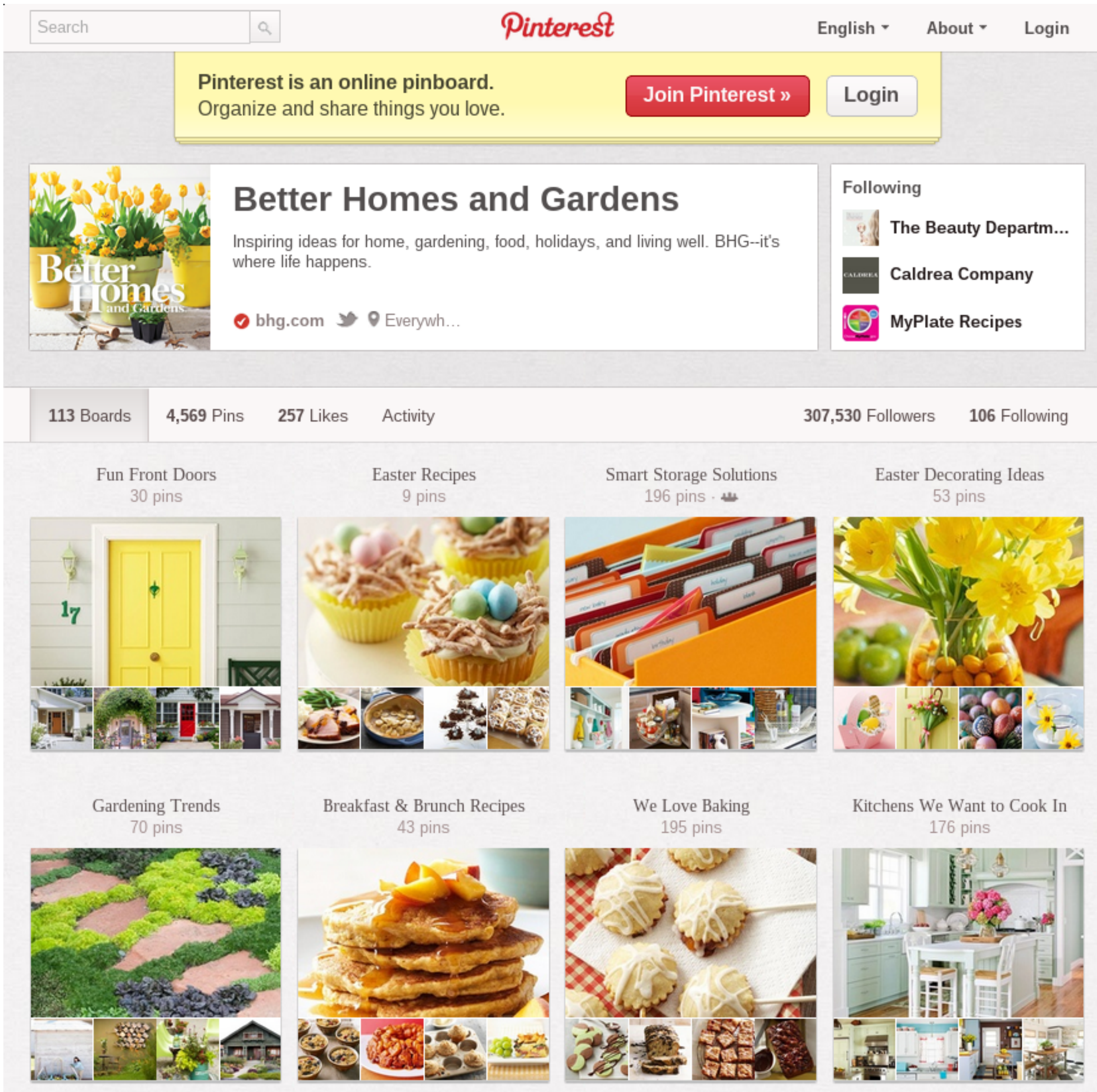

Figure 12. Better Homes and Gardens page within Pinterest.

4.3 Theory and Hypotheses

4.3.1 Traditional Media and customer growth

The extant marketing literature has focused on the role of traditional media in marketing outcomes Corcoran (2009). Traditional media refers to any advertising by a firm, for example advertisements taken out on television, radio, or in magazines. Traditional media has existed for many years, and as such has a storied history of research in economics sales and marketing. Traditional media success has been linked to increased sales (Assmus et al., 1984), customer engagement and satisfaction (Anderson and Sullivan, 1993), and customer acquisition and retention (Fornell and Wernerfelt, 1987).

There may be a number of reasons that traditional media activities attract followers. First, and perhaps most obvious, is that traditional media represents the primary method by which companies tell customers about the products that they have for sale (Aaker and Biel, 1993). Within CBN, we define traditional media as the digital content that together makes up the digital self-expression of a company or individual user. It is the combined total of all the digital content that populates the pages that users present to the world. For example, the pictures displayed in Figure 12 all advertise products either for sale or related to articles featured on Better Homes and Garden's commercial website. When users visit the CBN pages of a particular CBN user, their encounter involves the traditional media that the user has placed in that location.

H1a: More traditional media on CBN pages is associated with more CBN followers.

As with traditional forms of advertising, all digital content is not created equal. Obviously some advertising campaigns flop, while others flourish. However, within CBN it is possible to identify content that may be new or fresh. CBNs allow content to be uploaded from an outside website or shared from one user to another. Content uploaded from outside the CBN may be unknown to a larger number of users and therefore more interesting or compelling. Conversely, content that is copied from another CBN user (or repinned, to use the CBN terminology) exists on at least one more page within the CBN. As a result, users may have other avenues through which to view this content, or they may have seen it previously on another user page. Thus, we hypothesize that original content, i.e. content newly uploaded to the CBN, will be more valuable for attracting followers that that copied from other CBN users, giving us the following hypothesis.

H1b: Original CBN content will be more strongly associated with CBN followers than copied content.

As with traditional forms of advertising, all digital content is not created equal. Obviously some advertising campaigns flop, while others flourish. Marketing researchers have struggled for decades to identify exactly what makes a traditional marketing campaign "pop". Such questions are outside the scope of this research, and we are unable to evaluate whether one type of content is objectively better than another.

However, within CBN it is possible to identify content that may be "new" or "fresh". CBN allow content to be both uploaded via the previously mentioned browser applet, as well as shared across users. Content uploaded via the browser applet is new to the CBN. As such, it may be unknown to a larger number of users and therefore more interesting or compelling. Conversely, content that is copied from another CBN user (or *repinned*, to use the CBN terminology) exists on at least one more virtual board within the CBN. As a result, users may have other avenues through which to view this content, or they may have seen it previously on another user board. Thus, we hypothesize that original content, i.e. content uploaded via the CBN browser applet, will be more valuable for attracting followers that that copied from other CBN boards, giving us the following hypothesis.

4.3.2 Socially Earned Media

Socially-earned media in a general sense consists of digital communications about a company, person or brand appearing in digital spaces such as blogs, social network sites and forums (Kwai Fun and Wagner, 2008; Parameswaran and Whinston, 2007). Socially-earned media can consist of positive blog posts about a company, Facebook discussions around a company or the actions that they take, or other user-generated content like online customer reviews on Amazon.com (Stephen and Galak, 2012). In this study, focus on the socially-earned media that arises within the CBN system. Socially-earned media in this context refers to any comments that other users post about the traditional media content of other users. Socially-earned media is also generated when users show support for traditional media content by copying it from another user's page onto their own.

Socially-earned media can impact CBN followers in several ways. First, recent research has suggested that because socially-earned media typically involves an endorsement from trusted sources, for example friends or known online social network members, the message carried by socially-earned media may be highly influential (Chu and Kim, 2011). Additionally, socially-earned media may be able to reach more selective pockets of interest than traditional media (Stephen and Galak, 2012). Traditional media is distributed through channels with massive exposure.

However, traditional media channels may not reach customers who are especially engaged with a product or a brand (Stephen and Galak, 2012). By contrast, socially-earned media is very good at dividing customers into groups of individuals who are very engaged around niche topics. Customers in these groups are often very committed and loyal to a particular product or brand. In this way, socially-earned media helps companies to reach the long-tail. This is especially true within curation-based networks, where digital content can be collected from many users quickly, and may be seen by users that have never visited the pages of the one who originally pinned the content. Because these environments are designed to produce and spread socially-earned media, we hypothesis that not only will socially-earned media lead to more network followers, but users will also see a greater benefit from socially-earned, rather than traditional media.

H2a: More socially-earned media will be associated with more CBN followers.

H2b: Socially-earned media will be more strongly associated with CBN followers than traditional media.

When considering socially-earned media, it is important to understand that the effects of socially-earned media may exhibit differences when compared to traditional media. Specifically, socially-earned media may be dependent on time. Because socially-earned media relies on a response from the CBN community, we can anticipate that socially-earned media generation will lag behind the traditional media marketing efforts of a CBN user. Socially-earned media may also benefit from the effect of an information cascade (Watts, 1999).

Information cascades occur as a message is passed on from one person to another (Bikhchandani et al., 1992). With each step in the diffusion of the information, new network links open up so that information can progress faster and faster. Eventually, the number of nodes for information diffusion reaches a critical mass, and information diffusion occurs very quickly. We may expect therefore that the effect of socially-earned media would show an increasing relationship with time. From this we propose the following hypotheses:

H2c: Socially-earned media will exhibit an increasing exponential relationship with time.

4.3.3 User Engagement

We define user engagement here as the number of following relationships that a particular user has entered into with other CBN users. One of the strengths of online social networks in general, and CBN in particular, is that they also allow for two-way user interaction. To use a company example, they allow a company to follow its customers, just as the customers follow the company in turn. In this way, online social networks decrease the distance between company and customer. They break down the barriers between companies and their customers and promote interaction not possible offline (Trusov et al., 2009).

For example, within an online social network, companies can deliver advertising messages to customers, and then receive immediate customer response to the advertising message. When companies take the extra step to follow not only their own current customers, but also potential customers, they may realize a number of tangible marketing benefits. They are able to see the interests of these users, and observe their interactions with each other, as well as with other companies inside the CBN.

Studies have shown that observing and responding to customer feedback makes customers feel that they have a voice within the company, and leads to higher levels of customer loyalty (Gallaugher and Ransbotham, 2010; Hart and Sharma, 2004). A strategy of customer engagement within social media "underpins brand positioning and perception, establishes a clear message, conveys corrections, promotes contests, distributes time-sensitive information, and even recruits customers, partners and staff." (Gallaugher and Ransbotham, 2010, p. 199). From this we hypothesize that those users that show a greater degree of user engagement will have more CBN followers.

H3: User engagement will be positively associated with CBN followers.

4.3.4 Content Producers and Aggregators

We are also interested in looking at some group differences between content producers and content aggregators. We define content producers as those users that primarily use the CBN to distribute content that they produce themselves or content that they have purposefully collected from outside of the CBN. Content aggregators, by comparison, primarily fill their pages with content collected from other CBN users.

In the network economy, a content producer like the New York Times creates content that reports on news stories and events. They produce a huge amount of content that is consumed by customers through a variety of different media channels. Content aggregators, of which Google News is a well-known example, produce no content, yet aggregates the content of the New York Times and other news outlets to provide a "one stop shop" for people looking for news (Dellarocas et al., 2013).

Examining the issue of content production vs. content aggregation is interesting and timely, given that while the web is populated through the advertising efforts of content producers like Better Homes and Gardens and the New York Times, the vast majority of online revenue goes to content aggregators like Google (Jarvis, 2008). CBN have profound implications for the future of content aggregation. Specifically, CBN place the tools for content aggregation into the hands of individual CBN users. This has the effect of greatly increasing the number of content aggregators in the digital marketplace. However, the nature of CBN content curation preserves links to originating websites, so that while CBN users aggregate and share digital content, they do create socially-earned media for content producing websites. This ultimately means that content producers need to become active within the CBN community, so that they can take ownership of their content and its distribution. In this way content producers can create a more desirable network position from which to attract followers and convert these CBN followers to actual paying customers. We can expect to see some reasonable differences between content producers and aggregators, simply based on their respective natures. In terms of traditional media, we expect that followers attributable to original content will be higher for content producers. This is because they rely more heavily on new and original content as opposed to content aggregators. Conversely, content aggregators should see greater benefits from copied content.

H4a: Follower numbers for content producers will be impacted more by original content, rather than copied content.

H4b: Follower numbers for content aggregators will be impacted more by copied content, rather than original content.

Finally, we examine the role of user engagement on content producers and content aggregators. As stated above, user engagement provides a means of following and keeping track of other users and the CBN community. While this ability is no doubt valuable to both content producers and content aggregators, aggregators in particular have a true need to track and manage the content curation activities of the CBN community. This is because the pages of content aggregators are populated largely by content posted by other users, and for this content to be as relevant, timely and enticing to potential followers as possible, it is important that the content be compiled from good sources. For this reason, we propose the following hypothesis.

H5: User engagement will be more strongly associated with the followers of content aggregators than content producers.

4.4 Methods

4.4.1 Data Collection

To test the hypotheses presented above, we collected data from Pinterest, the largest and fastest growing curation-based online social network. The data was compiled from publicly-available information over a period of eight weeks from September to November of 2012.

The data set consists of all of the "actions" taken by a sample of 1820 Pinterest users. Actions here refers to all of the activity taken by each user in terms of posting content to Pinterest, organizing this content, and connecting and communicating with other Pinterest users. In addition, we collected data on the community response to the actions of these users. While the unit of analysis here is the Pinterest user, it was necessary to collect this data at the individual "pin" level of analysis. Collecting data in this way allows us to see the way that the Pinterest community responds to the content and activities of the users in the sample. Ultimately, these pins were all attributed to a user in the sample, which provides a sum total of the community response for *all* the user activity for a particular period.

All told, the sample included over 1 million pins that were examined each week for a period of eight weeks. To collect this data required the use of webcrawling applications developed using the Scrapy [10] development framework. Scrapy provides and extends several Python classes that specialize in webcrawling and information retrieval.

[10] `www.scrapy.org`

Webcrawling techniques make use of “spiders”, programs designed to visit webpages and *scrape* data for further processing and analysis. Designing a webcrawling is a two step process that consists of creating rules for the way that spiders crawl hyperlinks, and rules for matching the text that needs to be collected. Crawling rules tell the spider which type of links to follow, and which to avoid. Great care must be taken when writing crawling rules to avoid creating a recursive loop that would lead to the spider crawling the same pages over and over. This is one area where Scrapy greatly improves upon most webcrawling applications. Scrapy maintains a list of all visited pages, and before crawling a page, Scrapy checks this list to see if the page has already been visited. If an entry for the page exists, then Scrapy will refuse to crawl the page, ensuring that recursive loops are avoided in the crawling process.

To aid with creating rules for crawling and text matching, Scrapy allows for two different text-matching languages, regex and Xpath. Regex selectors use the traditional system of matching text based on regular expressions. Regular expressions have been an important part of applications development for several decades. Xpath, meanwhile, is a newer language that selects nodes within XML documents, and can also be used to identify blocks of code written in HTML. Xpath is not as versatile as regular expression matching, but it is simpler, and when the text to be matched is well-formed and consistent, Xpath selectors can be very powerful. An example of an Xpath selector used in this study is `hxs.select('/html//div[@id="PeopleList"]//a/@href')`. In this example, the Xpath selector tells the spider to first search through the HTML code for a div with the id “PeopleList”. Once found, the spider is to follow every hyperlink within this div. In practical use, this allows the spider to search through all of a particular social network user’s followers, visiting each follower page in turn.

When Scrapy scrapes an item from a webpage, the document is stored using JSON(JavaScript Object Notation) text. JSON is a Javascript-based markup language that is designed to transmit data between servers and web-based applications. As such, it is used frequently by modern APIs (Application Programming Interfaces) on websites such as Twitter and Wikipedia's MediaWiki service [11]. Because of its popularity in web-based application settings, JSON is the standard output format provided by Scrapy.

To store the items as they come from Scrapy, we used the open-source MongoDB [12], a lightweight non-SQL database that is designed to handle data in the JSON format. Finally, in order to analyze the raw data from Scrapy, standard Linux tools including sed, awk and grep were incorporated into a series of bash scripts to clean the data into further analysis within the STATA statistics package.

4.4.2 Variables

Pinterest provides for a fairly large number of different user actions related to the curation of digital content. More importantly, the confines of the Pinterest system make it possible to track each action taken on the part of a Pinterest user, as well as how the other users within the Pinterest network respond to these actions. In the following sections, we discuss the way that each of the variables considered was operationalized for the study. First, we discuss our main dependent variable, follower network growth. Traditional media, socially earned media, and user engagement variables are then discussed.

[11] http://www.mediawiki.org/wiki/API:Main_page
[12] www.mongo.org

4.4.3 Followers

Pinterest provides a public count of a user's "followers". Followers within Pinterest are those other users that have decided to receive updates when a Pinterest user posts new content to their digital collections. Unlike some other social network sites, Pinterest users are allowed only limited communication with their followers. A user cannot send an email "blast" to all of their followers, for example. One advantage of this system is that for someone to follow another user on Pinterest does not require the user's permission, so people are free to follow as many users as they wish. This number of followers forms the dependent variable in our model. Past research has considered the number of a user's social network followers as a proxy for social media "buzz" or popularity (Susarla et al., 2012a).

