CREATING VALUE ON LARGE ONLINE SOCIAL NETWORK SITES (SNS): THE CASE FOR USERS, PROVIDERS AND ADVERTISERS

Prasanta Bhattacharya

(M.E., BITS Pilani (Pilani, India))

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF INFORMATION SYSTEMS

NATIONAL UNIVERSITY OF SINGAPORE

2016

Supervisor:

Assistant Professor Tuan Q. Phan

Examiners:

Professor Atreyi Kankanhalli

Associate Professor Ke-Wei Huang

Associate Professor Gal Oestreicher-Singer, Tel Aviv University

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis. This thesis has also not been submitted for any degree in any university previously.

P. Bhattacharya.

Prasanta Bhattacharya
29 July 2016

ACKNOWLEDGEMENTS

A doctoral program is a long journey and no journey is complete without extending a token of gratitude to the many individuals who continue to guide and enable us through this challenging road.

First and foremost, I would like to express my deepest gratitude to my doctoral advisor, Prof. Tuan Q Phan who has been instrumental in guiding me through my journey at NUS, and in helping me mature into a more productive researcher. I have gained tremendously over the past four years from our regular interactions and engagements on projects, papers, and other lab activities. In particular, his keen interest in seeking out challenging research problems and his insistence on pursuing ambitious academic targets continues to impress and inspire me.

Next, I would like to convey a sincere note of thanks to my doctoral committee members and examiners, Prof. Atreyi Kankanhalli, Prof. Ke-Wei Huang, and Prof. Khim Yong Goh. While Prof. Kankanhalli's pointed suggestions on the theorizing and positioning of the essays have immensely helped in developing this dissertation, Prof. Huang's and Prof. Goh's constant guidance on improving the empirical strategies has been equally valuable over the years. As a result of their constant support, I have improved substantially in producing research of acceptable quality.

A conducive working environment is key to nurturing high productivity and research aptitude. For this reason, I would also like to extend my deepest appreciation towards to current Head of the Dept., Prof. Jungpil Hahn, as well as the former Head, Prof. Hock Hai Teo in providing me with the necessary opportunities and environment to grow as a researcher. A special mention also to Prof. Ke-Wei Huang and Prof. Isam Faik for coordinating the IS Brown Bag Seminar and the IS Teaching Seminars, which have provided many Ph.D students with great opportunities to present and discuss their ongoing research from time to time with peers, as well as the faculty.

I have been extremely fortunate to have had the opportunity of working with very supportive and encouraging collaborators from outside NUS, over the tenure of this thesis. In particular, I would like to extend my heartfelt gratitude to Prof. Edoardo Airoldi (Harvard Univ.) and Prof. Xue Bai (Univ. of Connecticut). Also, I would like to extend a special note of thanks to my good friend and long-time collaborator, Rishabh Mehrotra (UC London) who has been a constant source of exciting research ideas.

My time at NUS has been extremely joyous, the primary reason for which lies with the friends I have made within the IS Dept. and beyond. My deepest thanks to all my peers, seniors and juniors at the IS Dept. for being so supporting and encouraging throughout this journey. They have always been my best friends, strongest supporters and finest critics, all at the same time. Thanks, guys!

A concluding note of thanks to my family, back in India, who've made innumerable sacrifices to help me reach this point in my academic journey. However, I hope they will not mind if I choose, instead, to dedicate this dissertation to my junior and senior high school teachers, the unsung heroes of our education system. Thank you teachers, for all that you have done for me. You are all fondly in my mind as I write this note of acknowledgment today.

TABLE OF CONTENTS

SUMMARY

Social media and, in particular, online social network sites (SNS) have emerged as some of the most popular venues on the Internet. With the rapid growth of SNS, a wide range of stakeholders from individual users to the platform owners, digital marketers and policy makers have become interested in understanding how user engagement on these platforms generate value for them. While there has been a growing stream of scholarly work in the area of value creation in organizational settings (Kang et al. 2007; Lepak et al. 2007; Tsai and Ghoshal 1998), this dissertation contributes towards developing (i) an understanding of what constitutes “value” on large and online social networks, and (ii) how users engage with each other, as well as with the platform to create such value for the different stakeholders.

The SNS users seek and obtain various types of gratifications for themselves by participating on the platform (e.g. subjective well-being, information, entertainment, social status etc.) (LaRose and Eastin 2004; Nadkarni and Hofmann 2012; Raacke and Bonds-Raacke 2008), but they also create value for the SNS provider, and advertisers through the content they generate, the social connections they form, and through other activities like joining "fan" pages etc. For the SNS provider, the market valuation of their platform is directly linked to the amount of user activity on their platform, and thus, a higher user engagement on the SNS directly correlates to a higher economic value for the SNS provider (Gobry 2012; La Monica 2016). Similarly, for advertising brands and digital marketers, the SNS provides a useful new channel to reach out to existing as well as potential customers with informational as well as promotional content (K. Goh et al. 2013; Lee et al.

2014). In this way, a higher engagement on the SNS also creates a fertile ground for brands and advertisers to reap significant economic benefits.

Through this dissertation, I take a computational social science approach to studying three instances of such value creation on the SNS, aimed at (i) external brands and advertisers, (ii) the SNS users and (iii) the SNS provider respectively.

In my first study, I analyze the act of joining a brand-sponsored fan-page on a large SNS and present evidence that the brand's customers, on average, spend less and buy fewer items upon joining the brand page. I conjecture that this is perhaps driven by the fact that such brand pages on the SNS are often strategically used by the brand to lure users with utilitarian benefits, like price-discount coupons etc. However, using a suite of text-mining and econometrics based methods I identify users who are also likely driven by self-presentational motives to join these brand pages. I show that for these set of users, the reduction in expenditure upon joining the brand page is largely attenuated. Advertising brands as well as the SNS providers can leverage insights from this study to suitably target valuable users on the SNS.

In my second study, I analyze how users create value for themselves on the platform by broadcasting content and forming stable social connections, and model the evolution of these two value creation processes over time. Specifically, I study how network effects (e.g. homophilous friendship formation, peer-influence etc.) impact users' content generation behavior on the SNS, by jointly estimating the evolution of the user's network structure and the user's content posting behavior. My results show that while users choose to form new friendship ties on the platform based on similarity in posting behavior, they tend to diverge from similar others once they become friends. The results offer novel theoretical insights on the interplay between homophily and social influence in the context of content generation on the SNS, while also offering actionable recommendations to SNS providers on the design of suitable friend recommendation systems, and content filters.

In my third and final study, I focus on the third stakeholder in the value creation process viz. the SNS provider. I exploit the introduction of new privacy controls by a major SNS provider as a quasi-experimental context to study the impact of this intervention on users' content generation behavior. While past studies have highlighted the importance of studying privacy-related interventions in online contexts, this is among the first empirical attempts at quantifying its effects on user behavior in a real-world setting across both public as well as private channels. Through my analyses, I find that while users do not change the volume of public content generation, they significantly reduce the volume of private conversations they have on the platform. Moreover, we show that this effect is stronger for users who are not too privacy conscious, and weaker for users who tend to be more privacy conscious. The results from this study offer theoretical insights into how users in a networked context react to privacy related interventions, and how SNS providers can anticipate responses to such feature changes on their platform.

Through my dissertation, I seek to not only answer important questions about how users create value for different stakeholders on the SNS, but also emphasize how computational methods can complement more traditional social science approaches in trying to effectively answer these questions. For example, the first study employs a combination of text mining techniques and reduced-form econometric models. The second study uses a stochastic structural model, while the third study applies this structural model in a real world quasi-experimental setting to make stronger claims on causality.

LIST OF TABLES

LIST OF FIGURES

Chapter 1. INTRODUCTION

1.1 Background and Motivation

The past decade has seen a phenomenal growth in the emergence and popularity of social network sites (SNS) on the Internet. For instance, Facebook, the world's largest online social network has consistently ranked within the top 3 most visited websites in recent years, and as many as 3 out of the top 10 most visited websites on the Internet are SNS[1]. What is perhaps more interesting is that many online platforms that were traditionally not SNS-based (e.g. e-commerce) are now incorporating social network features into their platform to increase user engagement (e.g. by using Facebook/Twitter widgets etc.) (Ismail 2011). While most of the early studies on SNS have focused on an individual's motives for using the SNS (boyd and Ellison 2007; Park et al. 2009; Raacke and Bonds-Raacke 2008; Steinfield et al. 2008; Wilson et al. 2012), and on the outcomes of SNS use (Livingstone 2008; Loss et al. 2013; Wilcox and Stephen 2013), past work has afforded scant focus on understanding and quantifying the value generated on the SNS for the three major stakeholders - the individual user, the SNS provider and external businesses, like advertisers and marketers. However, value creation on the SNS is an area of great interest for businesses. However, it is important to first conceptualize the meaning of "value" in the context of large online SNS. Broadly speaking, value is the net worth of a product, service or entity in terms of something else.[2,3] On the SNS, value is created primarily through

[1] http://www.alexa.com/topsites
[2] For example, if a loaf of bread costs $3, that is the value of the bread in that particular currency.
[3] Value is slightly different from profit, which is the difference between benefit and cost. One

user participation, although it is a challenging exercise to quantify this value in any relevant metric (e.g. the revenue it generates for the SNS provider). Recent media reports suggest that there is a fair amount of uncertainty among firms and users, in understanding what the true value of participation on the SNS is (Dekel 2013; Wilson 2015). In this dissertation, I contend that a SNS acts as a dual-sided platform (Eisenmann et al. 2006) and allows users to create value for themselves by generating informative and/or entertaining content, forging new social ties, and engaging with platform features. In turn, however, the user also creates a lot of value for the other SNS stakeholders, namely the SNS providers and advertising brands. A key source of this externality comes from the large amounts of content that is being generated on the SNS every minute.

Over the past few years, SNS, and social media in general, have emerged as a leading source of user-generated content (Chen et al. 2012). There now exists over 3 billion active Internet users in the world, of whom, close to 1.7 billion have active social media accounts (Bullas 2015). Twitter, the world's leading microblogging platform, now has over 280 million active users who collectively generate over 500 million tweets every day (Bullas 2015). This spurt in user engagement on the SNS has opened up two major opportunities for both IS practitioners and IS researchers to conceptualize and quantify value created on the SNS. First, social media platform owners and digital advertisers can now leverage this new source of information about their consumers to improve product targeting and advertising effectiveness (Hoffman and Fodor 2010). For instance, a recent report suggests that in 2015,

---

can generate value without creating profit, and vice versa, although this reverse is relatively uncommon social network sites (SNS) worldwide earned over $25billion in advertising revenues alone (Statista 2015). Second, user activity on social media platforms produces large amounts of digital traces that can be exploited by social science researchers to analyze and generate novel theoretical insights on how users engage with certain SNS features (e.g. community pages) as well as with other users on the SNS (Cioffi-Revilla 2010; Lazer et al. 2009).

Drawing on these two opportunities, my dissertation attempts to shines light on how users achieve well-being and other personal gratifications for themselves, while also generating economic value on the SNS for organizational stakeholders (e.g. SNS owners, digital marketers). More specifically, I conduct three empirical investigations that help to illustrate the roles of, and interplay between, the following three processes that are instrumental in generating such value viz. (i) content generation, (ii) network (*or friendship*) formation, and (iii) interaction with platform features (e.g. brand communities, privacy settings etc.)[4]. Through my dissertation, I build upon and contribute to a growing stream of literature that illustrates how analyzing such online behavior is not just an important theoretical exercise, but also has strong practical utility for organizations as a method to understand the digital activities, product preferences and privacy-seeking behavior of their consumers. (Aral and Walker 2012; Goldfarb and Tucker 2011; Kumar et al. 2013; Rishika et al. 2013; Tucker 2014).