4.4.4 Traditional Media

We consider two types of traditional media, original content and copied content (repins). Within a CBN, content can be posted to a user's page in two ways. Content can be uploaded by the user. We refer to this type of content as *Original Pins*. Original pins are those pieces of content which the user has brought into the CBN from outside, either by uploading the content from their computer or by copying the content from an outside website.

CBN users may also curate content from the content collections of other users. In the CBN terminology, this is called a *repin*. Repins are distinct from original pins in that the user is not the first to upload the content into the CBN, but rather merely copies the content from one CBN page to their own. We are interested in studying the impact of new, original content compared to repins. To answer this question, we need a way to distinguish between the two. Ultimately the solution to this problem involved a regular expression match while the pins are being crawled. Any pin added to Pinterest has a date stamp, that reads “Pinned x days ago” for pins that are original, and “Repinned x days ago” for repins. When crawling the pins, original pins were flagged with a boolean value of one, while repins were coded as zero.

4.4.5 Socially Earned Media

As users of Pinterest post content to their collections, the Pinterest community responds by copying, commenting on, and showing support for this content.Pinterest keeps track of all of these actions, so that each piece of digital content, and collectively each user, amasses a collection of socially earned media as they engage in digital content curation. We collected data at the pin level, so that the data set contains socially earned media numbers for each piece of content on a user’s page.

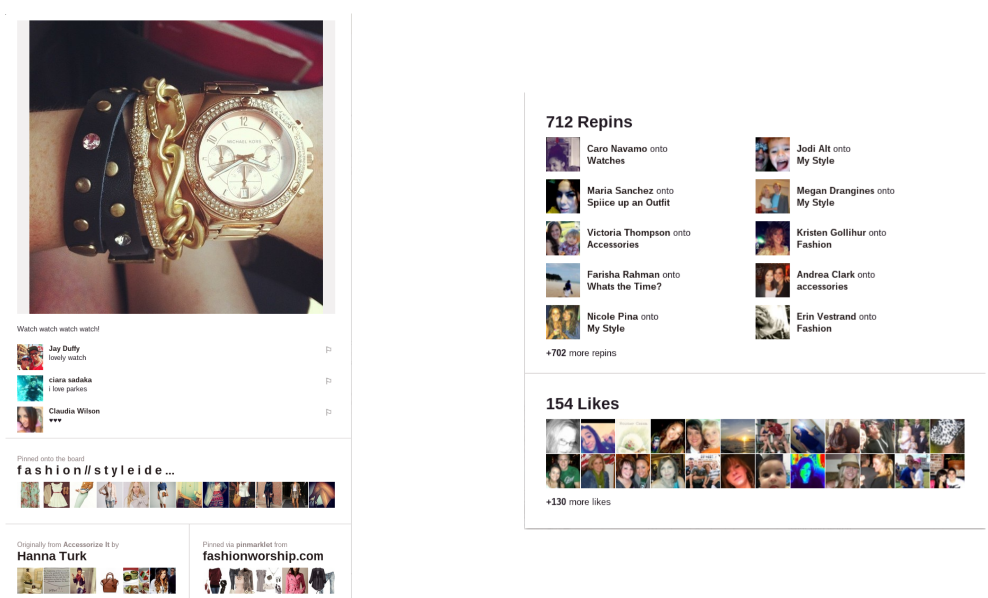

Figure 13. A Pinterest pin and the socially-earned media it has accrued.

Socially-earned Repins are the most common type of socially earned media within the Pinterest system. Just as a user can choose to repin content from another user with which to decorate their pages, a user's own content can be repinned in kind. The difference is subtle. A traditional media repin occurs when a user, for example user x copies a piece of content a from user y's collection. User x has now placed content a on their page, and user y gets credit for a socially earned repin.

Socially-earned Comments can also be added to any pin within another user's digital collections. It is not possible to add private comments to a pin. Any comments are visible to all other users. Interestingly, within the Pinterest system, the average number of comments collected by a pin is quite low (see Table 7). Pins are much more likely to accumulate repins. Across our entire sample, we observed 40 socially-earned repin for every socially-earned comment. Figure 13 shows a sample pin taken from one user's page. Note the presence of indicators for socially-earned repins and comments.

Variables of Interest	**Description**
Customer Acquisition	
Followers	Number of CBN users following this user.
Traditional Media	
Original Content	New content from outside CBN.
Copied Content	Content copied from other CBN user pages.
User Engagement	
Following Relationships	The other CBN users that this user "follows"
Socially Earned Media	
Earned Repins	Number of times content is copied by other users.
Earned Comments	Number of comments content has received.

Table 6. Description of how variables were operationalized for the CBN environment.

4.4.6 User Engagement

We consider a third type of media inherent to the Pinterest system. User engagement in this study is synonymous with the number of other CBN members that a user is following. Following users within the CBN offers numerous opportunities to observe the activities and communications of these users. Thus, following other users shows a commitment to the CBN community. Just as the follower count for any particular CBN user is public, so too are the number of others that the user is following in turn.

4.5 Empirical Analysis and Results

We begin our analysis by examining some descriptive statistics. Eventually our goal will be to analyze panel data in all periods simultaneously. Summary statistics for each of the variables are presented in Table 7.

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
Followers	279.593	611.968	10	13386	14790
Traditional Media					
Original Content	102.289	197.725	0	3864	14790
Copied Content	451.865	675.022	0	10994	14790
Socially Earned Media					
Earned Repins	1010.137	2771.428	0	67898	14790
Earned Comments	25.28	126.828	0	4594	14790
User Engagement	182.459	681.732	0	19113	14790

Table 7. Summary statistics for followers, traditional and socially-earned media, and user engagement

4.5.1 Descriptive Statistics

Within the data, a great number of users have a small number of followers. At the other end, some users have a great many followers. The same is true of all variables in the model. To address this problem of normality, we will consider the natural log of all variables in subsequent analysis. Natural logs are often used with secondary data when they do not impact the interpretation of questions of interest. Any recorded values of zero were retained as zeros after the log transformation.

4.5.2 Base Econometric Model

For ease of reading, we will not refer to the natural log of variables in the model or results. However, it should be assumed that we are working with the natural log of all variables. We assume that socially-earned media, because it requires a response from the user community, may in fact take some time to exert an influence on followers.

To model this effect, we need to introduce lag variables corresponding previous time periods. This is problematic, however, because the sum total of socially-earned media in one period is highly correlated with the total socially-earned media in another period. We can circumvent this problem, however, by considering variables that do not measure the aggregate of socially-earned media, but only the change in socially-earned repins and socially-earned comments from one period to the next. Note that by including the lag variables we reduce our observations, since it is not possible to lag the data beyond the first observed time period. We therefore begin by testing the following base model:

$$Followers_{i,t} =$$

$$\beta_0 + \beta_1(OriginalContent_{i,t}) + \beta_2(CopiedContent_{i,t}) + \beta_3(Following_{i,t}) + \beta_4(EarnedRepinChange_{i,t-1}) + \beta_5(EarnedCommentChange_{i,t-1}) + \gamma_i + \delta_t + \epsilon_{it}$$

where i and t represent a particular user at a certain time period. To control for any unobserved heterogeneity Ordinary Least Squares (OLS) regressions are used with fixed-effects (γ_i) (Greene, 2004). Panel analysis with fixed effects is appropriate here given that there is some risk of unobserved heterogeneity in the data. To test for the appropriateness of fixed effects, we performed a Hausman test that compares the model using fixed or random effects. The test statistic was significant, indicating that the fixed effects model is appropriate. We also control for any time trend by including a time period variable corresponding the current panel period (δ_t).

Additionally, it is likely that our model posses some heteroskedacitity and autocorrelation. Since the sample includes a large number of users, some of whom have very different amounts of followers, traditional media etc., it is reasonable to assume that we would see differences in the variance around these variables for different categories of users. For this reason, we fit a model that incorporates fixed effects with robust standard errors (White, 1980). Our dependent variable is the number of CBN followers in a particular time period. Independent variables in the model correspond to original pins and repins as traditional media, earned repins and earned comments for social media, and following relationships as a measure of user engagement.

The purpose of this base model is to show, in a general setting, how the total cumulative amounts of each variable impact a user's follower count. To test hypothesis H1a, we examine the coefficients for traditional media. We find that original content on a user's page in a time period impacts CBN followers in that period, providing partial support for hypothesis H1a. Specifically, original pins (β_1) had a coefficient of 0.005 ($p = .002$). Copied content (β_2) was not a significant predictor of CBN followers ($0.0007, p = 0.663$). Thus hypothesis H1a is partially supported. The coefficient for original content is significantly greater than that of copied content at a 95% level of confidence, indicating that the observed impact of original content was indeed greater than that of copied content, providing support for hypothesis H1b.

To evaluate hypothesis H2a, we must look at the socially-earned media coefficients for earned repins and earned comments (β_4 and β_5, respectively). Hypothesis H2a is partially supported. The change in socially-earned repins from period $t - 1$ to period t (β_4) had a significant impact on followers ($0.003; p < .001$). However, the coefficient for earned comments (β_5) was not significant ($-0.0003; p = 0.734$). Since copied content and socially-earned comments were not significant in this base model, testing hypothesis H2b simplifies to a comparison of the coefficients associated with original content and socially-earned repins. At the 95% level, the confidence intervals for these two coefficients overlap. Therefore in this base model we did not find support for H2b. Finally, we may evaluate hypothesis H3 by looking at the value of β_3. We observed a significant relationship between user engagement and followers ($0.112, p < .001$). This was the highest observed coefficient in this base model. All of these coefficients are summarized in Table 8, model 1.

4.5.3 Model Extensions

To evaluate Hypothesis H2c, it is necessary to evaluate the effect of socially-earned media over an extended length of time. Because we are working with a panel data set, we have the ability to create additional lag variables looking back several time periods. In this way we can model the impact of accumulated socially-earned repins and comments. This allows us to ask whether or not increased amounts of socially-earned media, earned at different periods of the data collection, exert an impact on a user's followers. To conduct this analysis, we create new variables that correspond to the change in socially-earned repins and socially-earned comments from period $t-1$ to period $t-2$ and from period $t-2$ to period $t-3$.

$$Followers_{i,t} =$$

$$\beta_0 + \beta_1(OriginalContent_{i,t}) + \beta_2(CopiedContent_{i,t}) + \beta_3(Following_{i,t})$$
$$+ \beta_4(EarnedRepinChange_{i,t-1}) + \beta_5(EarnedCommentChange_{i,t-1}) +$$
$$\beta_6(EarnedRepinChange_{i,t-2}) + \beta_7(EarnedCommentChange_{i,t-2}) +$$
$$\beta_8(EarnedRepinChange_{i,t-3}) + \beta_9(EarnedCommentChange_{i,t-3}) + \gamma_i + \delta_t + \epsilon_{it}$$

Recall that in our base model we found support for the impact of socially-earned repins, but not socially-earned comments. In this model, despite the presence of additional variables around socially-earned comments, the situation is largely the same. None of the socially-earned comment variables were significant. We did however observe that the impact of socially-earned repins is significant in all three lag periods (see Table 8, model 2). However, despite seeing a significant impact of socially-earned repins in each period, the coefficients remained relatively stable, and showed no indication of the hypothesized increasing trend. As a result, we did not find support for hypothesis H2c. This model does however offer some support for hypothesis H2b, that socially-earned media has a stronger impact on followers than traditional media. When we incorporate the cumulative effect of socially-earned media into our model, we see the role of traditional media decline, to the point that in this lag model the coefficients for both original content and copied content are not significant. User engagement β_3, however, continued to have a powerful effect in this model.

Our next extension to our base model involved looking at group differences between content producers and content aggregators (models 3 and 4 in Table 8, respectively). For the purposes of this analysis, we defined a variable for user originality, defined as the ratio of original content to total content for that user. Users with an originality score less than one, meaning that they copied content more often than they posted original content, were classified as content aggregators. Users with an originality score greater than one, indicating that they posted new content at least as often as they copied existing content, were placed in the content producers group.

After dividing the users into these groups, we again calculated descriptive statistics for each group (Table 9). We next fit the base regression model to the data. The results show some important differences between the ways that the two groups benefited from content curation actions. Contrary to our hypothesized relationship, we did not observe a significant benefit from traditional media for content producers (Table 8, model 3), Thus, hypothesis H4a was not supported. In fact, the only variable for which we observed a significant benefit for content producers was socially-earned repins $(0.006; p < 0.001)$.

Variable	Base(1)	Lag(2)	Prod(3)	Agg(4)
Original Content	$0.004^{***}(.001)$	$0.007(.003)$	$0.009(.02)$	$0.004^{*}(.002)$
Copied Content	$0.0007(.002)$	$0.007(.004)$	$0.009(.006)$	$0.0006(.001)$
Following	$0.113^{***}(.031)$	$0.072^{***}(.02)$	$0.04(.02)$	$0.145^{***}(.038)$
LERC1	$0.003^{***}(.0003)$	$0.002^{***}(.0005)$	$0.006^{***}(.001)$	$0.003^{***}(.0003)$
LERC2		$0.002^{***}(.0006)$		
LERC3		$0.003^{***}(.0003)$		
LECC1	$-0.0003(.001)$	$0.0001(.001)$	$-0.002(.002)$	$0.0001(.0012)$
LECC2		$-0.0007(.001)$		
LECC3		$-0.0012(.001)$		
N	12890	9149	1832	10939
R^2 (within)	0.37	0.30	0.46	0.37
R^2 (Overall)	0.33	0.38	0.20	0.38
$* * *p < .001; * * p < .01; *p < .05$ Notes: LERC = The lag of earned-repin change; LECC = The lag of earned-comment change				

Table 8. Panel data regression results

Next, the model 2 was fit to the group of content aggregators. Contrary to hypothesis H4b original content was actually more strongly associated with followers for the aggregator group than copied content (coefficients 0.004 for original content, vs. 0.6 for repins). The change in socially-earned repins was significant in the aggregator model $(0.003; p < 0.001)$, while socially-earned comments was not. Finally, user engagement had the highest observed coefficient of any variable in the aggregator model $(0.144; p < 0.001)$.

Finally, we then considered differences between the role of user engagement for the two groups. Hypothesis H6 was supported. User engagement was the largest coefficient in the content aggregator model, while not playing a significant role in the producer model. Table 4 shows the results of our hypothesis testing.

	Producers			Aggregators		
Variable	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**
Customer Acquisition						
Followers	382.2	879.8	2196	263.6	553.3	12552
Traditional Media						
Original Content	215.6	296.8	2196	83.5	168.0	12552
Copied Content	87.6	171.8	2196	520.5	710.7	12552
Socially-Earned Media						
Earned Repins	934.8	2160.0	2196	1034.2	2877.6	12552
Earned Comments	49.0	174.4	2196	21.4	116.8	12552
User Engagement						
Following	197.9	1099.3	2196	180.9	582.7	12552

Table 9. Summary statistics for content producers and aggregators.

Hypothesis	Supported?
H1a: More traditional media is associated with more CBN followers.	PARTIAL
H1b: Original content will be more strongly associated with CBN followers than copied content.	YES
H2a: Greater amounts of socially-earned media is associated with more CBN followers.	PARTIAL
H2b: Socially-earned media will attract more followers than traditional media.	PARTIAL
H2c: Socially-earned media will exhibit an increasing exponential relationship with time.	NO
H3: User engagement will be positively associated with CBN followers.	YES
H4a: Follower numbers for content producers will be influenced more by original content, rather than copied content.	NO
H4b: Follower numbers for content aggregators will be influenced more by copied content, rather than original content.	NO
H5: User engagement will be more strongly associated with the followers of content aggregators than content producers.	YES

Table 10. Results of hypothesis testing

4.6 Discussion

This section discusses some of the most important findings from the study. The findings can be placed in three main categories.

(1) Findings related to the general impact of traditional and socially earned media, as well as engagement actions.

(2) The impact of socially earned media *over time*.

(3) Observed differences in the role of traditional media, socially earned media, and engagement actions attributable to group differences between content producers and content aggregators.

Based on the results from our base model analysis (model 1), we can conclude that traditional media, socially-earned media and engagement actions all play a part in determining the CBN followers of users in our sample. From this we can conclude that more is better in terms of curation-based network marketing. Consistent with past research, and knowledge and experience from marketing practice, we observed that a larger amount of effort was associated with follower gains.

Taken on its own, this finding is not unexpected, and in and of itself it is of limited practical value. Content curation actions within the CBN are costly, just as content is expensive to produce and collect. User engagement actions also require an investment in terms of time and effort. For these reasons, it is necessary to consider the strategic preference of certain actions over others. In our analysis, the highest impact on follower growth was observed for user engagement actions. This finding is interesting because the curation-based network environment makes customer engagement possible on a level not typically afforded to organizations. User engagement actions allow for content sharing reciprocity, and let users monitor the actions of current and potential followers within the network. As Gallaugher and Ransbotham (2010) argue, monitoring customer interaction is an important part of social media marketing strategy. The authors recommend taking maximum advantage of the social network system's monitoring features to track user dialogue both between the firm and its customers, as well as customer dialogue with other competing firms, and finally customer-to-customer interaction. Entering into following relationships within the curation-based network make this monitoring process much easier. Following relationships generate a stream of activity updates that keeps a user up-to-date on the actions and activities of all of those they are following.

Another reason for the strong showing of user engagement may have to do with the quality of repin content available to users that engage with the CBN community. As users commit to following relationships, they gain the ability to see as another user adds content. Moreover, they can see in real-time, which content is gaining popularity in the network. This allows users to pick popular content to repin. Our group analysis provides some support for this idea. The observed coefficient for content aggregators was more than four times that of content producers, indicating that customer engagement played a much more important part in determining follower growth for content aggregators. The lower coefficient for content producers may be due to an over-reliance on traditional media actions on the part of content producers. We interpret this as a content search benefit granted to content aggregators, who rely more on the content of other network users. Content producers, themselves less reliant on existing CBN content, would seem to not benefit as directly from user engagement.

We see this as a significant implication of this study. It is reasonable that content producers may have good reasons for only using their own content in their collections.

However, it is important that producers not view the CBN as a one-way communication channel. Even if these producers do not repin content, they are still able to observe which types of content elicit the greatest community response. This will allow them to tailor their content presentations to the needs and wants of the community.

In addition to the strong showing of user engagement, we found that traditional media played an important part in determining CBN followers. Specifically, we observed an impact of original digital content. In our base model we were able to statistically determine that original content played a greater role in determining CBN followers than copied content. When the sample was segmented into content producers and content aggregators, however, we observed that content producers receive no discernable benefit from their traditional media marketing efforts. Content aggregators, on the other hand, saw a large benefit from the use of traditional media, specifically original content. This finding shows that within the CBN context, new content is generally preferable to repins, and that the community rewards users who take the time and effort to scour the web for fresh and exciting content. However, it seems that a mix of content may actually be the best strategy. It is possible that both original and repin content are subject to diminishing returns, and that an ideal traditional media strategy would not invest over-much in either of these areas.

The next findings that we discuss deal specifically with socially-earned media. The study provides some surprising evidence that the format of socially-earned media may be extremely important in a CBN context. In our base model, we found that textual socially-earned media in the form of user comments had no significant impact on a user's followers. This finding is very surprising given the established effectiveness of textual comments as electronic word-of-mouth (eWOM) in other research settings (Chevalier and Mayzlin, 2003; Godes and Mayzlin, 2004; Mudambi and Schuff, 2010). Consistently, this research finds that textual eWOM has an impact on a number of firm and consumer goals. Our research extends and contributes to this by revealing that textual eWOM may be sensitive to context. We believe that the CBN context, because it allows for such robust digital media sharing, naturally suppresses the role of textual communication to some degree. As CBN technologies continue to develop, it may be that we will witness a shift away from textual media to more content-based forms of communication.

One implication of this finding for practitioners is that electronic word-of-mouth communication via digital content may be less open to interpretation. Textual discourse, because of its very nature, is open to interpretation, and numerous meanings may be given to text because it is lacking in many of the audio and visual cues present in face-to-face communication (Anderson and Pérez-Carballo, 2001). As a result, many online opinion forums are prone to obscuring a marketing message at best, or going completely off-message at worst (Dellarocas, 2003). The comments section of YouTube, for example, is notorious for rapid degeneration (Lange, 2007). Repins have an advantage over text in that each repin represents an exact copy of the original traditional media. This gives marketing practicioners more control over exactly what information is spread to other customers. For a company, the ability to place specific marketing information, in the form of pictures of products for example, within the curation-based network, allows the company to control exactly what other customers see and share. Whenever a piece of this content is repinned, only the picture itself is copied. The textual comments do not copy over. This allows the marketing message to be reset every time that a user repins a piece of content.

The extensions to our base model add some complexity to these interpretations however. Another major finding of this study concerns impact of socially-earned repins and comments over time. This was done through the use of lag variables corresponding to a total of three weeks' time (8 model 2). From this model we report two major findings. First, even when accounting for the cumulative time effect of socially-earned media, we observed no significant impact of socially-earned comments. Meanwhile, socially-earned repins exert an influence over CBN followers that is powerful and long-lasting. This finding implies that companies engaging in curation-based marketing must take a long view when evaluating the performance of marketing campaigns. In our base model, we were not able to statistically determine whether original content attracted more followers than socially-earned repins. However, when adding in the compounding effect of repins over time, socially-earned repins were shown to be more important for attracting followers. This is consistent with past research that has shown that socially-earned media campaigns take time to develop (Misner, 1994; Trusov et al., 2009). Corporations will need to understand and incorporate these time effects to adequately develop and evaluate effective social-media marketing campaigns.

Finally, we investigated socially-earned media separately for content producers and content aggregators. This analysis revealed some surprising results. Content aggregators saw a significant benefit from three different sources; socially-earned repins, original content, and user engagement via following relationships. Content producers, meanwhile, only benefited from the socially-earned repins that they were able to solicit from the CBN community.

This finding sheds some important light on the competition between content producers and content aggregators. As previously stated, content aggregators are currently commanding complete control of the online advertising market, and realizing huge revenues while content producers struggle to stay profitable. We believe that a big part of the reason for content producer difficulties may stem from an over-reliance on efforts to display and promote their own content. Such efforts are naturally at odds with the two-way communicative nature of the internet. Aggregators, who are not locked into using a bounded set of content for advertising purposes, are free to pull from the nearly limitless supply of interesting content found all over the web. As a result, they can benefit not only from their own efforts, but by associating themselves with popular and influential members of online communities. Content producers have much to learn by studying the actions of effective content aggregators. The results of this current study should be very interesting to content producers looking to achieve this goal.

In this chapter, we have put forth an investigation into the impact of marketing activities in curation-based online social networks. Using a unique data set collected over a period of 8 weeks and involving a mixed sample of companies and individual users, we developed and empirically tested a series of econometric models to investigate the impact of marketing media and user actions within a curation-based network environment. Past research has shown that gaining attention and attracting customers relies on an appropriate mix of marketing activities. We provide support for this research by showing that traditional media, socially earned media, and customer engagement all play a complex role in process of generating online "buzz". Our findings suggest that in this environment, textual discourse plays a diminished role, and that the primary means of buzz generation involves the sharing of digital content. In the next chapter, we extend the analysis conducted here by incorporating network data on the interconnections between CBN followers.

CHAPTER V

STUDY 3: CURATION-BASED NETWORK DIFFUSION

This chapter presents the third of our studies around curation-based online social networks and the process of digital content curation. In this chapter we study the diffusion of digital content through CBN network structures. In doing so, we extend our work in the previous chapter by examining not only digital content curation actions, but also the underlying network structures corresponding to a sample of CBN users. The chapter has two driving research questions. 1) To provide a framework for future research that will develop and test a method for predicting eWOM diffusion based on interest similarity and CBN structure, and 2) To show, in detail, how CBN data can be used to explicate the eWOM actions of successful users as well as successful network structures. Our investigation is motivated by the fact that businesses today are working to establish a presence in existing online social networks that serve as a starting point for marketing efforts and customer interaction (Evans and McKee, 2010). Numerous online communities provide examples of this phenomenon, from Facebook pages, Twitter feeds, and pages on curation-based sites like Pinterest. In fact, a recent survey found that as many as 80% of businesses are now maintaining some presence within these types of sites [1].

These moves on the part of companies make good strategic sense for several reasons. First, online social network sites increasingly represent the portion of the internet in which many web users spend the majority of their online time (Parameswaran and Whinston, 2007). Originally, online social network sites served a primarily social service, giving

[1] http://www.vendingmarketwatch.com/news/10732362/

people a place to meet friends and discuss the issues of the day (Ellison et al., 2007b). Over time, as new services, applications and functionality were introduced into these networks, there is less need for social network users to go anywhere else. Today, for example, it is possible to read news, follow current events, check weather sports and more, all from within the confines of a single online social network system.

The fact that customers are spending so much time in online social networks puts pressure on companies to develop marketing strategies tailored to these digital spaces (Trusov et al., 2009). Such strategies naturally need to take advantage of the unique bi-directional communication abilities of online social networks, and the ability of online social networks to span the distance between companies and their customers (Gallaugher and Ransbotham, 2010). However, despite organizational demand, few concrete strategies have been developed for practice, and research continues to struggle to explain effective marketing practices for these settings (Misner, 1994; Stephen and Galak, 2012).

The situation has led some companies to question the value of online social network marketing campaigns. We previously discussed the example involving GM, who pulled out of the social media marketing business, saying that the campaign was not profitable and the online discussion it elicited was too hard to predict and control [2].

This study aims to improve firms' understanding of the spread of rich media eWOM in curation-based online social networks. So far, we have devoted significant attention to a discussion of eWOM, its benefits and challenges. eWOM has been shown at times to be more effective than traditional marketing efforts (Thorson and Rodgers, 2006). Specifically, positive eWOM can increase customer opinions of products and brands (Mudambi and Schuff, 2010), customer loyalty and product sales (Chevalier and Mayzlin,

[2] `http://articles.latimes.com/2012/may/15/business/`

2003). However, it is well known in the marketing literature that capitalizing on eWOM communication can be difficult for a number of reasons (Trusov et al., 2009). Curation-based online social networks provide us with the ability to study eWOM diffusion in new and interesting ways. There are several important aspects of CBN that make it worthy of study. The first is the interest-based nature of link formation within the CBN.

For information diffusion to occur within an online social network, close network nodes must share some similarity around the information that is to be diffused (Lewis et al., 2012). Much of the excitement around eWOM is based on the assumption that "friend-based" online social network neighbors share similar interests. It is often further assumed that people who we are close to socially hold considerable influence over our tastes and preferences (Susarla et al., 2012a). However, there is increasing evidence that this assumption may not be wholly accurate.

Recent academic research has argued that the networks of many online social network users may lack the essential similarity between nodes that is necessary for eWOM diffusion. In one paper, Agarwal et al. (2009) examined whether it was possible to predict user interests within Facebook by looking at the users in the network around them. What they found was that social peers are not particularly useful for predicting interests, in that they do not necessarily share interests in any predictable way. Lewis et al. (2012) examined the issue from a similar standpoint. The authors were interested in the nature of influence, and whether social connections are particularly influential when it comes to tastes and preferences. Here again, they found that just being friends with someone carries little weight in terms of influencing what someone will buy. Because CBN are not networks of friends, peers, or coworkers, but rather networks of shared interests, we believe that the context of CBN is a more natural fit for the diffusion of eWOM information. Thus, the very nature of CBN may eliminate some of the problems identified by Lewis et al. (2012) and other authors.

CBN also help to reveal the causal mechanisms that govern network growth and the formation of online social network structures. Since CBN come about as a result of a collection of digital content curation activities, we are able to identify the actions that lead to network formation. This is in sharp contrast to the majority of online social network research, where the largely invisible way in which eWOM networks grow and diffuse marketing information poses significant constraints on our ability to make causal inferences (Garg et al., 2011). In many studies of eWOM, the underlying network structure is both invisible and exogenous (Jackson, 2005). As a result, studies that focus on network position and structural characteristics are rare. We discussed in chapter 2 the difficulties associated with collecting this type of data, and studies that examine diffusion without structural network data are not able to identify the social interaction that lies at the heart of information diffusion (Jackson, 2006; Mayer, 2009).

Without knowledge of the underlying network it is almost impossible to identify causal effects (Mayer, 2009). This means that, while we may be able to observe the marketing strategies produce eWOM, it is not possible to tie these actions to firm outcomes in any reliable way. Recent IS literature has attempted to address some of these problems. One important recent study examines content diffusion within YouTube (Susarla et al., 2012b). The study is significant in that it finds that key structural characteristics of a YouTube user's network are correlated with the popularity of the user's videos. Our work here is similar in some respects, with a key difference. While Susarla et al. (2012b) identifies the important role that interests and preferences pay in eWOM diffusion, their data collection does not allow them to endogenate interests, but only to control for interest-based differences.

We extend this work into the curation-based context, where interest-similarity is inherent in the network context, and thereby allows us to develop measures of interest similarity between network users, and study the way that interest similarity directly impacts eWOM diffusion. Additionally, the nature of the CBN environment allows us to observe, and thus model, network growth as a function of digital content curation actions and the community response to those actions. This allows us to endogenate the underlying CBN structure, and as a result make causal inferences that are not possible in the bulk of existing IS literature (Mayer, 2009). For this purpose we make an important extension to our existing data analysis from chapter three. In our previous work the *number* of network followers was used to operationalize the concept of a CBN network. In this study we examine data on the network of followers and their interconnections. Through the addition of network-level data, this chapter extends the analysis from chapter three in order to study the diffusion of digital content through a sample of actual real-world CBNs.

In the next sections, we discuss the methods used to identify and quantify user behavior and network structures within CBN. As before, we stay within the confines of the Pinterest CBN system. We incorporate network-level data about users in the Pinterest CBN, and study the networks of 30 different users, all of them companies marketing products through CBN. These companies are analyzed across a number of different dimensions related to both their social networks and their digital content curation activities. The data collected around these companies allows us to present an example of the type of analysis possible using data mined from CBNs to show how eWOM actions in a CBN environment can be uncovered and visualized to aid our understanding of the process of eWOM diffusion. We provide much greater detail here than in other dissertation chapters on the collection of data and the nature of the analysis used both here and in Chapter 4. Next, the necessary extensions to collect network level data are discussed. Finally, we discuss some preliminary analysis of this data that makes use of the UCINET (Borgatti et al., 2002) and the NetDraw package for network visualization. The chapter concludes with some discussion of limitations and a plan for future research and extensions.

5.1 Theory Base for Research - Strategic Network Formation

The study of social networks did not begin in the field of Information Systems. Sociologists and mathematicians have both sought to understand the nature of network formation for many years (Milgram, 1967). One of the first seminal studies of social networks in sociology was Milgram (1967). This study used an ingenious method of letter mailings to try to explicate the underlying social connections that people use when diffusing information. As the study of social networks has become more complex, interested researchers have naturally gravitated towards online settings, where information systems allows for data collection on social network activities not possible offline (Mayer, 2009). Social interaction outside the boundaries of an information system is complex and unstructured. People socialize in an endless variety of ways and their interactions take place within any number of groups and settings. This makes it extremely difficult to model network formation offline in a rigorous and meaningful way.

The situation of online social networking offers some help in this regard. Online social networks track user behavior and the interactions between network members. As a result, much of the complexity involved in network formation is reduced, so that the problem of studying online social networks is reduced to one of data collection.

Recent years have witnessed some significant progress in the development of tools for data collection from online social network systems. The Scrapy webcrawling framework, for example, which we made use of in Chapter 4, is also a valuable tool for collecting social network data. As tools for collecting online social network data have improved, they have brought about an increase in researcher interest around online social network phenomenon (Susarla et al., 2012a). In particular, studies into the way that social networks form have become very popular recently in areas including sociology, economics, and computer science. This extant research on social network formation forms the theoretical basis for our work in this third dissertation study.

Typically, studies consider the process of network formation as either *random* or *strategic*. There are important differences between these two concepts, and each conceptualization offers advantages in particular settings. Studies of random network formation have their origins in the graph theory literature originally developed for the realm of mathematics. These studies model network formation based on rules determined by assumptions built into algorithms. Random network formation is beneficial when attempting to predict how networks will form based on certain criteria or rules that can be adequately approximated by algorithmic means (Erdős and Rényi, 1961). Random network formation techniques have been usefully applied to a number of areas including biology, genomics and chemistry. However, for studies of networks involving human actors, studies of strategic network formation are probably more common. This is the area of literature that we will make use of in our current work.

The concept of strategic network formation comes from the field of economics, where the tools of game theory are applied to social network data in order to model network formation as a set of strategic decisions on the part of rational actors (Mayer, 2009). Actors are assumed to make these decisions based on a set of resource constraints. Further, actors have specific utility functions and choices under which making network connections is favorable. If the costs of the connection outweigh the benefits, then no connection is formed. The analysis is based on two key assumptions. First, networks are thought to provide some benefit to the agents, thus it is possible to rank network structures as “good”, “better”, or “bad” in terms of how well they are suited to a particular objective (Jackson, 2005). Additionally, the agents themselves make decisions about network connections, and thus networks can be predicted based on equilibrium concepts from traditional game theory (Jackson, 2006). This second assumption is particularly well-suited to online social networks, where the system always allows users to add or remove connections at will.