1.2 Value Creation on Social Network Sites

An important contribution of this dissertation is in understanding what

[4] By "platform features" I refer to the list of all the other platform features that are not related to the two primary processes on the SNS viz. content generation and friendship formation.

constitutes value on the SNS. The idea of value creation has been central to organizational research, and several studies in the past have looked at different facets of this value creation, such as the role of relational archetypes and organizational learning (Kang et al. 2007), inter-firm networks and the accumulation of social capital (Tsai and Ghoshal 1998), the idea of value creation and capture across multiple levels of analysis (Lepak et al. 2007), and informational competencies across horizontal levels in an organization (Peppard et al. 2000). Across the studies, a number of definitions have been proposed to explain what it means to create value for organizations. For instance, Lepak et al. (2007) conceptualizes "value" quite simply as the difference between the benefits derived and the costs incurred. Moran and Ghoshal (1996) define value creation from a resource-based view of the firm (Barney 1991; Wernerfelt 1984) and argue that value emerges as a result of combining and exchanging firm resources, while Tsai and Ghoshal (1998) extend this framing to contend that product innovation can be a good indicator of value creation. Argote and Ingram (2000) focus on the creation and transfer of knowledge, and the existence of people-embodied knowledge as the source of value creation and competitive advantage for the firm. This is also consistent with the concept of two-sided markets and how different user groups create value not just for themselves, but also for other stakeholders in the cycle. For instance, Eisenmann et al. (2006) provide the example of how video game developers (i.e. one user group) would only create video games for platforms that have a large share of game players (i.e. second use group), because these game developers need a critical mass of players to recover their programming costs and make profit (Eisenmann et al. 2006). In turn, and on the other side, game players favor platforms with a lot of high quality games. Due to these synergistic effects on value proposition across the different stakeholders, platforms like the SNS are able to enjoy increasing returns over time.

There are two key learnings from these previous studies on value creation that I have adapted to my research on large online social networks. First, I obtained a generalized and usable definition of what it means to create value across different levels of analysis and across contexts. Second, I developed an insight that value is not restricted to a specific functional unit within an organization, but rather emerges as a result of the interplay among several units and entities, consistent with the idea of a two-sided market (Eisenmann et al. 2006; Peppard et al. 2000; Tsai and Ghoshal 1998).

I draw on the above studies to now discuss what it means for users as well as the SNS providers to create value on the SNS. Specifically, I focus on the role of content generation, network formation, and platform features as key elements that drive value creation for the three SNS stakeholders - the individual, the SNS provider and external businesses (e.g. brands and digital marketers). This inquiry into understanding value creation on the SNS is not just motivated by the need to develop theoretical insights in this area, but also has timely business application. As mentioned earlier, recent media reports suggest that users and firms face a lot of uncertainty when it comes to understanding whether and how value is created for them on the SNS (Dekel 2013; Wilson 2015). I summarize the value propositions for the key players below.

1. SNS Users: The registered users on any SNS platform form the lifeline of the SNS, and engage with the platform by (i) producing and consuming content on the site, (ii) forming new and severing existing[5] social ties, and (iii) interacting with platform features like privacy controls, brand communities, friend recommendation systems, search engines etc. There have been some prior work on understanding the value users derive from participating on the SNS. These mostly include a sense of belongingness and subjective well-being, access to information and entertainment, social support, and self-presentation (boyd and Ellison 2007; Hogan 2010; Marwick and Boyd 2011; Nadkarni and Hofmann 2012; Park et al. 2009; Raacke and Bonds-Raacke 2008; Wilson et al. 2012).
2. SNS Provider: While users derive a number of useful gratifications for themselves from SNS, their value proposition also directly contributes to value creation for the SNS providers like Facebook, Twitter, LinkedIn etc. Two direct sources of value for SNSs are (i) market valuation of the company that stems primarily from the amount of user engagement on the platform, and (ii) revenue generated for the SNS provider from paid services offered to external businesses (e.g. paid social ads, analytics for brand pages on the SNS etc.). However, while SNSs and digital markets recognize the value-generating potential of SNSs, there is still no consensus at quantifying how much value is generated from these different modes of engagement, and how the value generated might vary in reaction to specific feature changes on the platform (Barry 2012; Creamer 2012; Ray 2010; Syncapse 2010; Terlep et al. 2012).

[5] Although the phenomenon of severing ties is very rare on most major SNSs like Facebook and LinkedIn

3. External Businesses (e.g. Advertising brands and digital marketers): Similar to SNS providers, advertising brands and digital marketers also capitalize on increased user engagement on the platform in the form of higher content generation and increasing number of social ties. More content implies better personalization for targeted advertising, while more social ties implies better design of social ads, and greater chances of word-of-mouth based product diffusion on the platform. Further, SNS providers create value for external businesses by developing platform features that allow brands to reach current and potential customers on the SNS with ease. For instance, brand pages on Facebook and Instagram offer customers a chance to stay informed about latest information and promotions from the brand, while also allowing them a chance to self-present to their SNS audience by signaling their endorsement of the brand (e.g. by "liking" their posts etc.).

In the following section, I introduce the research context for this dissertation, and provide an overview of the three essays that explore how value is created for the advertising brands, SNS users and the SNS provider respectively.

1.3 Research Context

This dissertation comprises three empirical studies that were conducted using a set of large-scale real-world datasets from a major SNS as well as an Asian fashion retailer. The details of the data sources have been masked from this publication on account of non-disclosure agreements. The following are summaries of the three research studies that have been illustrated in later chapters.

In my first study, I investigate the economic value created for brands when individuals join brand pages on a large SNS. Drawing on theories of self-presentation (Goffman 1959a; Hogan 2010), I show that, by identifying and measuring self-presentational motives behind participating on the SNS (Kim et al. 2012; Ma and Agarwal 2007), we can better explain economic outcomes such as offline product purchases. The specific research questions I seek to answer in this study are the following: (*i) Do users spend more or less offline, on average, upon joining SNS brand pages?, (ii) How can we categorize self-presenting users on the SNS?*, and finally, *(iii) Do users who self-present more than others on SNS spend more or less offline upon joining the brand page?* To answer these questions, I leverage two real-world datasets and use a collection of text-mining based methods to construct the focal variables, which are then used in a set of econometric models to explain purchases made by the user. Through my analyses, I observe a group of users who show significantly high information divergence across public vs. private channels on social media i.e. users who show a large disparity in valence and content of information between their public and private channels. I hypothesize that these type of users might be those who are driven by self-presentational motives to participate on the SNS and engage with SNS features like brand pages. By matching this SNS usage data for self-presenting users with their offline purchase data collected from an offline fashion apparel retailer, I was able to make inferences on the effect of joining brand pages for these group of users. The results from this study show that while individuals are likely driven by utilitarian motivations (e.g. promotions or discounts) for joining SNS brand pages (Muntinga et al. 2015a; Rishika et al. 2013) and, on average, spend less upon joining the brand page, this decrease is significantly attenuated for individuals who are more self-presenting than others. Thus, as a potential theoretical contribution, the findings also shed light on the offline economic impacts of online self-presentation which has not been investigated in prior work on self-presentation or impression management. In addition, the findings are also invaluable to brand owners and social media marketers who are interested in understanding the value of SNS features, and how the SNS users interact with these features to generate value for not just themselves (e.g. signaling brand endorsements etc.), but also for the brand. Recent industry reports on the profitability of such brand pages show that there currently exists no consensus on the usefulness of such platform features (Barry 2012; Ray 2010; Syncapse 2010). Thus, the current study can potentially contribute a timely solution to this problem.

In my second study, I analyze the role of the users' network structure in influencing how they create value by generating content on the SNS. Users on the SNS create value for themselves through broadcast posts (e.g. status updates on Facebook), as well as private conversations with their friends on the platform. Sustaining a high level of discourse and an active community of networked peers on the platform is critical to building a sense of subjective well-being for the users, who also derive informational, entertainment and self-presentation related benefits from participating on the platform. Thus, with rapid growth of SNSs, it has now become an imperative for platform owners and marketers to investigate these social factors that drive content production and friendship formation on their platforms. I contend that the content producing behavior of users on the SNS is influenced not just by their personal attributes like age and gender, but also by their social network structure. However, quantifying the effect of network structure on content generation is a challenging empirical exercise since network formation in the real world is an endogenous process. This is partly because of the co-dependency of the social network structure and the content production behavior on each other, an example of Manski’s "reflection problem" in social networks (Manski 1993). However, empirically separating the effects of network structure and content generation on each other is essential to accurately quantifying the role of each in generating value on the SNS. To address this question, I develop a stochastic and dynamic structural model for the co-evolution of online network formation and content generation. The model builds upon and extends prior work in this area by Snijders et al. (2007) and Steglich et al. (2010) in several key ways and is applied to a unique panel dataset obtained through collaboration with one of the largest online social networks in the world. Based on the study results, I offer two significant theoretical contributions to the literature on peer effects in social networks. First, I find that both homophily as well as heterophily, based on observable behavior, can exist at different stages of network evolution. More specifically, the results show that while individuals befriend others who are similar in content production (i.e. homophily) during the friendship formation stage, they gradually diverge in their content production behavior from these similar others over time (i.e. heterophily). Second, I find evidence that both homophily and influence based on public posting behavior is dependent on the current state of the posting behavior i.e. high and low posters show varying susceptibility to both homophily and influence. Such behavioral dependence of peer effects has not been reported in prior work. The results from this study not only offer a statistically disciplined approach to modeling the co-evolution of network structure and posting behavior, but can also be used to generate actionable recommendations for SNS owners, social media marketers and advertisers. For example, the results can inform and guide the design of better and more adaptable friendship recommendation engines for SNS platforms, while also informing social media marketers on how to effectively seed marketing information, and target valuable users on the SNS.

For my third study, I leverage the co-evolution model, as introduced in the previous study, as well as a quasi-experimental empirical design to analyze the impact of a major privacy-related feature change by a large SNS, on the users' content generation behavior – a key driver of value creation on the SNS. While SNSs have introduced a number of features over the previous years to provide individuals with a higher level of control over their content and their audience, it is still unclear how such privacy controls affect the individual's self-disclosure patterns and friendship formations. However, this is an important question to answer as such interventions can potentially increase or decrease value for the SNS users, via better engagement and more personalized ads, as well as the platform owners, by altering the amount of personalized content produced on their platform. The findings from the study show that while SNS users do not show any significant change in the volume of public posts generated on the platform, they significantly reduce the volume of private messages exchanged, in the week following the privacy interventions. I also perform a quasi-experimental analysis to show that this effect is less pronounced for users who had very high privacy consciousness prior to the intervention, as compared to others users. I put forward some possible explanations for the findings and highlight the potential implications. I contend that the findings from this study would help us better understand how the value generation on SNS is affected in response to privacy-related feature changes which are becoming increasingly popular. Moreover, platform owners would also benefit from an understanding of whether their privacy related interventions have the desired outcome.

1.4 Potential Contributions

Through this dissertation I hope to contribute towards understanding how value is created on large SNSs by the processes of content generation, network formation as well as engagement with platform features. This value creation is directed at not only the individual user herself, but also toward the SNS providers and the external businesses, like advertisers and marketers. The following are the key contributions of this dissertation.

First, this dissertation is among the first to perform a set of large-scale investigations into the role of, and interplay among, content generation, friendship formation and platform features (e.g. brand communities and privacy controls) using real-world data from a major worldwide SNS. Further, I improve on prior work on the topic by integrating user contribution data across public as well as private channels, as well as with third party data providers. The inclusion of multi-channel data provides us with a more accurate description of the user behavior on the platform.

Second, the essays in this dissertation contribute to our understanding to how users on the SNS create value for three stakeholders viz. the user herself, the SNS provider, and external businesses (e.g. advertisers). For instance, the first study in this dissertation offers prescriptive suggestions to brands on how to effectively manage their presence on the SNS using brand pages, and on targeting valuable users on these pages. Similarly, the third study also offers revealing insights to SNS providers on the impact of the introduction of new privacy controls on the platform, on the users' content generation as well as network formation. I contend that such findings are instrumental to businesses in understanding how user participation on the SNS, and with specific SNS features, can increase or decrease their economic value propositions.

Third, this dissertation sheds light on the importance of recognizing and factoring in the role of the network that users are embedded in. I contend that the role of network structure in influencing user behaviors on the SNS (e.g. content generation) has been largely under-studied in previous work. This dissertation offers methodical contributions in analyzing behavioral outcomes on the SNS, after controlling for any change in underlying network which often acts as a potential confound. Specifically, in Chapter 3, I quantify the effect of network structure on content generation while controlling for the fact that content generation might affect network formation at the same time. Similarly, in Chapter 4, I try to correct for the confounding network change while estimating the effect of privacy control interventions on content generation on the platform.

Fourth and lastly, this dissertation spawns several novel empirical questions about value creation on the SNS that I hope to address in future work. Specifically, in later work, I will investigate the role of emerging SNS features (e.g. Facebook reactions) and policies (e.g. new Facebook group policies) on content generation, friendship formation and other modes of engagement such as joining brand pages, joining non-brand pages, participation in political discussion groups etc. Further, I hope to also learn more about how the network structure within such communities on the SNS, as well as outside these groups, affect such behavioral outcomes for the users. A number of potential future directions have been illustrated in Chapter 5.

Chapter 2. IMPACT OF ONLINE SELF PRESENTATION AT THE SNS ON OFFLINE PURCHASE BEHAVIOR

2.1 Introduction

With the emergence and growth of social media platforms, brands have become increasingly conscious of the business value generated from these channels through the vast amounts of user generated content (UGC) created every second. Previous research has emphasized that UGC is often invaluable for product targeting, personalized advertising and other brand marketing purposes (Archak et al. 2011; Chevalier and Mayzlin 2006; Liu 2006). Moreover, company-sponsored communities on these social media platforms have evolved into important venues for bidirectional communication between brands and consumers (K. Goh et al. 2013; Kumar et al. 2013; Lee et al. 2014). For instance, brand pages on Facebook and product pages on Instagram feature brand-sponsored profiles and allow fans and potential customers to learn about, and interact with the brand. Users engage with brands as well as with other users through these brand pages on Social Network Sites (SNS) to seek and obtain various benefits such as product-related information, product promotions and to also signal their endorsement of the brand to their SNS audience (Kalehoff 2013; May 2011; Muntinga et al. 2015b; Rishika et al. 2013). While there has been a recent flurry of recent industry reports highlighting the need for social media marketing (Angelova 2013; Gerber 2014) and questioning the offline value of such online engagements (Phang et al. 2014; Rampell 2010), a number of major concerns have surfaced.

First, while early studies show the potential of user generated content

(UGC) in predicting important social and economic outcomes (Chevalier and Mayzlin 2006; Godes and Mayzlin 2004; Luo et al. 2012; Metaxas and Mustafaraj 2012), recent studies highlight the various perils of using public online data, such as increased prediction errors and low explanatory power (Lazer et al. 2014; Wong et al. 2012). While previous studies have focused on the impact of UGC like product reviews, social media posts, blog articles, etc. on the behavior of consumers, the current study investigates how such content can be analyzed to make inferences about the content producer, and her offline behavior. Such insights can be valuable to marketers who want to target users on social media sites such as Twitter, Facebook, Instagram, Google+ and others, as well as other online platforms like user review sites, community forums, and social mobile applications.

Second, while previous research has predominantly focused on explaining marketing outcomes such as product sales using UGC measures, few have focused on the importance of specific motivations which guide content generation and social media engagement, e.g., self-presentation behavior (Bughin 2007; Toubia and Stephen 2013). In social psychology, the study of self-presentation has been gaining importance since Goffman's seminal work on the dramaturgical approach, where he argues that individuals engage in performances, "which occurs during a period marked by his continuous presence before a particular set of observers and which has some influence on the observers" (Goffman 1959b). Through such performances, individuals strategically manage their impressions in public to reap social benefits. However, while self-presentational behavior on social media communities is a popular and well-studied motivation for engagement (boyd and Ellison 2007; Hogan 2010; Toubia and Stephen 2013), little is known about its economic implications in the offline setting. However, there has been recent work that hint at the economic importance of self-presentation or image-seeking aspirations on social media (Muntinga et al. 2015b; Rishika et al. 2013). We contend that this is especially true for contexts that lends itself to self-presentational motives e.g. fashion products(Bellezza et al. 2014). In other words, it is not clear whether individuals who self-present more on online social media platforms would have distinct economic behaviors in certain offline settings, as compared to those who self-present less.

Third, while recent reports suggest that only a small fraction of users engage with a brand online after joining its online brand page (Creamer 2012; Ray 2010), it is unclear how the online engagement of these users translates to offline revenue, the eventual value proposition that brand owners aspire for. There have been some recent attempts to quantify the impact of social media engagement (K. Y. Goh et al. 2013; Kumar et al. 2013; Rishika et al. 2013). I depart from these existing studies by investigating user’s propensity to join brand pages on the SNS, for specific types of users (e.g. high self-presenting users). Secondly, I account for user-level heterogeneity by using their participation data from the entire SNS, and not just from inside the social media brand page.

In the current study, I investigate the relationship between the users’ engagement in brand pages and their offline purchase behavior from a fashion retail store. While Goh et al. (2013) and Rishika et al. (2013) studied the effects of engagement within brand communities (e.g., Facebook fan pages) on offline sales, I seek to investigate the impact of joining the brand page on offline sales. Specifically, I categorize social media users based on their specific motivation to participate in the SNS, and show that such motivations impact offline sales in distinct ways. By analyzing backend user-level data of public posts and private conversations from a large and popular SNS, as well as the user's offline transaction data from the loyalty card database of a popular fashion retailer, I seek to understand how high and low value buyers behave on the SNS and brand pages. In addition, I incorporate text-mining techniques to further study potential motivations for joining the brand page. For example, Muntinga et al. (2015) and Rishika et al. (2013) point out that, while some users join brand pages for utilitarian reasons (e.g., promotions and discounts), others might join with self-presentational motives such as brand signaling and endorsements (Kalehoff 2013; May 2011; Muntinga et al. 2015b). Drawing on these insights, I propose text-mining methods to identify and separate users who are more *self-presenting* on the SNS than others, and investigate whether such users have a distinct purchasing behavior offline. Specifically, I ask the following research questions: (i) Do users spend more (or less) offline upon joining an online SNS brand page? (ii) How can we identify self-presenting users on social media platforms using computational approaches?, and finally, (iii) Upon joining the brand page, do users who self-present more than others on SNS spend more or less offline than users who self-present less?

My results provide empirical evidence that individuals on social media brand pages have different trajectories of offline purchase behavior depending on their motivations to join the brand page. I find that individuals, on average, spend \$4.72[6] less per month offline upon joining the online brand page. However, this impact is moderated by the individual's extent of self-presentation on the SNS, such that customers who are more self-presenting spend \$3.85[7] per month more than customers who are less self-presenting. For the analyses, I propose a new measure of self-presentation as the difference in sentiment between one's public posts and private conversations on the overall SNS as well as within the brand pages. This sentiment divergence measure in my study captures the difference in emotional valence expressed by the user in public versus private channels[8] at a given point in time. I contend that individuals who display a higher value of sentiment divergence are more disposed towards self-presenting as compared to other individuals. For example, high divergence users might be very unhappy in private channels, but choose to present a more positive image in public channels, or vice-versa. The notion that individuals tend to strategically manipulate their behavior depending on the specific audience has been well documented in recent research on self-presentation in both online as well as offline contexts (Hogan 2010; Leary and Kowalski 1990; Marwick and Boyd 2011). Further, more recent work have shed light on how individuals strategically alter or fake emotions to reap benefits in social interactions (Andrade and Ho 2009a).

My results not only provide evidence on how and when brand pages are useful for marketers to drive better Online to Offline (O2O) conversions, but also offer a new approach of exploiting UGC on social media to uncover

[6] Nearly 28.49% of the average monthly expenditure per quantity sold in my data sample
[7] Nearly 23.23% of the average monthly expenditure per quantity sold in my data sample
[8] Public channels on SNS include broadcast methods of posting and sharing content publicly with all friends and followers. In contrast, private channels only allow for peer to peer conversations, such as private messages on Facebook or direct messages on LinkedIn.

individual-level differences among the content producers on these platforms. Through this research, I propose a set of new NLP-based metrics, e.g., sentiment divergence, to capture individual psychographics, and thus offer a cheap and efficient way of targeting relevant customers on the SNS. I contend that while recent studies and media reports have highlighted the importance of SNS brand pages as an emerging marketing medium, this paper is among the first to help quantify the economic value of joining a brand page on the SNS for different user groups. Further, prior research has highlighted the perils of using solely public information in making predictions. However, I also show the value of incorporating data on private communication among users[9], which is less likely to be influenced by some of the biases inherent with public data (e.g., herding effect, conformity, etc.) and hence, offers an important source of information about user attitudes and affect (Leary and Kowalski 1990). My NLP-based sentiment divergence measure captures this public-private dichotomy and draws upon self-presentation theories and self-disclosure tactics that describe how humans display divergent behavior in front of different audience groups (Goffman 1959b; Leary 1996)., the economics, information systems, and marketing research literature have been silent on the implications of such theories in their own contexts. Through this work, I show that such theories have an important role to play in the context of social media marketing and big data. For example, I find that individuals who display a higher sentiment divergence across the public and private channels spend more on average than others, upon joining brand pages on the SNS. I also conjecture that this sentiment divergence might be indicative of

[9] I suitably anonymize and protect such user data in the course of our analysis to preserve confidentiality and privacy

self-presentation behavior and thus, also emphasize the economic value of self-presentation on social media, a topic that has not been looked at in the extant literature on self-presentation. This association between self-presentation and heightened economic value is not just theoretically interesting, but is also valuable for online marketers and SNS platform owners who are interested in devising a more effective targeting strategy for members in online brand communities. Moreover, I am mindful that, with the exception of platform owners and a few prominent marketing agencies, it might not be tenable for other firms to get access to private communication data of their social media users. Hence, in addition to the sentiment divergence metric as described above, I offer alternate measures (e.g. egocentricity scores) that can be computed from public data alone, and offers a robust alternative to using sentiment divergence on the SNS.

In the next section, I review the relevant literature on the emerging role of brand communities, including brand pages, on social media, and some recent investigations on user-generated content on online platforms. I also revisit some popular studies that discuss self-presentation on offline and online platforms, as well as the incidence and importance of social biases in group activities.

2.2 Background and Related Work

In this section, I review past research from three related streams of literature. In the first section, I review recent work on how brands leverage brand communities on the SNS to generate value for themselves. In the second section, I look at how previous studies across various disciplines like information systems, computer science, marketing etc. have leveraged UGC on social media for its predictive value in a multitude of contexts. In the third section, I review studies from the social psychology and related fields that discuss the psychological theories surrounding self-presentation, both in online as well as offline contexts. Finally, in the fourth section, I discuss previous work that illustrates group-related biases associated with public contributions. Through the review of these four streams of literature, I intend to build a case for how UGC can be better exploited to identify self-presentational behavior on the SNS, which in turn would help us better explain specific outcomes related to the brand (e.g. offline purchases).

2.2.1 Role of marketer-driven brand communities on social media

The stark increase in SNS usage among young and older adults in recent years have attracted the attention of brand marketers who are now looking at social media as a fast-emerging channel to engage with their existing customers, and also to attract potential high-value customers (Moorman 2013; Phang et al. 2014; Rampell 2010). There are two major modes of engagement on social media platforms that brand owners typically use, i.e., paid social advertisements and brand communities. Recent reports suggest that more brands are increasingly focusing on the latter to increase their interactions with current and potential customers (Terlep et al. 2012). Even though a number of recent studies have focused on user engagement on these brand communities (Lee et al. 2014; Oestreicher-Singer and Zalmanson 2012), few studies have made connections between such engagement and individual-level sales. While customer engagements on social media platforms are easier to record and quantify, it is much harder to attribute customer value to that engagement – a major problem facing O2O commerce today. Furthermore, little is known about the value of a "fan". For instance, the value of a fan on a Facebook brand page was reported to be anything between \$0 (Ray 2010) to \$22.93 (Syncapse 2010), to as high as \$214.81 (Barry 2012). The lack of consensus in estimating the true “return-on-engagement” of social media brand communities stems mainly from the lack of availability of objective data about the brand community members and their offline purchase records.