Within models of strategic network formation, anytime a person makes a connection within a typical online social network, it involves some cost. This may be in the form of reduced privacy, greater effort involved in managing the network, etc. As a result, the agent works to create a network that is as large as possible, without becoming so large that privacy concerns and information sharing costs outweigh the benefits of adding network nodes. Users of online social networks continually make these types of trade-offs. A continuing question in the IS literature concerns the seeming lack of privacy concerns on the part of online social network users. Why do people continue to share so much personal information when this information continues to be compromised or used for unintended purposes? The reason is that online social network users derive benefits from the creation of online social network links, and that these benefits outweigh the perceived privacy dangers of link formation (Thambusamy et al., 2010). Creating links generates exposure for the user within the network. This is important, because a greater number of connected nodes imparts certain advantages within the network. For example, better connected individuals may have an easier time diffusing marketing information or receiving the best updates about new products or services. Within a CBN context, it is reasonable to assume that the amount of information diffusion would increase for larger networks, as these networks offer more nodes through which information can pass.

In order to study the strategic formation of CBN network structures, there are two major data collection challenges that must be overcome. First, it is necessary to collect and analyze data on the strategic actions of the actors within the network. In Chapter 4, we presented a study of just such a series of actions. CBN offer users the ability to create and consume various types of traditional and socially-earned media, as well as engage with other users in order to acquire network followers. These actions all serve to attract CBN followers, who network with one another and give rise to a CBN structure. Through the study of the digital content curation actions of the users in a CBN, it becomes possible to actually endogenate the CBN structure and therefore model digital content curation as a process of strategic network formation. The second major challenge involves collecting data on the CBN structure itself. This means identifying the relevant actors and their interconnections within the CBN. In the next sections, we discuss the modifications made to the Scrapy webcrawling framework that allows it to collect information on CBN structures.

5.2 Explicating CBN Actions and Structures

The analysis presented here is meant to be descriptive and somewhat exploratory. To this end, we present extensive details on the way that CBN data for the study was compiled and analyzed. Additionally, we dicuss the aspects of this data that make it unique and interesting to both researchers in the area of eWOM and CBN practitioners. To complement the descriptive nature of the data presentation, we pay considerable attention to several ways in which the analysis presented here could be easily extended, and outline the contributions of each of these extensions.

This study is meant to serve as a descriptive treatment of the methods and analysis of strategic network formation in the CBN. To this end, we provide a detailed example of exactly how the data for this study was collected around both 1) the strategic content curation actions of the CBN users in our sample, and 2) The structural data on the relationships between users in our sample. We begin by discussing the sampling methods that were used to select study participants, before discussing the strategic actions observed for these participants. Finally, we discuss the extensions to the Scrapy webcrawling framework that allow for the collection of structural CBN data.

5.2.1 Sampling

Recall that in Chapter 4 we were interested in the way that CBN actions led to a larger CBN network, as measured by a user's number of CBN followers.At the time, we were not interested in making assumptions about the underlying network structures corresponding to these users. For this reason, this analysis studied a mixed sample of users and companies randomly selected from within the Pinterest CBN. Here, we adapt our sample criteria and examine a study that consists exclusively of companies and corporations. This was done because the data collection requirements for the collection of network structure data are quite intense. These requirements placed strict limitations of the number of entities we were able to consider. For this reason, we chose to limit our sample to companies so as to have a large sample of companies and increase the relevance and findings of the study for an organizational audience.

We selected a sample of 30 different CBN users. These users were not selected at random, but rather because they met two specific criteria. First, they are all organizations with products or services that they wish to market through the CBN environment. For this reason, we would expect that the CBN pages of these organizations would be populated with mostly original content. This was an important criteria for this study, since we are interested in seeing how original content diffuses through the CBN. Repinned content can be difficult to trace back to its original source. Through the selection of companies that rely on original content, we are able to track content as it moves through the network by starting at the first point at which it enters the network. This makes the study of content diffusion not only easier, but more credible. The second criteria for selection was that the organizations and their CBN networks are small enough as to allow for data collection in a reasonable time, and to allow us to control for any elements of fame or reputation that might exist outside of the CBN environment. In other words, the companies in our sample are not "household names".

After selecting the companies for the sample, we first used the techniques developed in Chapter four to examine the traditional and socially-earned media activities of the companies. Companies were evaluated to determine their followers, their traditional media actions, and the community response to these traditional media activities (i.e. their accumulated socially-earned media).

Table 11 provides a summary description for each variable in our analysis. Table 12 shows descriptive statistics of the media variables for the companies in the sample. This first step was primarily done to ensure that the companies selected for the sample behaved in ways that were consistent with our selection criteria. As stated above, we did not want to include a company that was already above a certain number of followers, or earning excessively large or small amounts of socially-earned media. This first data collection step allowed us to control for these situations.

Variables of Interest	**Description**
Followers	Size of a particular curation-based user's network.
Traditional Media	
Original Content	New content from outside Pinterest brought into the network.
Copied Content	Content Copied from other Pinterest user pages.
Socially Earned Media	
Earned Repins	Number of Times content is repinned by other users.

Table 11. Description of variables for study 3.

As can be seen from the table, the companies had an average of 438 followers each. This is well within what might be expected from a small business without large name recognition. Thus, through the selection of this group of users, we believe we have effectively controlled for the effects of outside sources of CBN fame. Notice also that all of the companies made heavy use of digital content, with most companies posting original content as opposed to repinned content more than 50% of the time. Based on this, the companies would all be in the category of *content producers*, to use the term from Chapter 4.

This was also intentional, as we wanted to study a sample of companies that would be closely aligned with the characteristics of the typical business found in CBN. Most businesses fall into the category of content producers. These businesses use the CBN as a place to display original content around their products and services, rather than advertise the products and services of other companies through repinned content (Carr, 2012).

Variable	Mean	Std. Dev.
Followers	331	307.201
Traditional Media		
pins	718.793	751.274
original	455.931	347.92
Socially Earned Media		
earned repins	243.172	237.345

Table 12. Summary statistics for traditional and socially-earned media for companies in study 3.

5.3 Data Collection: Extending the Scrapy Web-crawling Framework

The statistics in Table 12 were created using the same webcrawling applications developed for our analysis in Chapter 4. however, by extending the Scrapy framework, we can achieve a more sophisticated level of webcrawling that allows us to analyze the structural characteristics corresponding to the networks of users in our sample. This involves a two step process. First, we use Scrapy to collect the usernames of all the followers of those companies in our sample. This is done by writing a webcrawler that can look at each page of followers for one of the companies in our sample, and compile a list of the URLs corresponding to these followers for subsequent crawling. We then go to a further level of depth, and using this list we aalso collect the followers each of those users. As a result, the number of users that Scrapy is forced to crawl grows exponentially with each level of depth, so that even small networks take a large amount of time to analyze. It should be noted that we also took the potential load of our webcrawling applications on

Pinterest's own servers into account, and implemented artificial delays into the crawling process so that the resource demands placed on Pinterest would not be much greater than that of a typical Pinterest user.

As part of this collection process, Scrapy collects two important pieces of information. First, each follower is assigned a *Company ID* which corresponds to the sample company to which this follower is associated. The *Company ID* field allows us to easily segment the webcrawling data post-processing and divide the data file up into segments corresponding to each company. This also allows Scrapy to potentially crawl a Pinterest user twice if necessary, should a follower appear in a relationship with more than one sample company. In general, a very useful feature of Scrapy is its use of an index that prevents recursive scraping. If we were crawling data from an e-Commerce website, such a feature would be invaluable, since there would be little reason to ever need to visit a page twice. Recursive scraping can expend server and bandwidth resources unnecessarily, as well as potentially cause the whole webcrawling process to fail if the spider finds itself in an endless loop. However, while this feature is beneficial is some crawl settings, it creates problems for collecting social network data. Because social networks often have multiple paths leading to the same pages, it may be just as important to know *how we arrived* at a page as what information it contains.

In addition to keeping track of the sample company associated with a particular follower, each follower is also assigned a value for *Follower_Of*, a value which tells us which Pinterest user this person is actively following. For the first level of the crawl, each follower will naturally have the same value for *Company_ID* and *Follower_Of*, though at the second level, when we crawl the followers of these followers, *Follower_Of* can correspond to any user from the first crawl level.

Using these two pieces of data, it is possible to eliminate any users who are not following the target sample company. This results in a dataset that is comprised only of those users that 1) Follow the target company, and 2) May or may not follow other users who *are also* following the target company. Figure 14 shows a representation of this process. In the Figure, we have a sample company and three unique CBN users. The spider starts by crawling followers 1 and 2 from the sample company follower list. The spider then crawls the followers belonging to follower 2. The dotted line between follower 2 and follower 3 represents an extraneous out-of-network crawl. This user only follows follower 2, and unlike follower 1, is not connected to the sample company.

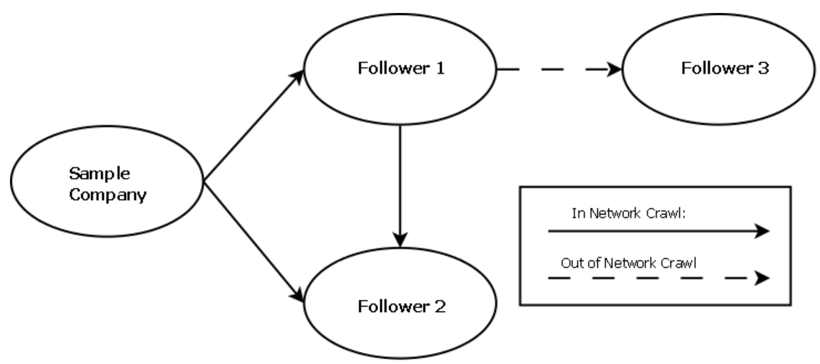

Figure 14. Conceptual model of the crawl process.

Obviously, data collection such as this is time consuming and somewhat inefficient. We are only interested in those users that are followers of the companies in our sample, or followers of a user that is a follower of our company. Yet until we look at each and every one of the followers of a particular user, we cannot know for sure whether the user belongs to this group.

To describe this process in greater detail, it may be helpful to examine some pseudo code taken from one of the Scrapy spiders originally written in Python:

```
class Spider:
        followers = Extract Array of all Pinterest
            Followers
        Company_ID = Set Company_ID to the URL of the
            current page.
        follower_urls = Get array of urls for the company
            's followers.
        *Here, we loop over each follower in the sample,
            getting their
        unique user_id and Follower_Of values.
        for follower in followers:
             item = FollowerItem()
             item['Company_ID'] = Assign Company_ID to
                 this user
             item['Follower_Of] = Get ID of person this
                 user follows.
             *this is equal to Company_ID at the first
                 crawl level.
             item['User_ID'] = Populate value for the
                 follower's own
             unique ID
             output item - print to data file.
             Turn Page.
        for follower_url in follower_urls:
             Generate URL for Scrapy
             Issue a request to crawl each of these
                 followers
             *This begins the second crawl level
```

Upon visiting a page of followers, the spider first populates an array with the HTML *div* tags that contain the relevant information on that particular follower. Another array is populated with links to the follower pages of each of the users in the followers array. The spider then populates several variables with values for the Company_ID, Follower_Of and User_ID. These values are returned as output, and the spider will turn to the next page of followers if needed. Finally, the spider uses the follower_urls array to make requests for the follower pages of additional users, and the process begins again.

The most difficult part of this data collection involved the creation of URL markers that corresponded to pages within Pinterest. Unlike most websites, Pinterest does not present data in “pages”. Rather, Pinterest shows a continuous stream of followers, and as the user scrolls down a page, the server automatically populates the screen with additional followers. It is not possible within Scrapy to simulate a process of scrolling down the page. As a result, it was necessary to create URLs that the Pinterest server would recognize in the same way as this scrolling process. The system uses a series of markers that correspond to the number of followers currently on the screen. For example, a typical URL for a page of followers might look like:

```
http://pinterest.com/username/followers?=page2\&
    marker=24\
```

The marker system is not intuitive. However it is predictable, and with careful observation it was possible to design a script that would create URLs with the correct markers. To do this required observing the actual calls to the server as a user scrolled down the page. Firebug, a community-created addon for the Firefox web browser, was employed for this purpose.

Through the use of this spider, it was possible to collect data on the complete follower networks of the thirty companies in our sample. These networks are most accurately referred to as *ego networks* (Everett and Borgatti, 2005). An ego network consists of the ego, in this case one of the companies in our sample, all the actors (or alters, to use the ego network terminology) tied to the ego and the ties that they share (Prell, 2011). Ego networks represent a major area of social network analysis that can be used whenever it is difficult or impossible to identify all the actors in a whole network (Otte and Rousseau, 2002). In the context of Pinterest, the size of the total network makes a complete crawl of the whole network impossible. For this reason ego network analysis is suitable (Prell, 2011).

Ultimately, the output from the above spider returns a network of all the followers of one of the companies in our sample. This creates a network structure similar to the left-most network diagram in Figure 15. To identify the interconnections between the followers it was necessary to do another crawl on the second level of followers, i.e. the "followers' followers". Linux stream editing tools *awk* and *sed* were then used to parse out the information of only the second level followers who were also followers of our target companies. This created the network structure seen on the right in Figure 15.

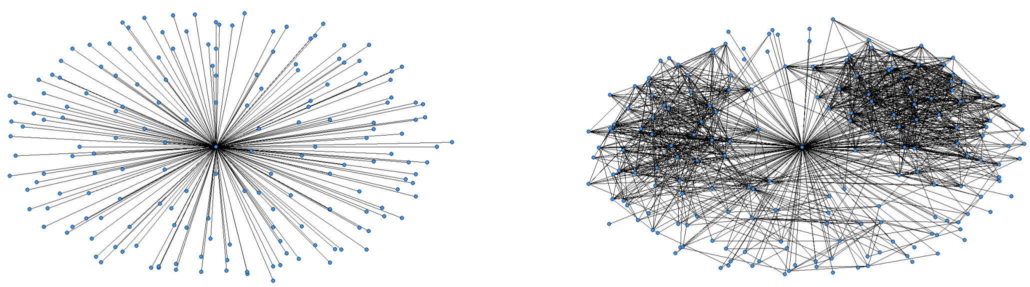

Figure 15. A "star" network structure (left) and ego network showing follower relationships.

5.4 Social Network Structural Analysis

The ego networks in Figure 15 were created using the UCINET program (Borgatti et al., 2002). UCINET provides a number of useful tools for social network analysis. In addition to network visualization tools that can provide descriptive depictions of network structures such as those seen in 15, UCINET can calculate a wide number of ego network variables for use in social network analysis.

The program takes input in the form of a text file and then converts this file into a relational matrix, which expresses the connections between users in the network. Table 13 shows an example of the data format required by UCINET to produce figures such as those seen in 15. The header information in the first lines of 13 tells UCINET the number of unique nodes in the network and that this is an edgelist, which is an array that gives the edges of a particular network graph [3]. From Table 13 it should be clear how we have made use of our variables of Company_ID and Follower_Of. Company_ID insures that each follower will have a row in the text file that links it back to the sample company, which occupies the position of ego in the center of the graph. User_ID and Follower_Of values allow us to fill in the remaining rows of our input file, rows which correspond to the other links between network alters as seen in the right-hand network of Figure 15.

DL n=119	
format = edgelist1	
labels embedded:	
data:	
beauxtithes	toshajackson
josiemae1	toshajackson
littlelizgirl	supersmartcarbo
ljhighby	supersmartcarbo
beauxtithes	josiemae1
broadwaythresh	lindadoane
broadwaythresh	rachelbaransi

Table 13. Example of file format required by UCINET

We calculated several different statistics for each of the companies in our sample. Table 14 shows summary statistics for the network size, the number of ties, and the density of each company network in the sample. Network size is simply the number of

[3] http://reference.wolfram.com/mathematica/ref/EdgeList.html

unique individuals that occupy nodes in the network. The number of ties represents the number of connections between network nodes. Finally, ego network density is a measure of the number of ties that exist, divided by the maximum number of ties that could exist. Expressed mathematically for an ego network i, it is simply $\mathrm{d}_i = \frac{T}{\frac{n(n-1)}{2}}$, where T represents the number of ties that exist in the network, and n stands for the number of alters in the network.

Variable	**Mean**	**Std. Dev.**
Size	331.556	209.562
Ties	711.778	1465.02
Density	0.687	0.651

Table 14. Summary statistics for companies in study 3.

Network size is an important variable in social network analysis, since larger networks offer the potential of a greater number of nodes for information diffusion. However, just looking at the size does not tell the whole story, and this is exactly why structural data allows for a richer type of analysis than numerical data (Jackson, 2006). Here, we are able to superimpose the concept of ego network density onto our existing notions of network size. Ego network density is measure of how closely tied are the alters in the network. If density is high, then many of the actors in the network are connected to one another (Prell, 2012). Within a dense network, the large number of connections between individuals implies that people have numerous ways of accessing the same information. Networks with high density are good for sharing knowledge (Borgatti and Cross, 2003). However, dense networks may not be as good for information discovery, because the many ties between individuals lead to a situation in which people often share the same knowledge. Since we are interested in the spread of eWOM about products and services, it is important that we are able to distinguish between networks that are good at information sharing and those that promote information discovery. 14 shows the summary statistics for the social network measures of the companies in our sample.