There have been a number of studies in the recent past that have used individual-level data from social media brand communities to answer related questions. For instance, researchers have investigated the relative influence of user-generated and marketer-generated content in influencing sales (K. Y. Goh et al. 2013). Studies have also analyzed how specific characteristics of user- and marketer-generated content on these brand pages impact engagement levels (e.g., the probability of a "like", or a "share" etc.) (Lee et al. 2014; de Vries et al. 2012). A number of important studies have departed from the content-perspective and analyzed the "social" nature of these brand communities. Specifically, these studies have explored the role of the community members in influencing engagement levels and product evaluations on the brand page (Laroche et al. 2013; Naylor et al. 2012). Lastly, a number of studies have tried to address the question of quantifying return-on-investment on social media marketing. More importantly, these studies provide us with metrics on how to estimate a user's value on these platforms, in terms of the revenue generated, the associated word-of-mouth, etc. (Kumar et al. 2013; Rishika et al. 2013). However, few have looked at individual-level engagement both on the brand pages as well as throughout the SNS, outside of the brand page. In the current study, I leverage user data from not just the brand page, but also from other user and non-user pages. The objective in this study is to create of a profile of the user based on their self-presentational tendencies, and not brand-specific self-presentation alone.

Thus, obtaining user-generated content from throughout the SNS, including but not limited to the brand pages helps us establish a more accurate and complete profile of the social media users, which I then use in specifying my econometric models.

2.2.2 Public user-generated content and inconsistent value predictions

The emergence of online communication and, in particular, on social media has dramatically increased online engagement and word-of-mouth (WOM), or UGC on online platforms (Dellarocas 2003). These WOM interactions have been used to predict movie and television success (Asur and Huberman 2010; Chintagunta et al. 2010; Godes and Mayzlin 2004; Rui and Whinston 2011), election outcomes (Metaxas and Mustafaraj 2012), product sales (Chevalier and Mayzlin 2006; Ghose and Ipeirotis 2011; K. Y. Goh et al. 2013) and even firm equity values (Luo et al. 2012). The WOM information comprises mainly of crowd-contributed reviews, ratings and social network chatter, which have been shown to display high predictive power. While earlier studies on WOM have focused on the quantitative aspects of user-generated content (e.g., volume, valence, ratings, etc.), more recent studies show that qualitative characteristics of the content (e.g., sentiment, readability, subjectivity, etc.) have better predictive power (Ghose and Ipeirotis 2011; Ghose et al. 2012; K. Y. Goh et al. 2013; Z. Zhang et al. 2012).

On the other hand, a number of existing studies have uncovered the limitations of using social media data in predicting various offline outcomes. Among the notable ones, Wong et al. (2012) report that Twitter data should be used with caution and that Twitter users differ significantly from non-Twitter users in terms of their relative preferences, for movies in their case. This highlights the importance of considering self-selection on SNS. The same study shows that a large volume of tweets does not necessarily predict box-office success as over half of them were found to be uninformative. In addition, while Asur and Huberman (2010) found the sentiment of tweets to be a strong predictor of movie success, other studies find that volume, and not valence, plays a vital role, implying thereby that any publicity is good publicity (Wasow et al. 2010). Such contradictions have surfaced in other contexts as well. For example, social media data was unable to predict pre-electoral polls in the US (O'Connor et al. 2010) and that the valence of consumer review text on product sites have been shown to have no correlations with actual sales (Liu 2006; Liu et al. 2010).

While there are potentially many reasons why publicly generated content on social media may not be a good predictor of different focal outcomes of interest, I focus my attention on two motivations for users to generate and post UGC publicly. First, content contributed online might be a result of an intrinsic motivation to generate content or reflective of a heightened self-presentation behavior (Bughin 2007; Toubia and Stephen 2013). Second, content that is public in nature is subject to group related influences. Individuals may have a tendency of modifying self-expressions to suit specific social contexts (Jones and Wortman 1973; Schlenker 1980). Thus, there exists a significant and positive correlation between the actions of the individual and the actions of group members and peers. I further discuss related work on these topics in the next two sections.

2.2.3 Self-presentation and content generation

Self-presentation theories highlight audience segregation as a key source of variation in the way we self-present (Goffman 1959b; Leary 1996). Similarly, Rosenberg's Evaluation Apprehension Theory (Rosenberg 1965) illustrate that humans behave differently when they perceive that they are being evaluated by others. This role of spectators or audiences in shaping an individual's self-expression is one of the common findings in impression management research (R. Baumeister et al. 1989; Jones and Wortman 1973; Schlenker 1980). Since public channels in online ecosystems are under perpetual observation and evaluation, users might significantly alter their content production behavior to gain social acceptance and positive evaluation (Hogan 2010; Marwick and Boyd 2011). This assertion is further supported by seminal studies in the area of social influence and social conformity (Asch 1951; Cialdini and Goldstein 2004).

Furthermore, a key value offered by social media is of enabling users to efficiently broadcast their information to a wide audience (Rui and Whinston 2011; Toubia and Stephen 2013). As a result, it may attract more self-presenting users. It is, therefore, not surprising that the need to self-present on social media sites has emerged as an important gratification sought by social media users (boyd 2004; Donath and boyd 2004). Social media websites are often designed to specifically encourage self-presentation related behaviors, as in the case of online dating sites and photos sharing sites. However, some social media sites also focus on more utilitarian purposes such as using health sites, reading blogs (Carpenter 2012; Toubia and Stephen 2013), and joining brand-sponsored pages for product discounts. While prior studies suggest that public content such as reviews, blog posts and newsgroup conversations may have value to content consumers (Chevalier and Mayzlin 2006; Dhar and Chang 2009; Godes and Mayzlin 2004), little is known about how utilitarian and self-presentation related motivations to participate on the SNS influence the offline purchase behavior of the content producers. Furthermore, individuals often join brand communities such as brand pages on SNS with self-presentation aspirations (Kalehoff 2013; May 2011). This is particularly true for some conspicuous domains such as fashion and apparel sectors, where users are driven by self-presentation related desires that are often reflected in their purchasing behaviors (Slama and Celuch 1995; Slama et al. 1999).

While the prevalence of self-presentation related behavior on social media is fairly well understood, a key problem that arises in analyzing data from such social media sites is in differentiating public content that is purely driven by intrinsic motivation versus that which is also driven by self-presentational motives (Toubia and Stephen 2013). While differentiating between these two competing motivations may be less critical from the content-consumer’s perspective, it is crucial to making accurate predictions and for user targeting. For instance, a person who writes a positive movie review purely for publicity-seeking purposes might not have liked the movie or even seen it. However, if the same person writes the same review but driven, instead, by purely intrinsic motivations, then the review can be used to better predict his movie viewing history as well as the quality of the movie and its overall success. Hence, the UGC in such cases can be effective for future targeting of marketing campaigns as well as for predicting sales.

2.2.4 Group-related biases in publicly generated user content

While self-presentational concerns have a significant influence on the semantics and sentiment of UGC, a second important factor is the presence of group-related biases such as herding. In addition to the content producer's intrinsic motivation to produce UGC, users' public actions are likely to be strongly correlated with the behavior of the group (Hyman 1942) due to herding behavior (Banerjee 1992; Bikhchandani et al. 1992; Shiller 1995). For example, Chen et al. (2011) establish the importance of observational learning in a group by showing that positive observational learning boosts product sales while negative observational learning has no similar effect. They suggest that the rationale is related to self-presentation motives. Similarly, through a series of experiments probing individual decision making in presence of a group, Asch (1951) uncovered strong evidences of conformity induced biases where the individual chose to follow the group decision even when the decision was clearly incorrect. These studies suggest that public content is rife with repetitive information which might not necessarily be indicative of the content producer's own preferences nor the product's true attributes, but, rather, can be merely reflective of the overall group behavior.

In the online context and on SNS, these studies suggest that a user's public and private content would be very different from each other owing to the presence of self-presentation and social context-specific concerns. In the absence of any such concerns, intrinsic motivations would dictate content generation and the public and private contents would be increasingly similar. However, if the user is more self-presenting, he may actively try to segregate content production based on his different audience groups (e.g., family and friends versus acquaintances) and this might lead to a higher divergence in the public and private content. Since the sentiment of content is an important qualitative attribute of the information, we are likely to observe divergent public and private sentiments due to the different groups of audiences in the two channels. This paper focuses on this self-presentation behavior that drives users to participate in online communities by posting publicly and joining brand communities. I suggest that using a combination of public and private data, we can quantify such self-presentation behaviors on SNS, and thus offer insights to help improve behavioral targeting of consumers in certain conspicuous domains such as the fashion and apparel industry.

2.3 Data Context

I obtained purchase data of 2,301 customers from a loyalty program database of a popular brick-and-mortar fashion apparel retailer in an Asian country[10]. The retailer has over 20 stores in the country selling men's, women's, and children's casual-wear clothes, moderately priced, equivalently, between 8 and 77 USD. Customers are *automatically* enrolled into the loyalty program after spending a threshold amount of 60 USD on a single order. The loyalty program provides price discount and money-back incentives to use the loyalty member card, as well as birthday promotions and other services. The retailer also hosts a brand page, on a popular SNS, where they disseminate marketing messages and promotions to the members of the brand page. Through collaboration with the SNS, I also obtained backend data for the loyalty

[10] Due to a non-disclosure agreement, I cannot reveal the identity of the retailer, its market of operations, and the name of its brand page on the popular SNS. The SNS similarly imposed such data confidentiality requirements for this research.

program members, of their social media activities, including public postings as well as user-to-user private conversations. This resulted in over 240 million pieces of textual content for about one year of SNS activity and offline purchases from November, 2010 to November, 2011. Figures 2-1, 2-2 and 2-3 show the distributions of average monthly expenditure, average number of items purchased per month, average public sentiment of SNS content, average private sentiment of SNS content, average public volume of SNS content, and average private volume of SNS content for my data sample.