These summary statistics in Table 14, together with the data in Table 12 provide some idea of the relative size and performance of each company in the sample. It is also possible to look at the correlations between density and traditional and socially earned media. Table 15 shows these correlations. Starred values represent a significant correlation at the (0.05) level of significance. Interestingly, we don't see a relationship between traditional media variables and density. This is expected, given that a dense network would not likely encourage any of our companies to post more traditional media. Neither would it be reasonable to expect that the presence of traditional media would create dense network structures. The situation for socially earned media is quite different. The correlations for socially earned repins and comments with density are both significant *and negative*, implying that as networks became less dense they generated greater amounts of socially earned media.

	Size	**Density**	**Pins**	**Original**	**Repins**	**Comments**
Size	1.0000					
Density	-0.3612	1.0000				
Pins	0.0096	-0.0640	1.0000			
Original	0.1274	-0.1853	0.9176	1.0000		
Repins	0.2163	-0.4120	0.4338	0.5677	1.0000	
Comments	0.4478	-0.4774	0.4768	0.6133	0.5815	1.0000

Table 15. Cross correlation of social network variables with traditional and socially-earned media.

It is interesting to note that density seems to exhibit some influence over the creation and spread of socially-earned media. Next, we can use UCINET to attempt to visualize the diffusion socially-earned media through the ego network structures we have already discussed. In order to do this, we need to combine the data in Table 12 and Figure 15 in a way that UCINET understands. For this purpose we can use social network node attributes.

*node	data
id	repins
accucutcraft	7
ookie0o	2
errydueholm	16
klwells	22
krasmusen	2
kristenmccon	1
laur931	2
ldl1968	22
liaohara	1

Table 16. Sample of attribute file for input into UCINET

Attributes can include things such as age, gender, or some aspect of a relationship between nodes (for example, how often does actor x seek information from actor y) (Borgatti and Cross, 2003). As part of the data collection for this work, we collected the origin of every piece of digital content shown on the pages of the users in our sample. We already made use of this functionality earlier in the dissertation when we distinguished original content from content that had been repinned. For repinned content, the webcrawling spider collects a value for the origin that corresponds to the user from whom the content was repinned. Using this origin value, we can examine all of the content seen on follower pages that was repinned from one of the companies in our sample. In this way, we can obtain a number that corresponds to the number of times that a particular follower has repinned content from one of our companies of interest.

This data is entered into UCINET through the use of an attribute file. This file contains information on the user ID of each alter in the network and the number of times their content was repinned from the ego company. Table 16 shows a sample of the attribute data file.

Through the combination of this attribute data and our existing network structure, we can now visualize an ego network for one of our companies with the repin activity superimposed over the network structure. In 16 nodes have been colored and scaled to size based on repin intensity. Yellow nodes are those network members that have not repinned any content from this company. Blue nodes have repinned at least one piece of digital content, with larger nodes exhibiting more repins than smaller nodes. The area of the network where we see the most repin intensity (i.e. that portion just north of the center) is comprised mostly of short ties, or ties between closely connected alters (Watts, 1999). Several other nodes on the periphery of the network also repin large amounts of content, yet these nodes have little or no connection to the company's larger network of followers. These users also represent great opportunities for targeted marketing, since they likely have access to large groups of potential customers that are today not familiar with the company's product offerings. These are referred to in the social network analysis literature as *long ties* (Watts, 1999).

Information like this is valuable both for firms and researchers. From the figure, we can see that repin intensity is greatest in a dense area of the network right around the ego. Interestingly, we also see that two nodes in particular are both well-connected and repin a large portion of this company's content. A good marketing strategy would identify such nodes and leverage resources to attract these types of users, are they appear to be both influential in the company's follower network and interested in the digital content that the company has to offer.

With Figure 16 we conclude our descriptive example of the type of analysis possible using data from CBN sites. This example was meant to serve as an exploratory study of the ways that that social network data can be combined with eWOM actions to improve on our understanding of eWOM diffusion in online social network contexts. In the next section, we discuss some ways that this preliminary analysis can be extended to address specific gaps in the existing eWOM literature.

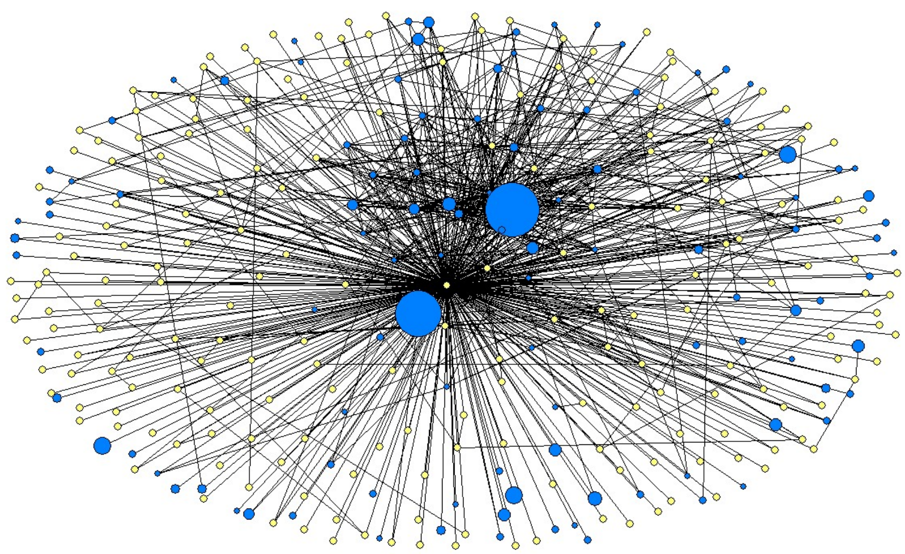

Figure 16. Ego network showing repin intensity for one user in the sample

5.5 Discussion

At this point, we have provided detail on the methods that we have used to develop a research framework for developing and testing a method for predicting eWOM diffusion based on interest similarity and network structure. Using this framework, we have assembled a data set capable of executing this goal. However, in order to fully address the research question for this study, it is necessary to extend the descriptive analysis discussed here in several ways. For example, we have shown several types of social network graphs produced through UCINET. In addition to the ability to draw social networks, UCINET has several features that allow for analysis of different node properies within a network.Using these features, we can greatly extend the types of analysis that are possible using the data collected for this dissertation chapter. In the next sections, we discuss three extensions that, once completed, will allow us to finally address the research question posed at the start of the chapter.

5.5.1 Interest Similarity

At several points in this dissertation, we have pointed out the limitations of traditional online social networks contexts in diffusing eWOM marketing information. One primary reason for this difficulty identified in the literature concerns the lack of mutually-shared interests between close network nodes (Agarwal et al., 2009). This issue of interest similarity is one that has posed a difficult for eWOM researchers for a number of years. In existing work, interest similarity is usually controlled for, without being explicitly measured or incorporated into the analysis (Garg et al., 2011).

The CBN context makes it possible for us to directly get at the question of interest similarity. Since CBN are built from digital content, it is possible to directly compare the content of one user to another, and thus obtain a direct variable of interest similarity between network nodes. This is possible using the data that we have already collected. Each piece of digital content within the Pinterest CBN has a canonical ID number that stays consistent across any number of user pages. Using this canonical ID, it is possible to develop a list that corresponds to all of the content on all of the pages of the users in the network. This list is effectively a summary of the interests of the network as a whole. We can then develop a variable that corresponds to the number of pieces of content on a user's page that are present on this list, and obtain a measure of commonality between the user interests and network-level interests. Using UCINET, this variable can then be analyzed as a network attribute, allowing us to answer questions about content diffusion within networks that exhibit varying degrees of interest similarity. In this waywe can analyze content diffusion as a relates to interest similary, thereby filling a large and important gap in existing eWOM research.

5.5.2 Repin diffusion over time

This extension to the analysis incorporates the panel nature of the data collection for this study. Recall that one of the goals of this research is to endogenate CBN network structures. To accomplish this, it is necessary that we be able to see the impact that certaincontent curation actions have on not just network growth, but the creation of network links. This allows us to make a major research contribution by identifying causal effects between digital content curation actions and certain CBN structural characteristics. Using this information, we can develop a model that can predict the type of network that will arise from certain content curation strategies. In our analysis presented so far, we have examined a network as a static entity for one time period.

The total collection for this study involved ten weeks of observation. As a result, the networks that we consider are not static, but changing entities. Actions taken in previous time periods manifest themselves in observable structural changes. The inclusion of multiple time period data will allow us to also examine the diffusion of eWOM to both *in-network* and *out-of-network* nodes. In-network refers to those nodes that are active followers of one of the users of our sample. However, it is very common that content is copied by users that are not actively following one of our users. The nature of our data collection makes these users invisible in a single time period. However, if users enter into a following relationship during our period of observation, we are then able to "see" them for all remaining data collection periods. This allows us to further endogenate network growth, by investigating the way that *out-of-network* eWOM diffusion ultimately leads to new follower acquisition.

For an example of how this analysis is conducted, consider again the company network shown in Figure 16. Figure 16 shows the diffusion of in-network repins that this company collected for a particular period. However, the attribute file used to create this figure a fraction of the total number of socially-earned repins attributable to the company . The difference between the repins in Figure 16 and the total number of repins represent out-of-network repins to users that are not company followers. In time period one, users from outside a company follower that copy content network would not be included in our data collection. Recall that the spiders used in this study are instructed to crawl only the content belonging to the followers of one of our companies. However, the spiders are intelligent enough to update follower lists each week and add in new crawl rules for followers that have joined a company follower network since the last crawl. The content from any new users would then be crawled, which would allow us to see the number of repins that were required for these new users to engage in a following relationship with our company.

5.5.3 Directed Networks and Interaction of eWOM Content

The networks that we have examined in this chapter all fall into the category of undirected networks. In an undirected network, we are not interested in questions of whether actor a follows actor b. However, some additional insights can be gained from studying social networks as directed networks. In a directed network, relationships between followers and those they are following are explicitly stated. This can be useful, especially for examining reciprocity of link formation, where a user follows someone who follows them in turn.

In many online social networks, link formation takes place in an undirected fashion. For example, if two Facebook users decide to become friends, then this is an undirected link. CBN networks are different, in that a user can follow someone without that person following them in turn. As a result, these directed networks can exhibit a hierarchy of social status, where one user may have many followers, while following very few people. One future research study could examine whether there is a benefit to having a large follower base with few followers, or whether reciprocal network relationships are more desirable. UCINET has the ability to treat any network as either directed or undirected, and our existing data preserves the relationship between followers and who they are following. For this reason, extending the analysis to the area of directed networks is easily possibly.

Finally, another important aspect of eWOM diffusion that we do not address in this study concerns the interaction between different types of eWOM content. In this chapter, we have primarily been interested in studying the diffusion of repin content through a network. In doing so, we have not addressed other variable factors that might influence the spread of repins. For example, if a piece of content receives eWOM comments, is it then more likely to be spread via repins? Past IS research has shown that textual comments can indeed lead to greater eWOM diffusion (Susarla et al., 2012a). Through the study of the interaction between textual and repin eWOM, we have the potential to validate the findings of extant IS research in the CBN context.

In this dissertation study, we have presented a descriptive analysis of eWOM diffusion in a CBN context. Our goal is not to provide an exhaustive study of all facets of eWOM diffusion, but rather to present CBN data collection and analysis in some detail, to show what is possible within this rather unique environment. We have examined eWOM actions, together with the underlying online social network structure, to show how eWOM in the form of rich media content diffuses through a CBN. Additionally, we have outlined a number of interesting and fruitful areas for future research possible with the data collection completed for this study. This provides a robust and far-reaching research plan that will make significant contributions in several areas of eWOM research.

As with all research, the dissertation suffers from some limitations that deserve mention. Perhaps the most important is that we consider only one CBN context. The selection of Pinterest as our research setting is appropriate given its status as the largest and fastest growing CBN, however the user population of Pinterest may keep it from being truly representative of society as a whole. First, while conceivably a network of any type of interest, Pinterest does exhibit a tendency to focus on arts, crafts, and *for the home* types of hobbies and interests. Other CBN that have been created during the time this dissertation was written have other niche specialities. One examples include Gentlemint, which tries to style itself as a Pinterest for men by focusing on male-oriented hobbies and interests.

It is important to point out, however, that there is nothing inherent to Pinterest that steers its user-base towards crafts or artsy interests. It just so happens that this group was the first to adopt Pinterest, and the site has been associated with these types of hobbies and interests from that point on. Not surprisingly given its specialty, the user population of Pinterest is predominantly female. Pinterest does not publish numbers on its user demographics, however some estimates have put the population of Pinterest at nearly 80% female. Pew Research reports that Pinterest is especially popular with women under the age of 50 [4]. This is not markedly different from Facebook, with a user-base that is also predominantly female. Another important limitation of the study concerns the not-altogether-random in the way that users were selected. Pinterest has a huge number of users, some of whom have massive networks of followers. We purposefully excluded these types of users from any of our studies for two reasons. First, the motivations of users with very large networks may be different from those of the typical user. Additionally, users with a large number of followers may exhibit characteristics that make it impossible to attribute their follower network to actions taken within the Pinterest CBN. In other words, they may be famous outside of Pinterest, and well-known within the larger context of online social media. In order to eliminate the potentially confounding effect of fame, we purposefully limited our sample to smaller networks. This limitation was also necessary for data collection purposes, as collecting data from hundreds of thousands of users would be both time and cost-prohibitive.

As a result, however, the study is limited in its relevance to large corporations that may have brand recognition. These companies may need to adopt different strategies in order to capitalize on their fame and status from other sources. The findings of this dissertation

[4] http://pewinternet.org/Reports/2013/

should be cautiously applied to such settings. Nevertheless, major corporations and famous social media personalities represent a small minority of total CBN users, and thus the dissertation study should be of great interest for the majority of companies, corporations, and individual users.

CHAPTER VI

CONTRIBUTIONS, IMPLICATIONS, LIMITATIONS AND FUTURE RESEARCH

In this final chapter we provide a summary of the contributions and planned extensions to the dissertation. In discussing the contributions of the study. The implications of the individual studies are discussed in turn, together with any implications that one dissertation study has for another. This final chapter also discusses several future research projects implied by the dissertation's key findings.

6.1 Study 1

6.1.1 Contributions and Implications for Theory

Study 1 makes several important contributions the extant IS research. Perhaps most importantly, we present the first study to actively study digital content curation in the context of online social networks. This phenomenon has thus far not been introduced to the IS literature, and our main goal in study 1 is to quantify, define, and study digital content curation and its implications for the larger area of online social networking. To this end, we began by studying the antecedents of user intentions to curate digital content.

Study 1 contributes a theoretical model of utilitarian content collection and social consumption of digital content. This model integrates theory from the areas of personal information management and sociology to account for the unique characteristics of the CBN platform. We are able to show that digital content curation intentions are motivated by two distinct desires. 1) The desire to discover and manage interesting digital content, and 2) the desire to use this content to create desireable impressions within CBN. One major contribution of the research is the finding that CBN organization ability is at least as important in determining curation intentions as social content consumption.

The finding that content curation intentions derive from both utilitarian content management and a desire to show off and display digital content has important implications for extending IS theory. Historically, the IS field has devoted much attention to the usefulness and usability of information technologies (Davis et al., 1989; Venkatesh et al., 2003). This is to be expected of a field that came about through a desire to understand managerial applications of technology. However, with the emergence of online social network technologies in recent years, much online social network research has thus far been populated by various social theories including social capital theory (Blau, 1964), social exchange theory (Thambusamy et al., 2010) etc. These research efforts have expanded the theoretical boundary of the IS field, and allowed for some very interesting and surprising findings.

However, our findings imply that with the rise of CBN, online social networks may be experiencing a shift away from socially observable online network usage and towards the management and presentation of content, and that the practical value of a CBN platform's content management utilities may be the most important in determining whether a design is successful. While social motivations for online social network usage will likely remain important as CBN technologies continue to diffuse, it is important that researchers adapt to the emerging CBN environment with theories that can incorporate the utilitarian side of these technologies. This finding is of particular importance to the literature around digital item acquisition (Hinz et al., 2010). Much of this research is motivated by a desire to understand why people acquire digital items, whether for prestige, utility, or some other need. Seldom does this research actively study the information management techniques that people use to locate, acquire, and keep track of their digital possessions. Our study extends this work by looking at the methods that people use for choosing and managing these items. We show that methods for discovering digital content, and importantly, for *re-finding* digital content are important features for systems for curating digital content.

The study also presents several important implications for IS research that come directly from the curation-based environments that the study explores. Digital content curation, by its very nature, has implications for such issues as privacy, copyright and digital rights management, and the very nature of internet usage. Owing primarily to the newness of CBN technologies, issues related to privacy, copyright and content management in the CBN environment have yet to be explored. In coming years, IS researchers should work to address these issues and define what constitutes acceptable use of digital content within CBN enivironments. For example, in terms of privacy, tt is generally assumed that users are more concerned about privacy when information is of a very personal nature. We found that in the CBN environment, when users felt information was less personal, or the CBN environment was more open to outside view, they were more likely to curate digital content. This finding shows that privacy is still an important part of CBN usage, and that CBN platforms may be somewhat sensitive to inclusion of personal content. More research is needed to determine if this finding is specific to our CBN context of Pinterest, or if the results will extend to CBN that are focused more on the curation of personal photos and content.