Figure 2-1: Average expenditure and quantities purchased by brand page members

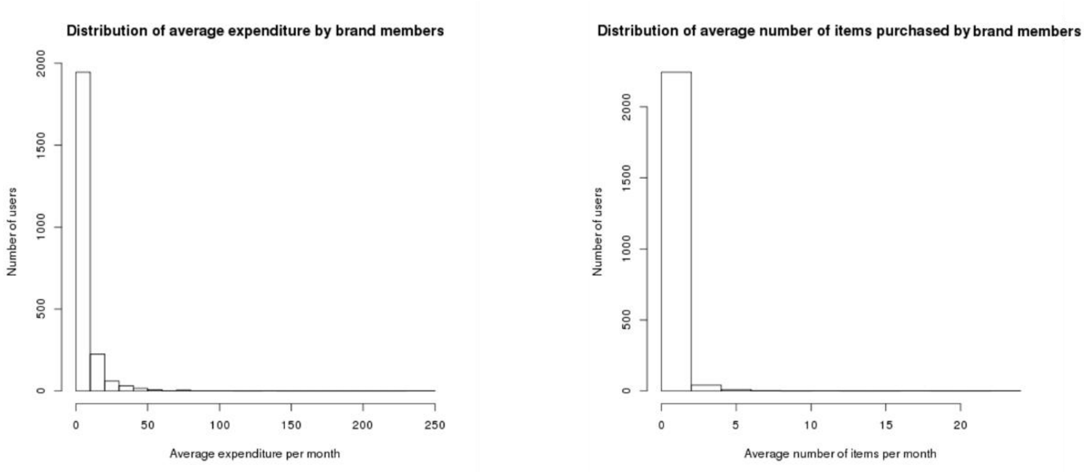


Figure 2-2: Average sentiment of public and private content on the SNS

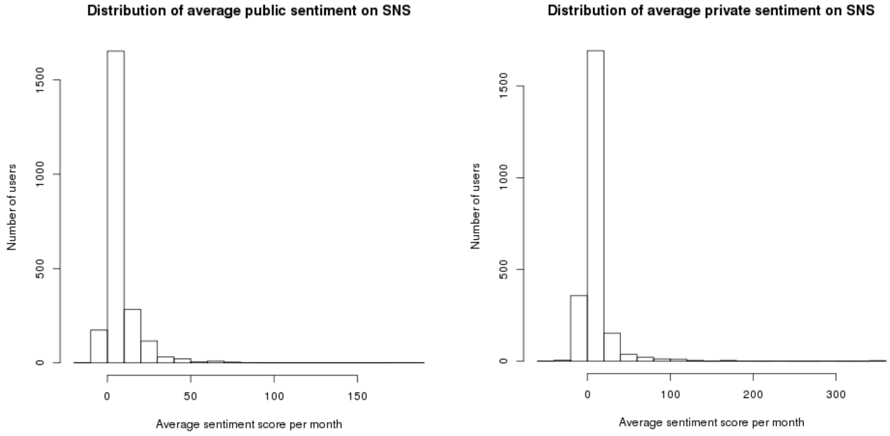


Figure 2-3: Average number of public and private posts on the SNS

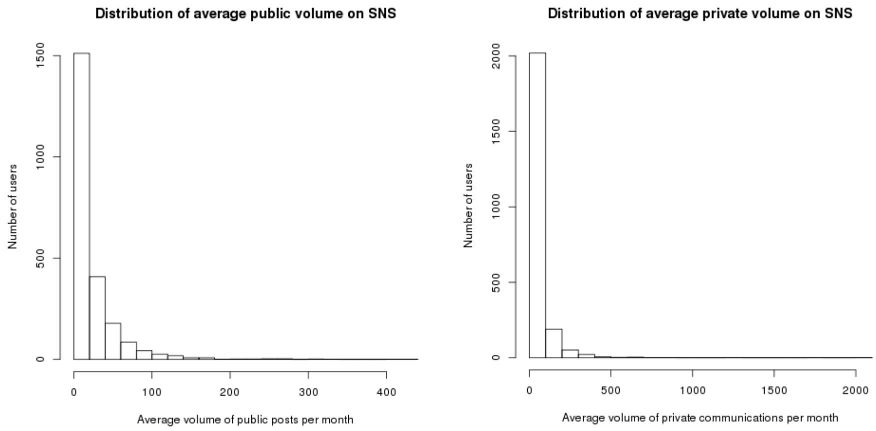


Table 2-1 provides a descriptive summary of the key variables used in this study. I define $Expend_{it}$ as the total expenditure (net of any price discounts, measured in the local currency) made by consumer *i* in month *t*. $PubSent_{it}$ and $PriSent_{it}$ denote the public and private sentiment scores for user *i* in month *t* for public content made on the user's profile and private content through one-to-one and one-to-many (e.g., group chat) messaging, respectively. The procedure to compute these sentiment metrics is explained in the next section. A sentiment score of 0 denotes a neutral sentiment. Similarly, $PubVol_{it}$ and $PriVol_{it}$ denote the total volume of social media contributions for user *i* in month *t* for public and private content, respectively. The $BrandPageJoin_{it}$ dummy is the key variable of interest in this section and provides us with an understanding of what happens when individuals join brand pages. $BrandPageJoin_{it}$ takes a value of 1 if the user *i* has joined the brand page in month *t*, and 0, otherwise. Please also note that if the $BrandPageJoin_{it}$ assumes a value of 1 in month t, it would keep retain a value of 1 in months t+1, t+2..etc.

The models also include a number of control variables to account for individual heterogeneity. Age_i and $SNSAge_i$ control for the biological age, in years, of the user i as well as her tenure on the social media platform in number of days since she registered, respectively. Similarly, I define $LoyaltyAge_{it}$ as the amount of time, in months, spent by the customer in the loyalty program of the retail store. The $NumFriends_{it}$ variable controls for the number of friends, or "degrees," of user i in month t. The $Promo_t$ variable captures the total number of offline marketing promotions[11] being run by the brand in month t, and controls for the monthly variations in the brand's promotional activity. I provide a table of the pair-wise Spearman's correlations among the model variables in Table 2-2, and find no evidence of multicollinearity in the data.

The next section leverages the above mentioned variables and proposes a set of econometric models to estimate the impact of joining brand pages and to uncover any moderating effects of self-presentation.

[11] News about most of these marketing promotions are communicated to the brand-page users through the SNS brand page

Table 2-1: Descriptive statistics of model variables

	Mean	Std. Dev.	Min	Max
Number of Users (Matched Content Producers and Loyalty Card Customers): 2301 **Period of observation:** Nov, 2010 - Nov , 2011				
Dependent Variables:				
Total Monthly Expenditure ($) *($Expend_{it}$)*	4.391	26.243	0.000	1105.360
Total Monthly Sales *($Quantity_{it}$)*	0.265	1.631	0.000	70.000
Independent Variables:				
Total Public Sentiment Score *($PubSent_{it}$)*	8.407	17.795	-98.000	398.000
Total Private Sentiment Score *($PriSent_{it}$)*	10.364	62.303	-622.000	3776.000
Volume of public UGC *($PubVol_{it}$)*	28.007	43.949	1.000	952.000
Volume of private UGC *($PriVol_{it}$)*	60.553	169.577	1.000	6524.000
Mean Divergence Score *($MeanDiv_{it}$)*	0.438	0.683	0.000	9.000
Standard Deviation of Divergence Score *($STDDiv_{it}$)*	0.366	0.572	0.000	8.000
Egocentricity Score *($Egocentricity_{it}$)*	0.0001	0.003	0.000	0.333
Age (years) *(Age_{i})*	31.295	8.635	13.000	106.000
SNS Tenure (days) *($SNSAge_{i}$)*	1273.647	285.482	64.000	2441.000
Store Loyalty Tenure (months) *($LoyaltyAge_{it}$)*	12.975	16.035	0.000	129.000
Degree *($NumFriends_{it}$)*	399.212	351.000	10.000	5234.000
Store Promotion *($Promo_{t}$)*	10.228	4.737	4.000	19.000

2.4 Empirical Analysis

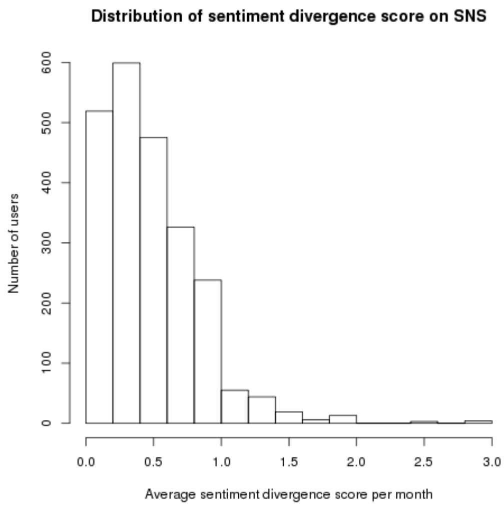
I conduct a three-part empirical study. In the first part, I investigate the impact of joining a social media brand page on the offline purchase behavior of the user, namely, the purchase expenditure. I employ a fixed-effects panel regression, as well as a time-varying propensity score estimation approach to perform the analyses. In the second part of the empirical analysis, I further investigate how self-presentation motivations can moderate the effects of joining the brand page. I devise a sentiment divergence measure to operationalize online self-presentation. To derive the sentiment scores, I favor a simple lexicon-based approach for sentiment mining (Li and Wu 2010) over more sophisticated machine-learning approaches because this algorithm can be employed in a scalable fashion. For each piece of content, I generate a set of three real-valued scores viz. positive polarity score (*psent*) and negative polarity score (*nsent*) as the proportion of positively valenced and negatively valenced words in the sentence, respectively. Next, I compute an overall polarity score (*osent*) as the difference between the *psent* and *nsent*. I then average the *osent* per user-month across public and private content due to sparsity issues to produce the $PubSent_{it}$ and $PriSent_{it}$, respectively. The sentiment divergence measure is then computed by taking the absolute measure of the difference between $PubSent_{it}$ and $PriSent_{it}$. The distribution of the sentiment divergence scores across users in my sample is shown in Figure 2-4 below. I then use this divergence measure of the user's public-private sentiment to explain monthly expenditures, while controlling for other potential relevant factors explained in the prior section.

Table 2-2: Inter-correlation matrix

	Expenditure	*PubSent*	*PriSent*	*PubVol*	*PriVol*	*MeanDiv*	*STDDiv*	*Fanpage Join*	*Age*	*SNSAge*	*LoyaltyAge*	*NumFriends*	*Promo*
Expenditure	1.000	0.031	0.015	0.001	0.004	0.050	0.030	-0.006	0.070	0.060	0.022	0.006	0.102
PubSent		1.000	0.204	0.468	0.188	0.195	0.112	0.120	0.106	0.177	0.060	0.289	0.074
PriSent			1.000	0.094	0.429	0.077	0.087	0.040	0.063	0.080	0.016	0.086	0.009
PubVol				1.000	0.259	-0.086	0.020	0.123	0.092	0.225	0.037	0.343	0.024
PriVol					1.000	-0.018	0.024	0.090	0.102	0.132	0.025	0.172	0.022
MeanDiv						1.000	0.307	0.171	0.233	0.225	0.098	0.048	0.028
STDDiv							1.000	0.172	0.212	0.229	0.095	0.074	0.021
Fanpage Join								1.000	0.448	0.552	0.304	0.279	0.069
Age									1.000	0.773	0.275	0.215	0.098
SNSAge										1.000	0.327	0.375	0.077
LoyaltyAge											1.000	0.080	0.050
NumFriends												1.000	0.046
Promo													1.000

Based on past studies which suggest that users produce content on social media sites due to intrinsic or utilitarian motives and due to “image-seeking” desires (Bughin 2007; Toubia and Stephen 2013), I conjecture that a higher (lower) sentiment divergence between public and private disclosures among users is indicative of greater (lesser) self-presentation behavior.

Figure 2-4: Distribution of sentiment divergence measure on the SNS



In the third and final part of the study, I perform a deeper investigation of whether self-presentation motivations are truly responsible for moderating the influence of joining brand pages on the offline purchase behavior of users. I apply text mining techniques from research on psycholinguistics to further characterize the self-presentation construct. First, I verify whether users who are more self-presenting, as reflected by higher levels of sentiment divergence, are also informationally any different from users who are less self-presenting across both public and private channels (Hogan 2010; Laserna et al. 2014; Marwick 2005). Thus, I conjecture that users who are highly divergent in sentiment will also differ from low divergence users on the distribution of words and topics used in their online conversations. Second, I replace the sentiment divergence metric in the regression analyses with a more direct measure of how self-presenting an individual is, by leveraging the volume of egocentric words mentioned by the user, denoted in this study as $Egocentricity_{it}$. I perform text mining on the user's public posts to measure the proportion of egocentric words (e.g., "I", "me", "our", etc.) spoken by an individual in a given time period, and used this as an alternate measure to the sentiment divergence metric, described above[12].

2.4.1 Results

I discuss the results from the empirical analysis in the sections below. In the next sub-section, I investigate the impact of joining a brand page on the offline purchase behavior of users. Following this, I look at the moderating effect of self-presentation by the users. Finally, in Section 2.5, I provide additional robustness tests to address potential confounds.

2.4.1.1 Impact of joining brand pages

In this section, I investigate the impact of joining the online brand page on the offline purchase expenditure, after controlling for time invariant and time varying covariates. I use the following panel fixed- and random- effects model specifications.

$$\begin{aligned} Expend_{it} = {} & \beta_0 + \beta_1 * PubSent_{it-1} + \beta_2 * PriSent_{it-1} + \beta_3 * PubVol_{it-1} + \beta_4 * PriVol_{it-1} + \\ & \beta_5 * BrandPageJoin_{it} + \sum_{k=6}^{10} \beta_k * Control_{it} + \alpha_i + \gamma_t + \epsilon_{it} \end{aligned} \tag{1}$$

[12] Egocentricity can be both a trait as well as a behavior. A person can be dispositionally more egocentric, in which case it is a trait. Alternately, a person could behave in a more egocentric manner under the effect of an intervention which activates his/her self-concept. Moreover, while a person is generally conscious about the semantics of what he writes, the specific linguistic constructs, like usage of pronouns is largely subliminal and happens at a subconscious level. That is why, writing styles have are increasingly being used in psychology research

where, $Control_{it} = \{Age_i, SNSAge_i, LoyaltyAge_{it}, NumFriends_{it-1}, Promo_t\}$

The estimation results are presented in Table 2-3. The panel fixed-effects model specification was preferred to the random-effects specification following the outcome of a Hausman test to compare the estimators.

Table 2-3: Explaining total expenditure after joining brand page

Predictors	**(a): Expend (FE)**	**(b): Expend (RE)**	**(c): Expend (FE)**	**(d): Expend (RE)**	**(e): Expend (FE)**
PubSent	-0.007 (0.011)	-0.003 (0.009)	-0.003 (0.011)	0.002 (0.009)	-0.013 (0.011)
PriSent	-0.002 (0.002)	-0.0004 0.002)	-0.001 (0.002)	0.0001 (0.002)	-0.0002 (0.002)
PubVol	-0.003 (0.007)	-0.005 (0.003)	-0.004 (0.007)	-0.007** (0.003)	-0.001 (0.007)
PriVol	0.00001 (0.001)	-0.0002 (0.001)	-0.0001 (0.001)	-0.0004 (0.001)	-0.0002 (0.001)
BrandPageJoin	-4.111*** (0.916)	-1.838***(0.558)	-4.719*** (1.015)	-2.193*** (0.666)	-6.620*** (1.178)
Age	-	0.034** (0.018)	-	0.035** (0.018)	-
SNSAge	-	-0.001 (0.001)	-	-0.001 (0.001)	-
LoyaltyAge	-	0.018 (0.016)	-	0.018 (0.016)	-
NumFriends	0.022*** (0.005)	-0.0003 0.001)	0.022*** (0.005)	-0.0003 (0.508)	0.012** (0.006)
Promo	0.110*** (0.029)	0.152 ***(0.032)	0.110*** (0.029)	0.152 *** (0.032)	0.111 *** (0.023)
MeanDiv			-1.846** (0.866)	-1.335* (0.858)	
STDDiv			-0.061 (0.344)	-0.114 (0.330)	
*BrandPageJoin * MeanDiv*			1.548* (0.860)	0.846 (0.851)	
Egocentricity					-200.919 *** (83.050)
*BrandPageJoin * Egocentricity*					266.491 *** (95.666)
Intercept	0.949 (1.894)	1.414 (1.228)	1.863 (1.948)	2.053 (1.288)	17.921*** (3.416)
Month Dummies	**Present**	**Present**	**Present**	**Present**	**Present**
Brand Page Dummy	**Present**	**Present**	**Present**	**Present**	**Present**
Observations:	16,935	16,935	16,935	16,935	16,935
Number of Consumers:	1,926	1,926	1,926	1,926	1,926
R-squared	0.109	0.101	0.110	0.102	0.128
Standard errors in parentheses	*** p<0.01, ** p<0.05, * p<0.1				

Note:

i. The subscript *t*-1 for the first few predictors indicates that the variable is lagged by 1 time period. This is done to avoid concerns involving reverse causality in the same time period.

ii. All measures of Standard Errors reported are robust.

iii. A post-estimation Hausman Test was performed on FE and RE estimation results to evaluate the consistency of the alternate estimator and only the FE estimator was found to be consistent.

We observe from Table 2-3 that after controlling for an individual's social media activity (e.g., volume and sentiment of public and private posts), the main effect estimate for the brand page join dummy is negative and significant. This implies that individuals on average spend \$4.72 less[13] after joining the brand page (Table 2-3(c)). Note that the sample size in the estimation drops from 2301 to 1926 users. This is likely caused by users who show no variance in the dependent variable i.e. made no purchases. Such observations are dropped during the regression process.

However, the conclusions from this analysis might be confounded in the presence of covariates that can potentially be correlated with the act of joining the brand page. For instance, certain dispositions (e.g., biological age), as well as situational attributes (e.g., current state of social media activity) could influence the page-joining self-selection pattern of individuals. To address this concern, I computed a propensity-score matching (PSM)-based treatment effects model, wherein I explicitly modeled the probability of joining a brand page as a function of the individual level covariates. More details on this robustness analysis are provided in Section 2.5.

The negative relationship between joining the brand page and the subsequent purchasing behavior can be better understood by unpacking the various motivations behind why individuals join a brand page on social media platforms. A recent study showed that 42% of users who "liked" a brand fan page on the popular social network site Facebook did so because they expected promotions and offers (Kalehoff 2013). A similar study reported that

[13] Nearly 28.49% of the average monthly expenditure per quantity sold in my data sample

52% of all fans of travel-related brand pages on social media sites like Twitter or Facebook “liked” the pages in hopes of better discounts (May 2011). The same study showed that companies are aware of this trend and around 77% of all content displayed on these brand pages pertained to promotional coupons. Therefore, it may be the case that individuals who join the brand page are better exposed to product discounts and they end up spending less for their purchases after factoring in the price discounts, than social media members who are not members of the brand page, i.e. not “fans”.

In the next set of analyses, I show that self-presentation, a leading motivation for joining and participating in social media brand pages, can offset some of the negative impacts driven by utilitarian motives.

2.4.1.2 Impact of joining brand pages for self-presenting users

While utilitarian reasons for joining a brand page is often the main purpose (Kalehoff 2013; May 2011), a second important motivation to join a retailer’s brand page is related to an individual’s self-presentational needs (Kalehoff 2013). Social media users often choose to join brand pages in order to signal brand attachment and product preferences to their friends on the platform. This is particularly pronounced for brands dealing in identity-signaling product categories like apparel and cars, as is the case with the brand in the current context. As a result, I investigate whether individuals who are more self-presenting than others alter their purchasing behavior on joining brand pages. I operationalize self-presentation using a sentiment-divergence metric, as detailed in the previous section, and explore if the influence of joining the brand page is moderated by sentiment divergence of the content producer (i.e.,

$MeanDiv$ and $STDDiv$ in the model below). I use the following panel fixed- and random- effects model specifications.

$$\begin{aligned} Expend_{it} &= \beta_0 + \beta_1 * PubSent_{it-1} + \beta_2 * PriSent_{it-1} + \beta_3 * PubVol_{it-1} + \beta_4 * PriVol_{it-1} + \\ &\beta_5 * MeanDiv_{it-1} + \beta_6 * STDDiv_{it-1} + \beta_7 * BrandPageJoin_{it} + \\ &\beta_8 * MeanDiv_{it-1} * BrandPageJoin_{it} + \sum_{k=9}^{13} \beta_k * Control_{it} + \alpha_i + \gamma_t + \epsilon_{it} \end{aligned} \quad (2)$$

where, $Control_{it} = \{Age_i, SNSAge_i, LoyaltyAge_{it}, NumFriends_{it-1}, Promo_t\}$

The main effect estimates of this variable, β_7, coupled with the interaction with the sentiment divergence variable, β_8, in model (2) provide the main impetus for this part of the analysis. The sentiment divergence scores $MeanDiv_{it}$ and $STDDiv_{it}$ are computed for each user at a monthly level. I use divergence as proxies to capture the self-presentation behavior of content producers online. $MeanDiv_{it}$ denotes the absolute difference between the mean public and private overall sentiment scores, *osent*, for a given user-month. Similarly, $STDDiv_{it}$ denotes the difference between the standard deviation of the public and private sentiment scores for a given user-month. The sentiment divergence score measures the amount of asymmetry in the users' public and private sentiment for the given month. As before, I argue that users who have a higher divergence may be more self-presenting in their social media behavior as compared to users who have a lower sentiment divergence. Thus, high divergence users tend to present content which is high in positive (negative) valence in public while simultaneously presenting high negative (positive) valence in private. Similarly, neutral sentiment may denote objective content. Drawing on Goffman's seminal concepts of audience segregation (Goffman 1959b), I suggest that users choose to display this asymmetry vis-a-vis to present a suitable self-image to the different audience sets. Consequently, a high divergence may be an indication of higher self-presentation behavior in individuals - those who are likely to develop and maintain multiple self-images online in an effort to effectively maintain their self-concepts in front of different audience groups. In contrast, individuals with low divergence have lower levels of audience segregation, i.e., their public and private self-images are similar. These individuals are not as self-presenting and might use social media platforms to seek more utility-related gratifications like information and entertainment (Park et al. 2009). To investigate the effects of divergence on purchasing behavior, I employ a fixed-effect (FE) and random-effect (RE) estimation of the model (2).

I am primarily interested in investigating whether the act of joining the brand page has different effects on high and low divergence users in terms of purchasing behavior. I conjecture that while low-divergence users might be attracted to the more utilitarian aspects of joining the brand page (i.e., staying informed about price promotions and offers), high-divergence users might be more attracted because the brand attachment helps in their image building motives. Consequently, the low-divergence users joining the brand page would end up spending less than the high-divergence users. The estimation results are shown in Table 2-3.

The interaction between joining the brand page and sentiment divergence, however, provides a counterintuitive insight that even though members of the brand page spend less after joining the brand page, this effect is moderated by the level of sentiment divergence in public and private content of the brand page members. More specifically, high divergence brand page members[14] spend $3.85 more[15] than low divergence brand page members, on joining the brand page. These results are consistent with a study which finds that 31% and 27% of the users mentioned that they "liked" brand pages to "share personal good experiences" and to share their "interests/lifestyles with others" respectively (Kalehoff 2013). These people are deemed to be more self-presenting than the other people who join the fan page for utilitarian reasons, such as for discounts. Thus, I find empirical support that even though many users might join brand pages for discounts, those who do so due to strong self-presentational desires, such as the high divergence individuals in my data sample, end up spending more than others with less of such aspirations.

The divergence in sentiments, however, while being an interesting phenomenon, might be indicative of other related behavior as well (e.g., mood swings etc.). Moreover, the divergence measure for capturing self-presentation relies on the use of private communication data that might not always be available to marketers. Thus, to ascertain the role of self-presentation, I perform text mining on the user's public posts to measure the proportion of ego-centric words (e.g., "I", "me", "our", etc.) spoken by an individual in a given time period. Recent studies show that individuals might use such ego-centric words implicitly and thus the degree of their use serves as an unbiased measure of how activated the self-concept of the user is (Laserna et al. 2014). As a result, individuals who are more self-presenting will tend to use

[14] For the purpose of interaction analyses, I assume a self-presentation score of 0 to denote low divergence users and a score of 2.487 (mean + 3*sd) to denote high self-presenting users.

[15] Nearly 23.23% of the average monthly expenditure per quantity sold in my data sample

a larger proportion of self-related or egocentric words in their conversations. Thus, I test model (2) using egocentricity scores as an alternate operationalization of the self-presentation construct. The results from this test are illustrated in Table 2-3(e) and are found to be consistent with the previous results with the sentiment divergence variable[16].

2.5 Robustness Tests

In the previous sections, I provided evidence to show that individuals who join brand pages spend less on average after joining. However, among those who join, high self-presenting individuals tend to spend more than low self-presenting individuals. In this section, I illustrate results from additional tests and methods to address potential endogeneity concerns, and to rule out some alternate explanations. In the next section, I use an alternative measure for purchase behavior of customers. Specifically, I use purchase quantity in place of purchase expenditure to investigate whether the change in expenditure might be due to the sale of expensive items, or due to sale of an increased quantity of items. Then, in Section 2.5.2, I illustrate results from a propensity-score matching (PSM) model to address possible identification concerns, in particular the self-selection of individuals into the brand page. Following this, I devise and illustrate additional methods using natural language processing (NLP) to further show that my divergence metric functions as a proxy for self-presentation, and not other forms of behavior. Finally, I use a panel autoregressive model to address concerns over reverse causality between the purchase expenditures in past months and

[16] The elasticities of expenditure for the brand page join variable and the mean egocentricity score variable are -1.449 and -0.004 respectively.

self-presentation in the current month.