6.1.2 Contributions and Implications for Practice

The finding that utilitarian motives play a major role in CBN usage has some important implications for managers. In current iterations of CBN, content management functionality is somewhat limited. Users can create collections of content, and apply headings or tags to this content, but there is little ability to search or identify relevant themes within content collections. Rather, the emphasis thus far has been on providing users with ways to show off their content. This speaks to the social consumption side of digital content curation, while downplaying the importance of content organization. In the next iteration of CBN interface technology, it would benefit companies to add greater functionality around organizing, searching, and managing digital content collections. This finding has another important implication. Human discussion normally contains some element of context, so that comments taken without context are of little value. As a result, the typical eWOM discussions that take place within online social network sites are often context specific. One comment is placed, which spawns another comment along the same context, and so on and so on. However, for companies interested in distributing a marketing message through a CBN, these comments may actually make a marketing message less portable. For there to be any value to the textual eWOM attached to a piece of content, then the entire conversation must be diffused along with the content in order to preserve the textual context.

Rich media digital content does not suffer from this limitation. Since pictures are themselves a richer communication medium than text, the diffusion of the picture itself allows more information to spread through the network faster. As a result, eWOM by means of digital content may actually diffuse quicker and easier than textual eWOM. As communication technologies continue to get faster, we may witness an increased reliance on communication via digital content, so that textual eWOM plays less of a role in online opinion formation.

6.1.3 Limitations

Study 1 suffers from a few limitations that deserve mention. First, while we believe that a study of the Pinterest CBN environment is appropriate and timely given its status as the most popular current-generation CBN, there may be reason to question the generalizability of the study. As a private company, demographic data on the Pinterest user population is not available, however marketing data has put the Pinterest population at 80% female. Our own survey of Pinterest users resulted in a sample that was 74% female. As of this writing, Pinterest and CBN in general are too new to determine whether this gender distribution is something inherent to the curation-based format or a function of Pinterest alone.

The method that was used to recruit participants for study 1 is also a limitation of the study. Because of the newness of the CBN environment, it was necessary to recruit participants who had a certain familiarity with the system and methods of content curation. This required that we screen our sample of participants by asking probing questions about their Pinterest usage. As a result, the sample could potentially exhibit some non-response bias, though our tests for this failed to detect such problems.

6.1.4 Future Research

Several interesting opportunities for future research are implied by the findings of study 1. First, several of the constructs used in our model are certainly deserving of future study. In particular, the construct of serendipitous information discovery has important implications for the development of online knowledge repository systems, of which CBN are a subset. Future research should work to further unpack this construct by studying its individual antecedents. While some research has attempted to do this (McCay-Peet and

Toms, 2011; Passant et al., 2008), much of this research is in an early stage. Additionally, this research is not focused on the CBN environment.

Second, the gender distribution of our sample, and the gender make-up of the Pinterest CBN, would make a gender study of content curation very profitable. In our theoretical discussion, we made a case for the female role of photo curation in the household. It may be that there are gender differences in the motivations for digital content curation, and that these account for the observed population differences in current CBNs. Such a study would be interesting for marketing and IS researchers, and would no doubt benefit practitioners as they develop current and future generation curation-based systems.

Another interesting area for future research would look at user differences in curation intention antecedents. In our study, we consider individual-level antecedents of digital content curation. Another question of interest would look at company intentions. CBNs play host to large numbers of varied companies and corporations, each with different motivations and goals for the CBN environment. It would be interesting to see the way that these different motivations impact the nature of digital content curation for these groups.

Finally, it is interesting that our analysis did not find support for the role of *reminding* in digital content curation intentions. This is interesting given that intuitively we would think that users would appreciate a system that was good at reminding them of useful and interesting content. We believe that the reason this construct appeared as negatively related to content curation intentions may stem from its subconscious nature. It is possible that, because reminding happens independent of concious user action, it is difficult for users to self-report a system's ability to remind. If this is the case, then a behavioral survey analysis is likely not the best way to evaluate this construct. Future research could examine the construct in an experimental setting in order to either refute or provide further validation for our finding.

6.2 Study 2

6.2.1 Contributions and Implications for Theory

The second dissertation study makes numerous contributions with relevance for theory and practice. First, the study is the first to consider the impact of digital content curation actions on growing a network of followers. As of this writing, the study has relevance for any company or individual looking to attract followers within online social media. However, as CBN continue to grow in popularity, the study should obtain special merit based on its specific treatment of the curation-based environment.

One aspect of the study that we predict will give it lasting relevance is its examination of a new type of eWOM in the form of socially-earned repins. We find that these repins exhibit a strong and lasting impact on follower acquisition in the CBN environment. This is important for research around eWOM in this environment. In contrast to much extant eWOM research, we found no support for the role of textual comments on follower acquisition. Instead, our investigation found consistent support for socially-earned repins, while uncovering no statistical evidence for the role of textual eWOM.

The examination of socially-earned repins also makes an important contribution to the small but growing area of research around socially-earned media. The majority of studies in this area consider at most one company or one specific industry. Study 2 is the first study to consider a large, random sample of companies and users. This gives the study generalizability and a scope that sets it apart from the extant research. Additionally, by considering a mixed sample of companies and users, the study contributes to extant theory on the nature of customer acquisition in online social media. We are able to see effective strategies of both users and companies. This allows us to see effective strategies for each group, but it also lets us identify effective mixed strategies that incorporate the best techniques taken from all types of users.

Another important contribution concerns study 2's treatment of original content and copied content. Past research in social media marketing has struggled to understand why some social media marketing messages succeed while others fail (Trusov et al., 2009). Study 2 contributes to this research by showing the impact of novelty, and endogenating the concept of novelty into a model of customer acquisition. We are able to conclude that novelty and original content plays an important part in customer acquisition. Additionally, we show copied content is not only less effective than original content, the study provides evidence that it may not even be effective at all.

6.2.2 Contributions and Implications for Practice

A primary contribution of study 2 is its treatment of socially-earned repins as a new form of eWOM. We have discussed the importance of this contribution for theory development, but it also has important implications for practice as well. Online social network technologies have created a time in which most companies are still struggling to understand and capitalize on changing virtual environments. For example, despite the intense amount of media attention leading up to Facebook's huge IPO, uncertainty about the company's ability to capitalize on the untapped potential of eWOM advertising [1] contributed to the stock performing poorly soon after going public. Moreover, declining revenue from traditional ad-based online advertising (Berman et al., 2011; Chatterjee et al., 2003), combined with the fact that people are now spending more time on online

[1] http://www.technologyreview.com/news/427972/

social networks (Golder et al., 2007), has made the demand for eWOM marketing strategies higher than ever. Despite this, there is still relatively little that is known about how to encourage people to talk about the things they like, and how to create systems that foster useful eWOM communication.

As evidence of the uncertainty around effective eWOM, consider that we found no support for socially-earned comments, despite the fact that research and practice has devoted large amounts of attention to studying the diffusion of textual comments online (Mudambi and Schuff, 2010). The implication of this finding is that textual eWOM, which has received such a storied treatment in the extant IS literature (Chevalier and Mayzlin, 2003; Godes and Mayzlin, 2004), may be less effective for attracting attention within CBN than socially-earned repins. This has implications for 1) the CBN users that are interested in gaining a network of CBN followers, and 2) the companies that actually produce the digital content. If textual eWOM truly has so little impact of content diffusion, then marketing efforts within the CBN need to focus on enticing users to repin this content onto their own pages. Additionally, CBN interface designers should continue to emphasize digital content sharing in their designs, while potentially downplaying the ability to add textual comments.

In addition to the study's finding that one type of socially-earned media outperforms another, we also find variance in the role of traditional media in follower acquisition. In our models, original content greatly outperformed copied content in terms of attracting followers. This makes intuitive sense, and supports decades of past research in the area of advertising and marketing (Kaye and Johnson, 2003). The finding has important implications for practitioners investing in social media marketing. Based on our analysis, companies and users alike benefit the use of new and exciting content. Such content is likely costly to find and produce, however, showing that social media marketing may be a more complicated and costly endeavor than it would first appear.

Finally, the study is the first to consider a sample of competing content producers and aggregators. Extant research offers theoretical or analytical treatment of the conflicting roles of production and aggregation in virtual spaces (Dellarocas et al., 2013). Our study is the first empirical study to offer an analysis of real-world data on the strategies that each group uses when advertising and competing for market attention. We have frequently discussed the role of content aggregation in online marketing, and some work has focused on how content aggregation impacts the growth and development of the world wide web. In today's *link economy* (Jarvis, 2008), which refers to the web of hyperlink connections that bind together the internet, hyperlinks have distinct business value. With the drastic decline in traditional advertising revenues, content producers are desperate to capture revenue from online advertisements. However, this is a battle that they are losing, as giant aggregators like Google are built as gateways to point users to content producers' sites, collecting revenue at each step along the way (Osnos, 2009).

Today there is an ongoing debate around property rights and revenue generation in these types of interactions. Content producers populate the internet with huge amounts of interesting and original content, and without them the internet would be a pretty boring place. However, advertising revenues often go to websites that are able to generate the most traffic, and this is an area where curation sites win easily. For example, the blog boingboing.com generates more traffic than the New York Times by curating content from the New York Times and other similar news sites[2].

[2] http://www.impactbnd.com/content-curation-inbound-marketing-imperative/

The situation is similar to the industry-shaping impact felt by Napster back in 1999. Before Napster, all the tools were available for people with access to the internet to share music and files. However, these tools were difficult to use and finding people with the files that you wanted was difficult (McCourt and Burkart, 2003). Napster did not make it possibly to share files, it made that filesharing incredibly easy. As a result, illegal filesharing changed from an annoyance to an industry-redefining phenomenon (Hong, 2013; McCourt and Burkart, 2003). Networks like Pinterest provide a set of tools that makes content aggregation easy. Many of these systems have tools that let users import content from other websites into their collections with the click of a browser button. Once the content is in the network, it can be viewed and distributed to any number of users, all without ever leaving the network system.

As a result, content producers have a vested interest in learning how to control and manipulate the CBN environment in order to compete with content aggregation services (Dellarocas, 2006). The current study should be interesting for content producers that are looking to either promote the *controlled* aggregation of their content, or build a hedge against the aggregation of their content into other sites. Methods to simply stop users from copying content from one place to the next are, just based on past evidence of the prevalence of content agregation, not likely to succeed. Rather, content producers can use the findings of this study to build content curation systems into their own websites, so that users are provided with an incentive to stay on the content producer's site. Our analysis sheds some light on why aggregators tend to outperform producers in most advertising environments. We show that one reason that content producers may struggle when competing against aggregators stems from the one-sided approach to advertising that many producers adopt. Whereas content producers tend to benefit from only one source (socially-earned repins of their posted content), aggregators successfully leverage content, socially-earned repins, and user partnerships. The findings of this study point to the need for content producers to adopt a much more stratified marketing strategy in online social media environments in general, and CBN in particular.

6.2.3 Limitations

The findings in this study are limited in three areas that deserve mention. First, the study suffers from the same potential population issues as study 1. However, we believe that by considering a random sample that includes companies and users, we have eliminated the potential for sample bias and considered a sample that is at least representative of the Pinterest network as a whole, and in fact all CBN users.

Another limitation stems from the fact that we are not able to account for socially earned media effects that may exist outside of the CBN. It is possible that some users in the sample are known outside of the Pinterest network, and that this level of fame may encourage follower growth that we are not including in our model. To account for this, we limited our sample to only include users and companies that were not “household names”. In this way, we made sure that, even if fame effects do exist, they are kept to a minimum. The question of how major companies and brands operate within a CBN is interesting, however, and future research could explore ways to isolate follower growth effects attributable to the CBN even in the face of large brand recognition.

Finally, a third limitation of the study stems from the lack of network level data on this population. Researchers are placing increased importance on network effects and the structural characteristics of online social networks when evaluating firm performance and marketing outcomes (Jackson, 2006). While we are able to endogenate network growth as a series of observable marketing activities, there may be structural network characteristics that account for variance in our econometric model. For example, certain network users may occupy favorable positions within the larger Pinterest network, and realize marketing benefits as a result of these positions (Burt, 2009). This limitation is addressed by means of the additional data analysis incorporated into the third dissertation study, and indeed the limitation is a major part of the motivation for study 3.

6.2.4 Future Research

The findings of study 2 recommend a number of future research opportunities. Perhaps the most striking finding from study 2 concerns the total lack of evidence for the role of textual comments in CBN follower acquisition. It is tempting to take this finding at face value, but given the large amount of extant work that points to the power of textual eWOM, more work is needed to determine whether this is an anomaly of our data set or this particular research environment.

In study 2, we consider only the total number of comments attributed to a particular CBN user. As a result, it is not possible to make assumptions about the contents of these comments, or the sentiment that may or may not be present in a particular comment. One potential way that this analysis could be conducted would involve examining the content of CBN comments to see exactly how they may be influencing the nature of content distribution within the network. This analysis would help to determine if certain kinds of sentiment, for example positive vs. negative textual comments, play a more significant role in follower acquisition.

Another opportunity for future research involves the difference between companies and individual users. In study 2, we considered groups divided along the lines of content production and aggregation. However, in this sample we see individuals that fit the criteria of a content producer, and companies functioning as content aggregators. In order to increase the relevance of the findings for the study, it would be interesting to consider the content curation strategies of a sample of companies compared to a sample of users. As we stated when developing the motivation for study 2, a network of interested followers is a benefit for anyone operating within online social networks, be they an individual or a corporation. Additionally, the techniques and strategies that work for individuals may help to inform the activities of companies operating within the network. Future research would consider the performance of both of these groups to develop mixed strategies for maximizing follower acquisition.

6.3 Study 3

6.3.1 Contributions and Implications for Theory

Our third dissertation study, while descriptive and preliminary in nature, still makes several important contributions. First, the study makes a major contribution to eWOM research by examining the diffusion of a new type of eWOM content in the form of socially-earned repins. This contribution is distinct from that of study 2. Unlike study 2, where we consider the impact of socially-earned repins, here we actually model and study the process of diffusion as digital content spreads from one user to another. In study 2, we were able to show that this process has important implications for customer acquisition. Through the extension of this analysis, we contribute to our understanding of the way in which socially-earned repins actually spread, allowing us to predict the direction and scale of content diffusion. This will allow researchers and practitioners alike to better anticipate the outcomes of eWOM marketing initiatives CBN.

The magnitude of study 3's contribution in this regard is compounded by the examination of a sample of 30 distinct ego networks. Past studies have considered much smaller samples, normally looking at only a single network (Susarla et al., 2012a). Studying such a large sample gives the research generalizability and greater internal validity. Additionally, the study of multiple networks allows us to compare social network analysis metrics such as network density, centrality, etc. across multiple CBN structures.

Another major contribution of the study is its development of a method for endogenating CBN network structure. As we have discussed in chapter 2, and again at the beginning of chapter 5, a major challenge in social network research concerns the fact that network structures are typically only viewed cross-sectionally. As a result, only a snapshot of the network is visible, making it impossible to observe the network member actions that created the observed network structure (Jackson, 2006). This is problematic because without being able to endogenate network structures as a function of observed actions, it is not possible to identify the causal effects that underlie network formation (Mayer, 2009). We make a major theoretical contribution by studying the digital content curation actions that preclude the formation of CBNs, allowing us to endogenate CBN structures and identify these causal effects. This will give the study important relevance research in the areas of economics and information systems. As online social networks become more prevalent, it will increase the need to understand not only how information diffuses through networks, but the processes through which networks come into being. This study will help to fill this gap in the extant literature.

Finally, the study makes an important contribution by studying the impact of interest similarity on information diffusion. As we have discussed, interest similarity has been identified as an important variable in determining the ease and likelihood of information diffusion through online social network structures (Agarwal et al., 2009). However, to date no real methodology exists for measuring interest similarity between network nodes. Study 3 fills this gap in literature by developing such a measure for the CBN environment. This method will improve our understanding of interest similarity and information diffusion through CBN network structures, as well as answer calls in the wider area of social network research for techniques that can reveal the antecedents of information diffusion in other network contexts.

6.3.2 Contributions and Implications for Practice

Many of the contributions discussed in the previous section also have important implications for practice. Examining information diffusion in CBNs will inform companies regarding how best to get their message out to customers. Additionally, the methods that we employ in study 3 allow companies to identify the most influential customers and members of their follower network. For example, consider the network presented in Figure 16. The blue network nodes are those nodes that have repinned content from the central network user. Large nodes are obviously important network members, since they have repinned a large amount of content. Central network nodes are also valuable, since they are connected to a high number of other network members, and content on these users' pages is likely to be seen by many of the company's customers.

However, companies may actually want to focus on blue nodes that occupy the periphery of the network. These network actors have shown an interest in the content that the central company has to offer, yet they are not connected to many of the company's followers. It may be that these nodes are connected to users that do business with the company's competitors, or that they offer great opportunities for expanding the follower network into unexplored parts of the CBN. Once these nodes have been identified using the methods outlined in study 3, they can then be allocated additional marketing resources in order to improve future diffusion of marketing messages.