2.5.1 Alternate measure of purchase behavior

In this study, I use purchase expenditure as a measure of an individual's offline purchase behavior. However, an increase in expenditure might be observed due to the purchase of expensive items, or due to the purchase of several units of moderately-priced or low-priced items. To separate these two alternative mechanisms, and to also provide an additional robustness to the existing analyses, I re-estimate the models using purchase quantity as the dependent variable. Specifically, I estimate the following model to test the effect self-presenting customers joining the brand page.

$$\begin{aligned} Quantity_{it} = {} & \lambda_0 + \lambda_1 * PubSent_{it-1} + \lambda_2 * PriSent_{it-1} + \lambda_3 * PubVol_{it-1} + \lambda_4 * PriVol_{it-1} + \\ & \lambda_5 * MeanDiv_{it-1} + \lambda_6 * STDDiv_{it-1} + \lambda_7 * BrandPageJoin_{it} + \\ & \lambda_8 * MeanDiv_{it-1} * BrandPageJoin_{it} + \sum_{k=9}^{13} \lambda_k * Control_{it} + \mu_i + \theta_t + \epsilon_{it} \end{aligned} \tag{3}$$

where, $Control_{it} = \{Age_i, SNSAge_i, LoyaltyAge_{it}, NumFriends_{it-1}, Promo_t\}$

The results from the estimations are illustrated in Table 2-4 and are found to be consistent with the estimation results from the previous section. Additionally, the model in (3) was re-estimated with an individual's egocentricity score as an alternate measure of self-presentation. The results are provided in Table 2-4 (column e), and are consistent with the previous results with purchase expenditure. These results serve as a robustness check for my conjectures on what drives users' offline purchase expenditure when they join brand pages. Additionally, they help us understand that the decrease in expenditure is driven by a corresponding decrease in the number of products purchased by the customer, and not by a change in the type of products purchased (i.e., shifting from more to less expensive products).

Table 2-4: Explaining total purchase quantity after joining brand page

Predictors	**(a): Quantity (FE)**	**(b): Quantity (RE)**	**(c): Quantity (FE)**	**(d): Quantity (RE)**	**(e): Quantity (FE)**
PubSent	-0.0003 (0.001)	-0.0001 (0.001)	-0.0001 (0.001)	0.0002 (0.001)	-0.0005 (0.001)
PriSent	-0.0001 (0.0001)	0.00001 0.0002)	-0.00004 (0.0001)	0.00004 (0.0002)	0.00003 (0.0001)
PubVol	-0.0004 (0.0004)	-0.0004 (0.0002)	-0.0004 (0.0004)	-0.0005 (0.00004)	-0.0003 (0.0005)
PriVol	-0.00003 (0.0001)	-0.00002 (0.00004)	-0.00004 (0.0001)	-0.00004 (0.00004)	-0.0001 (0.0001)
BrandPageJoin	-0.210*** (0.057)	-0.069** (0.032)	-0.250*** (0.062)	-0.094** (0.038)	-0.396*** (0.080)
Age	-	0.002** (0.001)	-	0.002** (0.001)	-
SNSAge	-	-0.00001 (0.0001)	-	-0.00001 (0.0001)	-
LoyaltyAge	-	0.001 (0.001)	-	0.001 (0.001)	-
NumFriends	0.001*** (0.0003)	-0.00004*(0.00002)	0.001*** (0.0003)	-0.00004* (0.00002)	0.001** (0.0003)
Promo	0.006*** (0.002)	0.010***(0.002)	0.006*** (0.002)	0.010 *** (0.002)	0.006 *** (0.002)
MeanDiv			-0.119** (0.050)	-0.090* (0.050)	
STDDiv			-0.007*** (0.021)	-0.007 (0.020)	
*BrandPageJoin * MeanDiv*			0.102** (0.050)	0.062 (0.049)	
Egocentricity					-13.542** (5.450)
*BrandPageJoin * Egocentricity*					20.546 *** (6.528)
Intercept	0.025 (0.095)	-0.020 (0.070)	0.084 (0.098)	0.022 (0.074)	1.094*** (0.190)
Month Dummies	**Present**	**Present**	**Present**	**Present**	**Present**
Brand Page Dummy	**Present**	**Present**	**Present**	**Present**	**Present**
Observations:	16,935	16,935	16,935	16,935	15,684
Number of Consumers:	1,926	1,926	1,926	1,926	1,917
R-squared	0.113	0.101	0.113	0.105	0.123
Standard errors in parentheses	*** p<0.01, ** p<0.05, * p<0.1				

Note:

i. The subscript *t*-1 for the first few predictors indicates that the variable is lagged by 1 time period. This is done to avoid concerns involving reverse causality in the same time period.

ii. All measures of Standard Errors reported are robust.

iii. A post-estimation Hausman Test was performed on FE and RE estimation results to evaluate the consistency of the alternate estimator and only the FE estimator was found to be consistent.

2.5.2 Estimating treatment effect of joining brand page using Propensity Score Matching

In earlier analyses, I provided preliminary evidence that individuals reduce their purchase expenditure upon joining a brand page. However, the treatment effect of joining the brand page could be largely over-estimated or under-estimated if the treatment is correlated with individual-level covariates (e.g., biological age, SNS tenure, social media activity, etc.). In this section, I address the exogeneity assumption of the treatment effect to provide a more accurate estimate of the treatment. A direct application of the propensity score matching estimator is complicated by the presence of time-varying covariates in my model, namely, the ones capturing an individual's social media activity (e.g., $PubSent_{it}$ and $PriSent_{it}$). Therefore, I estimate the treatment effect within each defined time period, contingent on the availability of close propensity score matches to the treated individuals in the same time period. First, I employ a logit model to predict an individual's propensity to join the brand page at a particular time period as a function of time-invariant covariates and time-varying covariates that are held constant within the specified period. I use Age_i, $SNSAge_i$ and $LoyaltyAge_{it}$ as time invariant covariates for matching, and use $NumFriends_{it}$, $PubSent_{it}$, $PriSent_{it}$, $PubVol_{it}$ and $PriVol_{it}$ as the time-varying covariates. Once the propensity scores are computed for each individual, I construct a suitable control group for each individual *i* who has joined the brand page, within each month, based on a nearest-neighbor match of *3* closest contenders. Contingent on finding one or more suitable matches in the control group for an individual *i* in month *t*, I perform a t-test to estimate the average treatment effect of joining the brand page. The results from the propensity score estimation based model are illustrated in Table 2-5 below. Based on estimations from a subset of months when I could construct close matches to the treated individuals, I provide more robust evidence that joining the brand page does indeed reduce purchase expenditure and quantity for the brand's customers.

Table 2-5: Average treatment effect of joining brand page using PSM

Timeline	Expenditure	Quantity
Month 1	-3.263** (1.690)	-0.083 (0.091)
Month 2	-5.844* (3.222)	-0.373* (0.216)
Month 12	-24.715*** (6.809)	-1.240*** (0.386)
Month 13	-27.616*** (7.644)	-1.236*** (0.423)
Estimator	Propensity Score Matching	Propensity Score Matching
Treatment model	Logit	Logit
No. of observations	1394	1394
No. of nearest matches	3	3

*** <0.01, ** <0.05, *<0.1

The propensity score matching hinges on selection based on observable attributes. There could, however, be unobservable factors that might influence both the treatment as well as the outcome, thereby violating the unconfoundedness or selection-on-observables assumption. While it is not possible to correct or control for these confounds, it is possible to perform a sensitivity analysis to quantify the extent of the bias, should these confounds exist. I follow the bounding approach as proposed by Rosenbaum (2002) to provide upper and lower bounds on how strongly an unobserved variable must influence the selection process to create a significant bias in the inference generation. The sensitivity analysis showed no violations to the unconfoundedness assumption even if the unobserved variable increased the odds of joining the brand page by over 100%.

2.5.3 Sentiment divergence and content divergence

In previous sections, I have focused on sentiment divergence. I have contended that social media platforms can effectively use the sentiment divergence metric as a proxy for self-presentation behavior to predict offline purchases – that there is value in using natural language processing (NLP) algorithms for behavioral targeting to predict consumer purchases. However, one could argue that sentiment divergence, alone, may not be sufficient evidence to identify self-presentation in individuals, but could potentially be capturing other aspects of human behavior. To address this concern, I show here that sentiment divergence correlates well with other types of qualitative measures, in content produced and topics discussed, that are generally associated with self-presentation. Using text mining techniques on the social media content, I highlight two findings. First, consistent with past research by (Hogan 2010), I show that public content is semantically different from private content. Second, within each channel of content (public/private), high divergence users talk about semantically different topics from low divergence users. I select high divergence users as those who have a divergence score that is three standard deviations greater than the mean divergence score. Low divergence users are those who have a divergence score of 0.

Using text-mining algorithms, I devise a method based on word distributions from public and private conversation corpuses. Firstly, I select keywords that reveal one's writing style and tone. Based on previous work (Laserna et al. 2014; Wang et al. 2013) and unlike other NLP algorithms, these words include some "stopwords" which are typically removed in traditional text-mining approaches. I believe that these words capture useful behavioral information on social media, especially due to the abbreviated nature of social media content. I focus on affect-sensitive words that form a large share of our online self-disclosure. I obtain a list of affective conditions that are experienced by individuals from the popular PANAS and STEM scales (Levine et al. 2011; Watson et al. 1988). I perform keyword search from the corpuses that fit into the definition of this list of affects. In total, I search for 5 positive affects (Energy, Contentment, Joy, Love, Pride) and 5 negative affects (Anger, Anxiety, Envy, Guilt and Sadness), as consistent with the above mentioned PANAS and STEM scales.

Furthermore, I compare the ego-centric versus alter-centric words, that is, words that refer to the self and those that refer to the others. As a departure from traditional NLP, I identify user centricity using words that are typically removed – the so-called "stop-words." These are generally not used due to their high frequency and little informative nature in topic modeling and other text-mining methods. However, in the current context, I believe they can be intelligently leveraged to classify ego-centricity. Therefore, ego-centric words tend to indicate a higher emphasis on the concept self with words like "I", "My", and "Our". Alter-centric words, on the other hand, indicate a higher emphasis away from the concept of self with words like "You", "His" and "Their". Finally, I focus on temporal words that indicate a sense of time with words like “Day”, “Week”, and “Year”. Recent studies have uncovered the relationship between time cues and concepts of self (Gino and Mogilner 2013; Kouchaki and Smith 2013). The studies found that implicitly activating the construct of time and particularly the early hours of the day, caused individuals to act in a more ethical and honest manner.

Second, I perform a deeper semantic analysis on my dataset at a corpus level. The *All Content* corpus serves as a global list of all public and private content generated by the focal users on the social media platform. Next, I split this into the *High-divergence Public* and *High-divergence Private* corpuses that include the public postings and private conversations of individuals who show high divergence on the platform. Similarly, I obtain the *Low-divergence Public* and *Low-divergence Private* corpuses. From the above corpuses, I extract the 500 most frequently used terms in each of the corpuses and record the frequency of the focal words (i.e., affective words, egocentric words, etc.) from these word lists. The analysis is done at the corpus level rather than the individual level due to sparsity issues. With 240,013,021 textual content and 72 gigabytes of public and private content, the large-scale and sparse data poses a significant computational and memory challenge for traditional NLP methods. Traditional methods such as Latent Dirichlet Allocation (LDA) topic modeling are not feasible because they require loading large corpuses into memory and computing large matrices in a non-parallel manner. My simple keyword-based method is appropriate for such “Big Data” because each piece of content can be computed individually, and in parallel. The algorithm is fast and requires little computer memory. To further improve the computational speed, I deploy a Map Reduce method to parallelize the analysis on a super computer cluster with over 10,000 nodes. I believe that this method is not only practical and feasible, but is valuable to marketers and platform owners as I show below.

The word frequency distribution for each of these categories across the channels and divergence levels are described in Table 2-6 and the distribution of scores visually illustrated using a heat map in Figure 2-5. First, we observe that the individual corpuses are semantically different from the *All Content* corpus. The combined corpus fails to represent and highlight the idiosyncratic aspects of the individual channels (public vs. private) and the individual divergence levels inside each channel (high vs. low).