In addition to the data visualization methods that we discuss in study 3, methods for identifying interest similarity are also of real practical value. Using these methods, companies can identify those network members that are most closely aligned with their products and brands. Armed with this information, companies could then target those network members that exhibit high interest similarity, yet have not begun to actively distribute company content. These users would make good targets for conversion, as their interest similarity would improve their rate of adoption of company content.

6.4 Conclusion

This dissertation has presented a wide-spanning treatment of digital content curation in the online social network environment. In so doing, we have examined many aspects of digital content curation across different research contexts and units of analysis. In addition to providing an introduction to the notion of digital content curation in social networks, a new and exciting phenomenon worthy of the attention of the IS community, the dissertation's individual studies also make important contributions. The dissertation thus sets the stage for a large amount of future research directions.

Behavioral researchers should be interested in the methods and conclusions outlined in study one. Study two will be of interest to data-minded researchers interested in the areas of social network data-analytics and econometric modelling of social network data. Finally, study three is relevant to a growing area of IS researchers interested in social network analysis.

Additionally, practitioners should find plenty of value in the methods and conclusions offered by this dissertation project. Online social network designers continue to wrestle with which methods and designs are most useful to social network users. All three dissertation studies inform this work by offering different perspectives of analysis and answers to the question of how best to leverage curation-based network technology, whether the user be an individual or corporation.

It is our hope that the dissertation project will inform research and practice around digital content curation, and promote the study of curation-based networking in years to come. As we have discussed throughout the dissertation, digital content curation has the potential to exert far-reaching and industry-shaping forces in a myriad of areas including online social network usage, marketing and online information diffusion. As curation-based social network technologies continue to gain in popularity and prominence, the findings of this dissertation should remain relevant and interesting to the IS community for quite some time.

REFERENCES

Aaker, D., A. Biel. 1993. *Brand equity & advertising: advertising's role in building strong brands*, vol. 1. Psychology Press.

Abrams, D., R. Baecker, M. Chignell. 1998. Information archiving with bookmarks: personal web space constructionand organization. *Proceedings of the SIGCHI conference on Human factors in computingsystems*. ACM Press/Addison-Wesley Publishing Co., 41–48.

Agarwal, A., O. Rambow, N. Bhardwaj. 2009. Predicting interests of people on online social networks. *Computational Science and Engineering, 2009. CSE'09. InternationalConference on*, vol. 4. IEEE, 735–740.

Ahern, S., E. Dean, N. Good, S. King, M. Naaman, R. Nair. 2007. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. *Conference on Human Factors in Computing Systems: Proceedings of the SIGCHI conference on Human factors in computing systems*, vol. 28. 357–366.

Anderson, E.W., M.W. Sullivan. 1993. The antecedents and consequences of customer satisfaction for firms. *Marketing science* **12**(2) 125–143.

Anderson, J., J. Pérez-Carballo. 2001. The nature of indexing: how humans and machines analyze messages and texts for retrieval. part i: Research, and the nature of human indexing. *Information Processing & Management* **37**(2) 231–254.

André, P., J. Teevan, S.T. Dumais. 2009. From x-rays to silly putty via uranus: serendipity and its role inweb search. *Proceedings of the 27th international conference on Human factorsin computing systems*. ACM, 2033–2036.

Arndt, J. 1968. Word-of-mouth advertising and perceived risk. *Perspectives in consumer behavior* 330–36.

Assmus, G., J.U. Farley, D.R. Lehmann. 1984. How advertising affects sales: Meta-analysis of econometric results. *Journal of Marketing Research* **21**(1) 65–74.

Bampo, M., M.T. Ewing, D.R. Mather, D. Stewart, M. Wallace. 2008. The effects of the social structure of digital networks on viral marketing performance. *Information Systems Research* **19**(3) 273–290.

Bandura, A. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review* **84**(2) 191.

Bandura, A. 1989. Human agency in social cognitive theory. *American psychologist* **44**(9) 1175.

Bandura, A. 2001. Social cognitive theory: An agentic perspective. *Annual review of psychology* **52**(1) 1–26.

Barreau, D., B.A. Nardi. 1995. Finding and reminding: file organization from the desktop. *ACM SigChi Bulletin* **27**(3) 39–43.

Bass, F. 1969. A new product growth model for product diffusion. *Management Science* **15**(5) 215–227.

Baumeister, R.F., M.R. Leary. 1995. The need to belong: desire for interpersonal attachments as a fundamentalhuman motivation. *Psychological bulletin* **117**(3) 497.

Becker, H.S. 1951. The professional dance musician and his audience. *American Journal of Sociology* 136–144.

Berman, S.J., B. Battino, K. Feldman. 2011. New business models for emerging media and entertainment revenue opportunities. *Strategy & Leadership* **39**(3) 44–53.

Bikhchandani, Sushil, David Hirshleifer, Ivo Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *The Journal of Political Economy* **100**(5) 992–1026.

Binder, Jens F., Andrew Howes, Daniel Smart. 2012. Harmony and tension on social network sites. *Information, Communication & Society* **15**(9) 1279–1297.

Björneborn, L. 2008. Serendipity dimensions and users' information behaviour in the physical library interface. *Information Research* **13**(4) 13–4.

Blau, P.M. 1964. Exchange and power in social life. *Contemporary Sociological Theory* 112.

Borgatti, S.P., M.G. Everett, L.C. Freeman. 2002. Ucinet for windows: Software for social network analysis. *Harvard Analytic Technologies* **2006**.

Borgatti, Stephen P, Rob Cross. 2003. A relational view of information seeking and learning in social networks. *Management Science* **49**(4) 433.

Bosman, Julie. 2006. Chevy tries a write-your-own-ad approach, and the potshots fly. *New York Times* **4**.

Boyd, Danah. 2009. Why youth (heart) social network sites: The role of networked publics in teenage social life. *D. Buckingham (Ed.), Youth, Identity and Digital Media* 119–142.

Brandtzæg, P.B., M. Lüders, J.H. Skjetne. 2010. Too many facebook "friends"? content sharing and sociability versusthe need for privacy in social network sites. *Intl. Journal of Human–Computer Interaction* **26**(11-12) 1006–1030.

Burt, Ronald S. 2009. *Structural holes: The social structure of competition*. Harvard University Press.

Campbell, Donald T, Donald W Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin* **56**(2) 81–105.

Carr, Kelby. 2012. *Pinterest Marketing For Dummies*. For Dummies.

Chang, P.C., L.M. Quiroga. 2010. Using wikipedia's content for cross-website page recommendations that consider serendipity. *Technologies and Applications of Artificial Intelligence (TAAI), 2010 International Conference on*. IEEE, 293–298.

Chatterjee, Patrali, Donna L Hoffman, Thomas P Novak. 2003. Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Science* **22**(4) 520–541.

Cheung, Christy MK, Dimple R Thadani. 2012. The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems* **54**(1) 461–470.

Chevalier, J.A., D. Mayzlin. 2003. The effect of word of mouth on sales: Online book reviews. Tech. Rep. 3.

Choudhury, M., H. Sundaram, A. John, D.D. Seligmann. 2009. What makes conversations interesting?: themes, participants and consequencesof conversations in online social media. *Proceedings of the 18th international conference on World wide web*. ACM, 331–340.

Chowdhury, S., M. Landoni. 2006. News aggregator services: user expectations and experience. *Online Information Review* **30**(2) 100–115.

Christofides, E., A. Muise, S. Desmarais. 2009. Information disclosure and control on facebook: Are they two sidesof the same coin or two different processes? *CyberPsychology & Behavior* **12**(3) 341–345.

Chu, Shu-Chuan, Yoojung Kim. 2011. Determinants of consumer engagement in electronic word-of-mouth (ewom) in social networking sites. *International Journal of Advertising* **30**(1) 47–75.

Chunara, R., J.R. Andrews, J.S. Brownstein. 2012. Social and news media enable estimation of epidemiological patterns early in the 2010 haitian cholera outbreak. *The American Journal of Tropical Medicine and Hygiene* **86**(1) 39–45.

Church, M., AF Salam. 2010. Facebook, the spice of life? *Proceedings of the 31st International Conference on Information Systems (ICIS)*.

Clark, J., S. DeRose, et al. 1999. Xml path language (xpath) version 1.0.

Cockburn, A., S. Greenberg, B. McKenzie, S. Jasonsmith, M.and Kaasten. 1999. Webview: A graphical aid for revisiting web pages. *Proceedings of the OZCHI*, vol. 99. 15–22.

Compeau, Deborah R, Christopher A Higgins, ON London. 1995. Self-efficacy: Development of a measure and initial test. *MIS Quarterly* **19**(2) 189–211.

Corcoran, S. 2009. Defining earned, owned and paid media. *Forrester Research* .

Counts, S., E. Fellheimer. 2004. Supporting social presence through lightweight photo sharing on andoff the desktop. *Proceedings of the SIGCHI conference on Human factors in computingsystems*. ACM, 599–606.

Cronbach, Lee J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* **16**(3) 297–334.

Cross, R., S.P. Borgatti. 2004. The ties that share: Relational characteristics that facilitate informationseeking. M. Huysman, V. Wulf, eds., *Social capital and information technology*. MIT Press, Cambridge, MA, 137–161.

Csikszentmihalyi, M., E. Rochberg-Halton. 1981. *The meaning of things: Domestic symbols and the self*. Cambridge Univ Pr.

Dallas, C. 2007. An agency-oriented approach to digital curation theory and practice. *Proceedings of the International Symposium on Information and CommunicationTechnologies in Cultural Heritage*. Earthlab, 49.

Davis, F.D., R.P. Bagozzi, P.R. Warshaw. 1989. User acceptance of computer technology: a comparison of two theoreticalmodels. *Management science* **35**(8) 982–1003.

Debatin, B., J.P. Lovejoy, A.K. Horn, B.N. Hughes. 2009. Facebook and online privacy: Attitudes, behaviors, and unintendedconsequences. *Journal of Computer-Mediated Communication* **15**(1) 83–108.

Dellarocas, C. 2003. The digitization of word of mouth: Promise and challenges of onlinefeedback mechanisms. *Management science* **49**(10) 1407–1424.

Dellarocas, C. 2006. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science* **52**(10) 1577–1593.

Dellarocas, C., Z. Katona, W. Rand. 2013. Media, aggregators and the link economy: Strategic hyperlink formation in content networks. *Management Science* **59**.

Dichter, E. 1966. How word-of-mouth advertising works. *Harvard Business Review* **44**(6) 147–160.

DiMaggio, Paul, Eszter Hargittai, W Russell Neuman, John P Robinson. 2001. Social implications of the internet. *Annual review of sociology* 307–336.

Dominick, J.R. 1999. Who do you think you are? personal home pages and self-presentationon the world wide web. *Journalism & Mass Communication Quarterly* **76**(4) 646–658.

Donaton, Scott. 2006. How to thrive in the new world of user-created content: let go. *Advertising Age* **77**(18) 38.

Döring, N. 2002. Personal home pages on the web: A review of research. *Journal of Computer-Mediated Communication* **7**(3).

Doyle, E., I. Stamouli, M. Huggard. 2005. Computer anxiety, self-efficacy, computer experience: An investigationthroughout a computer science degree. *Frontiers in Education, 2005. FIE'05. Proceedings 35th Annual Conference*. IEEE.

Dumais, S., E. Cutrell, J.J. Cadiz, G. Jancke, R. Sarin, D.C. Robbins. 2003. Stuff i've seen: a system for personal information retrieval and re-use. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 72–79.

Durrant, A., D. Frohlich, A. Sellen, E. Lyons. 2009a. Home curation versus teenage photography: Photo displays in the familyhome. *International Journal of Human-Computer Studies* **67**(12) 1005–1023.

Durrant, A., A.S. Taylor, D. Frohlich, D. Sellen, A. andUzzell. 2009b. Photo displays and intergenerational relationships in the familyhome. *Proceedings of the 23rd British HCI*

Group Annual Conference on Peopleand Computers: Celebrating People and Technology. British Computer Society, 10–19.

Dwyer, C., S.R. Hiltz, K. Passerini. 2007. Trust and privacy concern within social networking sites: A comparisonof facebook and myspace. *Proceedings of AMCIS 2007*. Citeseer, 1–12.

Ellison, N.B., R. Heino, J. Gibbs. 2006. Managing impressions online: Self-presentation processes in the onlinedating environment. *Journal of Computer-Mediated Communication* **11**(2) 415–441.

Ellison, N.B., C. Steinfield, C. Lampe. 2007a. The benefits of facebook âfriends:â social capital and college students'use of online social network sites. *Journal of Computer-Mediated Communication* **12**(4) 1143–1168.

Ellison, N.B., C. Steinfield, C. Lampe. 2011. Connection strategies: Social capital implications of facebook-enabledcommunication practices. *New Media & Society* **13**(6) 873–892.

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

Erdős, P., A. Rényi. 1961. On the strength of connectedness of a random graph. *Acta Mathematica Hungarica* **12**(1) 261–267.

Ericsson, K.A., R.T. Krampe, C. Tesch-Römer. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological review* **100**(3) 363.

Evans, D., J. McKee. 2010. *Social media marketing: the next generation of business engagement*. Sybex.

Everett, Martin, Stephen P Borgatti. 2005. Ego network betweenness. *Social networks* **27**(1) 31–38.

Fan, X., J. Mostafa, K. Mane, C. Sugimoto. 2012. Personalization is not a panacea: balancing serendipity and personalizationin medical news content delivery. *Proceedings of the 2nd ACM SIGHIT symposium on International healthinformatics*. ACM, 709–714.

Fetscherin, M., C. Lattemann. 2008. User acceptance of virtual worlds. *Journal of Electronic Commerce Research* **9**(3) 231–242.

Fornell, C., B. Wernerfelt. 1987. Defensive marketing strategy by customer complaint management: a theoretical analysis. *Journal of marketing research* **24**(4) 337–346.

Fornell, Claes, David F Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research* **18**(1) 39–50.

Foster, A., N. Ford. 2003. Serendipity and information seeking: an empirical study. *Journal of Documentation* **59**(3) 321–340.

Gallaugher, J., S. Ransbotham. 2010. Social media and customer dialog management at starbucks. *MIS Quarterly Executive* **9**(4) 197–212.

Gangadharbatla, H. 2008. Facebook me: Collective self-esteem, need to belong, and internetself-efficacy as predictors of the igeneration's attitudes towardsocial networking sites. *Journal of interactive advertising* **8**(2) 5–15.

Garg, R., M.D. Smith, R. Telang. 2011. Measuring information diffusion in an online community. *Journal of Management Information Systems* **28**(2) 11–38.

Garvey, P. 2001. Organized disorder: moving furniture in norwegian homes. D. Miller, ed., *Home possessions: material culture behind closed doors*. Berg Publishers, 47–68.

Gatignon, H., T.S. Robertson, et al. 1985. An exchange theory model of interpersonal communication. *Advances in Consumer Research* **12**(1) 534–538.

Giacalone, R.A., P. Rosenfeld. 1989. *Impression management in the organization*. Lawrence Erlbaum.

Giddens, A. 1991. *Modernity and self-identity: Self and society in the late modernage*. Stanford Univ Pr.

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

Goffman, E. 2002. The presentation of self in everyday life. 1959. *Garden City, NY* .

Golder, S.A., D.M. Wilkinson, B.A. Huberman. 2007. Rhythms of social interaction: Messaging within a massive online network. *Communities and Technologies 2007* 41–66.

Granovetter, M.S. 1995. *Getting a job: A study of contacts and careers*. University of Chicago Press.

Greene, William. 2004. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal* **7**(1) 98–119.

Grewal, R., G.L. Lilien, G. Mallapragada. 2006. Location, location, location: How network embeddedness affects projectsuccess in open source systems. *Management Science* **52**(7) 1043–1056.

Guy, M., A. Ball, M. Day. 2009. Missing links: The enduring web. *International Journal of Digital Curation* **4**(2) 135–143.

Haferkamp, Nina, Sabrina C Eimler, Anna-Margarita Papadakis, Jana Vanessa Kruck. 2012. Men are from mars, women are from venus? examining gender differences in self-presentation on social networking sites. *Cyberpsychology, Behavior, and Social Networking* **15**(2) 91–98.

Hart, Stuart L, Sanjay Sharma. 2004. Engaging fringe stakeholders for competitive imagination. *The Academy of Management Executive* **18**(1) 7–18.

Hasan, B. 2006. Delineating the effects of general and system-specific computer self-efficacybeliefs on is acceptance. *Information & Management* **43**(5) 565–571.

Higgins, E.T., G.A. King, G.H. Mavin. 1982. Individual construct accessibility and subjective impressions andrecall. *Journal of Personality and Social Psychology* **43**(1) 35–47.

Higgins, M.C., K.E. Kram. 2001. Reconceptualizing mentoring at work: A developmental network perspective. *Academy of Management Review* **26**(2) 264–288.

Hinz, O., M. Spann, I.H. Hann. 2010. Prestige goods and social status in virtual worlds. *ICIS 2012 Proceedings* .

Hong, Seung-Hyun. 2013. Measuring the effect of napster on recorded music sales: Difference-in-differences estimates under compositional changes. *Journal of Applied Econometrics* **28**(2) 297–324.

Hsu, M.H., C.M. Chiu. 2004. Internet self-efficacy and electronic service acceptance. *Decision support systems* **38**(3) 369–381.

Hu, M., B. Liu. 2004. Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledgediscovery and data mining*. ACM, 168–177.

Hu, Nan, Paul A Pavlou, Jennifer Zhang. 2006. Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of online word-of-mouth communication. *Proceedings of the 7th ACM conference on Electronic commerce*. ACM, 324–330.

Huwe, T.K. 1999. New search tools for multidisciplinary digital libraries. *Online* **23**(2) 67–74.

Jackson, L.A., K.S. Ervin, P.D. Gardner, N. Schmitt. 2001. Gender and the internet: Women communicating and men searching. *Sex roles* **44**(5) 363–379.

Jackson, M. 2005. A survey of network formation models: stability and efficiency. *Group formation in economics: Networks, clubs and coalitions* 11–57.

Jackson, M. 2006. The economics of social networks. *Econometric Society Monographs* **41** 1.

Jarvis, J. 2008. The link economy vs. the content economy. buzzmachine blog, june18, 2008.

Jeong, E., S. Jang. 2011. Restaurant experiences triggering positive electronic word-of-mouth (ewom) motivations. *International Journal of Hospitality Management* **30**(2) 356–366.

Jones, W., H. Bruce, S. Dumais. 2001. Keeping found things found on the web. *Proceedings of the tenth international conference on Informationand knowledge management*. ACM, 119–126.

Jones, W., S. Dumais, H. Bruce. 2002. Once found, what then? a study of âkeepingâ behaviors in the personaluse of web information. *Proceedings of the American Society for Information Science and Technology* **39**(1) 391–402.

Jung, T., H. Youn, S. Mcclung. 2007. Motivations and self-presentation strategies on korean-based" cyworld"weblog format personal homepages. *CyberPsychology & Behavior* **10**(1) 24–31.

Kavanaugh, Andrea L, Scott J Patterson. 2001. The impact of community computer networks on social capital and community involvement. *American Behavioral Scientist* **45**(3) 496–509.

Kaye, Barbara K, Thomas J Johnson. 2003. From here to obscurity?: Media substitution theory and traditional media in an on-line world. *Journal of the American Society for Information Science and Technology* **54**(3) 260–273.

Kim, Hee-Woong, Hock Chuan Chan, Atreyi Kankanhalli. 2012. What motivates people to purchase digital items on virtual community websites? the desire for online self-presentation. *Information Systems Research* **23**(4) 1232–1245.

Kim, Junyong, Pranjal Gupta. 2012. Emotional expressions in online user reviews: How they influence consumers' product evaluations. *Journal of Business Research* **65**(7) 985–992.

King, R.C., W. Xia. 1997. Media appropriateness: Effects of experience on communication mediachoice. *Decision Sciences* **28**(4) 877–910.

Kock, N. 2007. Media naturalness and compensatory encoding: the burden of electronicmedia obstacles is on senders. *Decision Support Systems* **44**(1) 175–187.

Kwai Fun, R., C. Wagner. 2008. Weblogging: A study of social computing and its impact on organizations. *Decision Support Systems* **45**(2) 242–250.

Lange, P. 2007. Commenting on comments: Investigating responses to antagonism on youtube. *Annual Conference of the Society for Applied Anthropology*, vol. 29.

Latane, B. 1981. The psychology of social impact. *American Psychologist* **36**(4) 343–356.

Lewis, K., M. Gonzalez, J. Kaufman. 2012. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences* **109**(1) 68–72.

Liaw, S.S., H.M. Huang. 2003. An investigation of user attitudes toward search engines as an informationretrieval tool. *Computers in human behavior* **19**(6) 751–765.

Lim, K.H., I. Benbasat. 2002. The influence of multimedia on improving the comprehension of organizationalinformation. *Journal of Management Information Systems* **19**(1) 99–128.

Liu, Yabing, Krishna P Gummadi, Balachander Krishnamurthy, Alan Mislove. 2011. Analyzing facebook privacy settings: User expectations vs. reality. *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, 61–70.

Lynch, C.A. 1995. Networked information resource discovery: an overview of current issues. *IEEE Journal on Selected Areas in Communications* **13**(8) 1505–1522.

Malone, T.W. 1983. How do people organize their desks?: Implications for the designof office information systems. *ACM Transactions on Information Systems (TOIS)* **1**(1) 99–112.

Mangold, W Glynn, David J Faulds. 2009. Social media: The new hybrid element of the promotion mix. *Business horizons* **52**(4) 357–365.

Marshall, C.C., F.M. Shipman III. 1997. Spatial hypertext and the practice of information triage. *Proceedings of the eighth ACM conference on Hypertext*. ACM, 124–133.

Mayer, A. 2009. Online social networks in economics. *Decision Support Systems* **47**(3) 169–184.

McCay-Peet, Lori, EG Toms. 2011. Measuring the dimensions of serendipity in digital environments. *Information Research* **16**(3) 16–3.

McCourt, T., P. Burkart. 2003. When creators, corporations and consumers collide: Napster and thedevelopment of on-line music distribution. *Media, Culture & Society* **25**(3) 333–350.

McPherson, G.E., J. McCormick. 2006. Self-efficacy and music performance. *Psychology of Music* **34**(3) 322–336.

Mehrabian, A., S.G. Diamond. 1971. Effects of furniture arrangement, props, and personality on socialinteraction. *Journal of Personality and Social Psychology* **20**(1) 18–30.

Miles, E.W., D.G. Leathers. 1984. The impact of aesthetic and professionally-related objects on credibilityin the office setting. *Southern Journal of Communication* **49**(4) 361–379.

Milgram, S. 1967. The small world problem. *Psychology today* **2**(1) 60–67.

Miller, D. 2001. *Home possessions: material culture behind closed doors*. Berg Publishers.

Miller, N. 2012. *The Facebook IPO Primer*.

Misner, I.R. 1994. *The world's best-known marketing secret: building your business with word-of-mouth marketing*. Bard & Stephen Austin, TX.

Mudambi, S.M., D. Schuff. 2010. What makes a helpful online review? a study of customer reviews on amazon. com. *MIS Quarterly* **34**(1) 185–200.

Murphy, C.A., D. Coover, S.V. Owen. 1989. Development and validation of the computer self-efficacy scale. *Educational and Psychological Measurement* **49**(4) 893–899.

Nault, B.R., M.B. Vandenbosch. 2000. Research report: Disruptive technologies-explaining entry in next generation information technology markets. *Information Systems Research* **11**(3) 304–319.

Ono, H., M. Zavodny. 2003. Gender and the internet*. *Social Science Quarterly* **84**(1) 111–121.

Osnos, P. 2009. What'sa fair share in the age of google? *Columbia Journalism Review* **47**(5) 25–28.

Otte, Evelien, Ronald Rousseau. 2002. Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science* **28**(6) 441–453.

Papacharissi, Z. 2002. The self online: The utility of personal home pages. *Journal of Broadcasting & Electronic Media* **46**(3) 346–368.

Parameswaran, M., A.B. Whinston. 2007. Research issues in social computing. *Journal of the Association for Information Systems* **8**(6) 336–350.

Passant, A., I. Mulvany, P. Mika, N. Maisonneauve, A. Löser, C. Cattuto, C. Bizer, C. Bauckhage, H. Alani. 2008. Mining for social serendipity. *Dagstuhl Seminar on Social Web Communities* .

Pheterson, G.I., S.B. Kiesler, P.A. Goldberg. 1971. Evaluation of the performance of women as a function of their sex, achievement, and personal history. *Journal of Personality and Social Psychology* **19**(1) 114–118.

Prell, Christina. 2011. *Social network analysis: History, theory and methodology*. SAGE Publications Limited.

Putnam, Robert D. 1995. Bowling alone: America's declining social capital. *Journal of democracy* **6**(1) 65–78.

Rafaeli, S., G. Ravid, V. Soroka. 2004. De-lurking in virtual communities: a social communication network approach to measuring the effects of social and cultural capital. *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*. Ieee.

Ransbotham, S., G. Kane. 2011. Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in wikipedia. *MIS Quarterly* **35**(3) 613–627.

Robert, L.P., A.R. Dennis. 2005. Paradox of richness: A cognitive model of media choice. *Professional Communication, IEEE Transactions on* **48**(1) 10–21.

Roberts, N., P.S. Galluch, M. Dinger, V. Grover. 2012. Absorptive capacity and information systems research: Review, synthesis, and directions for future research. *Information Systems* **6**(1) 25–40.

Rogers, E.M. 1983. *Diffusion of innovations, 3rd ed.*. Free Press, New York, NY.

Roloff, Michael E. 1981. *Interpersonal communication: The social exchange approach*. Sage Publications Beverly Hills, CA.

Safko, L. 2010. *The social media bible: tactics, tools, and strategies for business success*. Wiley.

Schau, H.J., M.C. Gilly. 2003. We are what we post? self-presentation in personal web space. *Journal of consumer research* **30**(3) 385–404.

Schellenberg, E.G. 2006. Long-term positive associations between music lessons and iq. *Journal of Educational Psychology* **98**(2) 457.

Schwab, A., A.S. Miner. 2008. Learning in hybrid-project systems: The effects of project performance on repeated collaboration. *The Academy of Management Journal* **51**(6) 1117–1149.

Seneca, T. 2009. The web-at-risk at three: overview of an ndiipp web archiving initiative .

Shang, Rong-An, Yu-Chen Chen, Hsueh-Jung Liao. 2006. The value of participation in virtual consumer communities on brand loyalty. *Internet Research* **16**(4) 398–418.

Shin, D.H. 2008. Understanding purchasing behaviors in a virtual economy: Consumer behavior involving virtual currency in web 2.0 communities. *Interacting with computers* **20**(4) 433–446.

Shklovski, I. 2012. Social network sites: Indispensable or optional social tools? *Sixth International AAAI Conference on Weblogs and Social Media*.

Sijtsma, Klaas. 2009. On the use, the misuse, and the very limited usefulness of cronbach's alpha. *Psychometrika* **74**(1) 107–120.

Sommer, R., H. Ross. 1958. Social interaction on a geriatrics ward. *International Journal of Social Psychiatry* **4**(2) 128–133.

Son, J.Y., S.S. Kim. 2008. Internet users' information privacy-protective responses: A taxonomy and a nomological model. *MIS Quarterly* **32**(3) 503–529.

Stephen, A.T., J. Galak. 2012. The effects of traditional and social earned media on sales: A study of a microlending marketplace. *Journal of Marketing Research* **49**.

Straub, D.W. 1989. Validating instruments in mis research. *MIS quarterly* **13**(2) 147–169.

Sumi, Y., J. Ito, T. Nishida. 2008. Photochat: communication support system based on sharing photos and notes. *CHI'08 extended abstracts on Human factors in computing systems*. ACM, 3237–3242.

Susarla, A., J. Oh, Y. Tan. 2012a. Influentials or susceptibles? analyzing cascades of word-of-mouth conversations in online social networks (working paper) .

Susarla, Anjana, Jeong-Ha Oh, Yong Tan. 2012b. Social networks and the diffusion of user-generated content: Evidence from youtube. *Information Systems Research* **23**(1) 23–41.

Thambusamy, R., M. Church, H. Nemati, J. Barrick. 2010. Socially exchanging privacy for pleasure: Hedonic use of computer-mediated social networks. *ICIS 2010 Proceedings*.

Thorson, K.S., S. Rodgers. 2006. Relationships between blogs as ewom and interactivity, perceivedinteractivity, and parasocial interaction. *Journal of Interactive Advertising* **6**(2) 39–50.

Trusov, M., R.E. Bucklin, K. Pauwels. 2009. Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing* **73**(5) 90–102.

Utz, S., N. Kramer. 2009. The privacy paradox on social network sites revisited: The role of individual characteristics and group norms. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* **3**(2).

Van Audenhove, L., S. Delaere, P. Ballon, M. Van Bossuyt. 2007. Changing content industry structures: The case of digital newspapers on epaper mobile devices. *11th International Conference on Electronic Publishing, Vienna, Austria*.

Venkatesh, V., M.G. Morris, G.B. Davis, F.D. Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS quarterly* **27**(3) 425–478.

Vilpponen, Antti, Susanna Winter, Sanna Sundqvist. 2006. Electronic word-of-mouth in online environments: exploring referral network structure and adoption behavior. *Journal of Interactive Advertising* **6**(2) 71–86.

Wasko, Molly McLure, Samer Faraj. 2005. Why should i share? examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly* **29**(1) 35–57.

Watts, D.J. 1999. Networks, dynamics, and the small-world phenomenon 1. *American Journal of Sociology* **105**(2) 493–527.

White, Halbert. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society* **48**(4) 817–838.

Yakel, Elizabeth. 2007. Digital curation. *OCLC Systems & Services* **23**(4) 335–340.

Young, K. 2011. Social ties, social networks and the facebook experience. *International Journal of Emerging Technologies and Society* **9**(1) 20–34.

Zhang, W., S.A. Watts. 2008. Capitalizing on content: Information adoption in two online communities. *Journal of the association for information systems* **9**(2) 73–94.

APPENDIX A

STUDY ONE INSTRUMENT AND FACTOR LOADINGS

Desire for Online Self-Presentation Seven-point scales anchored with “strongly disagree” and “strongly agree” (Kim, Chan, and Kankanhalli, 2011) [1]

(1) I want to establish a preferred image for myself in Pinterest.[.81]

(2) I want to present my image in Pinterest[.82]

(3) I want to project an image of myself in Pinterest[.96]

(4) I want to give a preferred impression about myself to others in Pinterest[.93]

Object-based presentation self-efficacy Seven-point scales anchored with “strongly disagree” and “strongly agree” (Kim, Chan, and Kankanhalli, 2011)

(1) I know how to express myself to others in Pinterest.[.69]

(2) I know how to present my image reasonably well on my own in Pinterest.[.87]

(3) I can decorate my Pinterest pinboards reasonably well on my own in Pinterest.[.87]

(4) I know how to create my image reasonably well on my own in Pinterest.[.91]

Digital content curation intentions Seven-point scales anchored with “strongly disagree” and “strongly agree” . (Added)

(1) The probability that I will pin content to my Pinterest page within the next six months is high.[.95]

[1] p-values for all items $< .001$. The first item for each construct does not have a p-value as it is automatically fixed to a value of 1.00

(2) My willingness to pin content to my Pinterest page within the next six months is high.[.96]

(3) The likelihood of my pinning content to my Pinterest page within the next six months is high.[.96]

Serendipitous Information Discovery Seven-point scales anchored with “strongly disagree” and “strongly agree”

(1) I feel that Pinterest enables me to make connections between different topics.[.69]

(2) I feel that Pinterest presents content in ways that invite me to explore across topics.[.85]

(3) I believe Pinterest encourages me to browse and explore.[.89]

(4) I feel that Pinterest lets me explore topics I do not normally examine.[.77]

(5) I believe that Pinterest shows me unexpected things that catch my eye.[.86]

(6) On Pinterest, I find myself pausing to look at things more closely.[.86]

(7) I want to click on things to see where they take me on Pinterest.[.76]

Information Re-Finding Seven-point scales anchored with “strongly disagree” and “strongly agree” (Adapted from Liaw and Huang, 2003)

(1) I believe it is easy to use Pinterest to store information I want to find again.[.82]

(2) I believe Pinterest is an effective tool for finding information a second time.[.93]

(3) I believe using Pinterest can help me "re-find" useful information. [.90]

Information Re-Minding Seven-point scales anchored with “strongly disagree” and “strongly agree” (Adapted from Liaw and Huang, 2003)

(1) I believe it is easy to use Pinterest to store information I want to be reminded of again later.[.85]

(2) I believe Pinterest is an effective tool for reminding me about information.[.93]

(3) I believe using Pinterest can remind me of useful information. [.93]

(4) I believe using Pinterest can help me remember things. [.88]

Virtual Community Involvement Seven-point scales anchored with “strongly disagree” and “strongly agree”

(1) Participating in the Pinterest community is one of the most enjoyable things I do.[.85]

(2) Participating in the Pinterest community is important to me.[.96]

(3) Participating in the Pinterest community is pleasurable to me.[.75]

(4) Participating in the Pinterest community means a lot to me.[.91]

Privacy Expectations Seven-point scales anchored with “strongly disagree” and “strongly agree” (Agarwal and Rodhain, 2002)

(1) I believe other Pinterest users have the right to look at what I pin.[.69]

(2) When I pin content to Pinterest, I typically do not have an expectation of privacy.[.86]

(3) Anything I pin to Pinterest is not exclusively my property. [.75]

(4) I do not consider Pinterest content to be personal, private information about me. [.85]