We note that the word distributions give us interesting insights into how high and low divergence individuals behave online. For instance, while we can find plenty of mentions of attentiveness and energy related words by high divergence users in both public and private channels, we find no such mentions in content by low divergence users. This is intuitive as users who are more self-presenting in nature might resort to using affect sensitive words that signal a positive self-image. This is consistent with previous work on self-regard which highlights that individuals in general, and particularly those with self-presentational aspirations, tend to have a need to maintain a positive self-regard (Allport 1955; Epstein 1973; James 2013; Maslow 1943; Steele 1988; Tesser 1988). While such findings tend to be culture-specific (Heine et al. 1999), the fundamental needs of desiring to maintain a positive self-view(R. F. Baumeister et al. 1989; Diener and Diener 1996), tendency to improve this self-view (Blaine and Crocker 1993; Greenwald 1980; Taylor and Brown

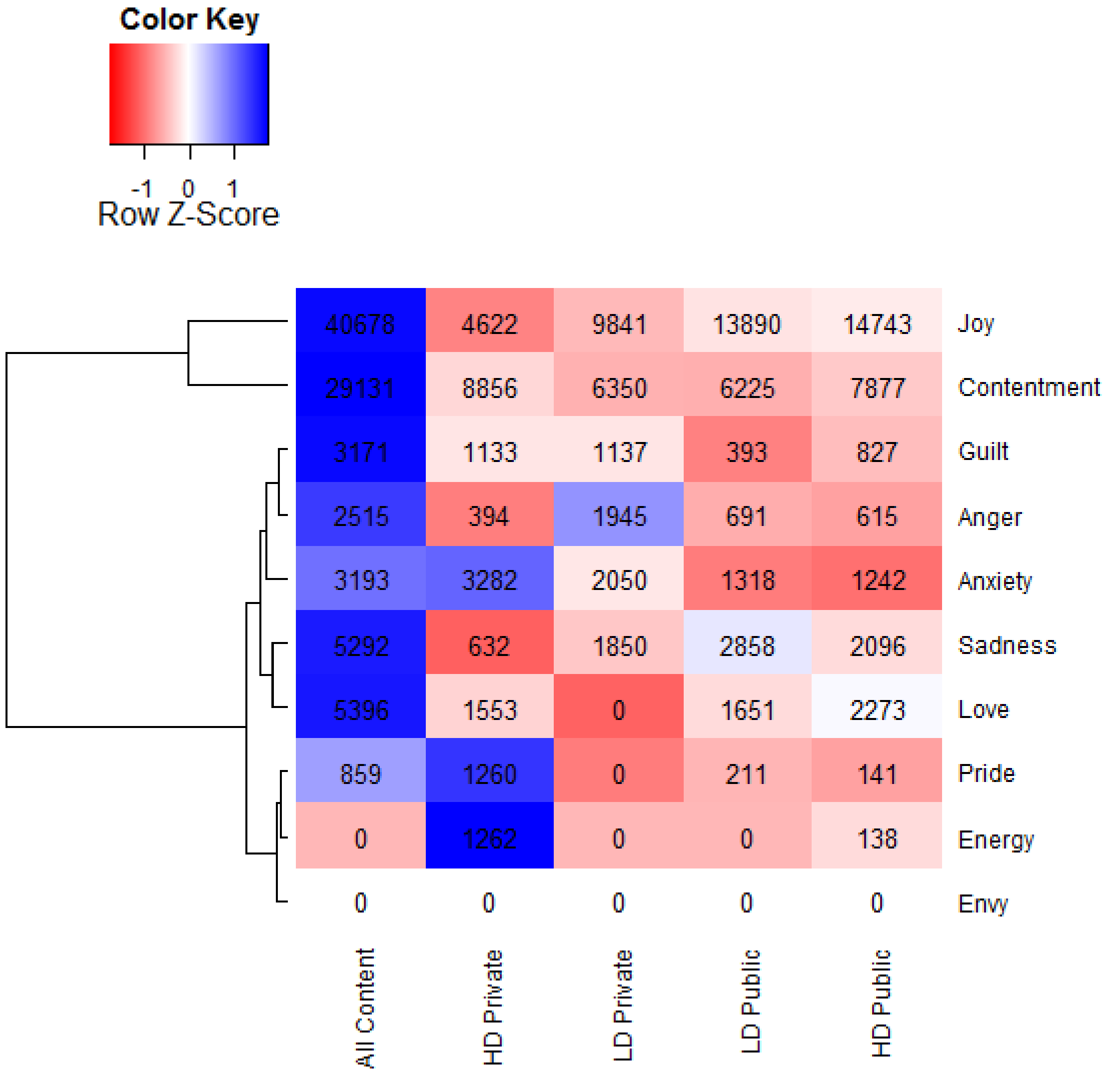
1988), and affinitiy towards any information that helps validate or improve this self-view (Baumeister and Jones 1978; Greenberg and Pyszczynski 1985; Steele 1988; Tesser 1988) tend to be universal, and this reflected in the nature of content produced by the individuals on the SNS. The same pattern is also observed for words relating to the contentment affect. For affects such as joy or love, however, public content by both high and low divergence users show greater mentions of these words than private content by these users. For negative affects, we can find more mentions in private conversations than in public postings. This is line with past studies which show that people are more comfortable talking about negative emotions in private than in public (Hogan 2010). For instance, low divergence users mentioned anger related words in private conversations nearly three-times as often as in public. Interestingly, however, high divergence users mentioned anger related words in public posts nearly 1.5 times as often as in private conversations. This implies that high self-presenting individuals might be more willing to express their anger in public to strategically make a point or gather social approval (Andrade and Ho 2009b)

Table 2-6: Word-frequency distributions for the various content corpuses

	HD Public	LD Public	HD Private	LD Private	All Content
Affect-sensitive words:					
1. ***Anger***	615.000	691.000	394.000	1945.000	2515.000
2. ***Anxiety***	1242.000	1318.000	3282.000	2050.000	3193.000
3. ***Attentiveness/ Energy***	138.000	0.000	1262.000	0.000	0.000
4. ***Contentment***	7877.000	6225.000	8856.000	6350.000	29131.000
5. ***Envy***	0.000	0.000	0.000	0.000	0.000
6. ***Guilt/Shame***	827.000	393.000	1133.000	1137.000	3171.000
7. ***Joy***	14743.000	13890.000	4622.000	9841.000	40678.000
8. ***Love***	2273.000	1651.000	1553.000	0.000	5396.000
9. ***Pride***	141.000	211.000	1260.000	0.000	859.000
10.***Sadness***	2096.000	2858.000	632.000	1850.000	5292.000
Total Positive Affect $^{3+4+7+8+9}$	25858.000	22159.000	17426.000	17328.000	78376.000
Total Negative Affect $_{1+2+5+6+10}$	4780.000	5260.000	5441.000	6982.000	14171.000
Pos/Neg Affect Ratio (PNAR)	5.410	4.213	3.203	2.482	5.531
11.***Egocentric Words***	1359.000 (1.870%)	1150.000 (1.796%)	2222.000 (2.800%)	3936.000 (5.784%)	10331.000 (3.964%)
12.***Alter-centric Words***	27024.000 (37.178%)	17806.000 (27.804%)	33547.000 (42.263%)	27959.000 (41.090%)	107763.000 (41.347%)
Egocentricity Measure $^{11/12}$	0.050	0.065	0.066	0.141	0.096
13.***Temporal Words***	10127.000 (13.932%)	14140.000 (22.079%)	15858.000 (19.978%)	11006.000 (16.175%)	40647.000 (15.596%)

Note: the percentages mention in brackets denote the proportion of words for that corpus.

Figure 2-5: Affect distribution across corpuses



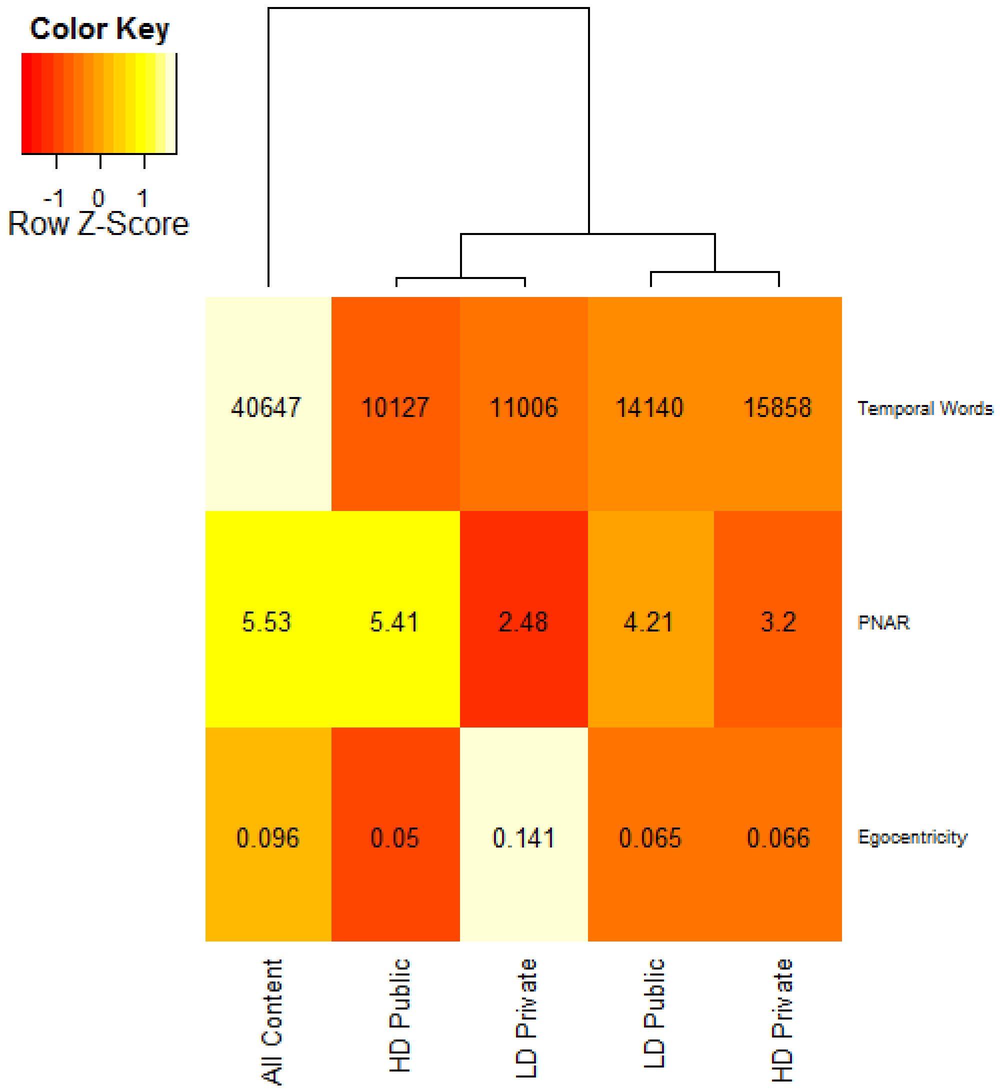
I also construct the ratio of the number of positively affective to negatively affective word mentions to form the Positive-to-Negative Affect Ratio (PNAR) metric. Unlike my sentiment score, the PNAR is based on all words and not just polar words. Thus, the latent meaning of non-polar words is also taken into consideration. Also, my PNAR metric is based on a richer definition of the affective state of the individual based on the PANAS scales. I present the group-wise PNAR scores in Table 2-6, and their score distribution in Figure 2-6. I find that the combined corpus has a PNAR ratio of 5.53 which is more than that for all content produced in private (3.20 and 2.48) as well public (5.41 and 4.21). We can also observe that high divergence (i.e., self-presenting) individuals have the highest PNAR ratio (of 5.41). As mentioned in the previous para, this is consistent with previous work on self-regard(Heine et al. 1999), self-esteem(Mehdizadeh 2010) and self-presentation(Baumeister and Jones 1978; Hogan 2010) which emphasizes that all individuals, and especially self-presenting individuals would seek to construct, project and maintain a positive self-image in public. This argument is further validated by our finding that that public content has a higher PNAR ratio than private content for both high and low divergence individuals, although this difference between public and private PNAR ratio is higher for high divergence people than for low divergence people.

We know from past studies on impression management that self-presentation manifests itself in several forms. I use one such form, egocentricity, as a measure to distinguish image-seeking people having high self-presentational intentions from those who have no such intentions. I measure egocentricity as the ratio of the mentions of self over others. I report these group-wise egocentricity scores in Table 2-6, and their score distribution in Figure 2-6. The results show that the All Content corpus has an egocentricity ratio of 0.1 which is mildly less than that of low divergence private corpus while being almost double that of the high and low divergence public corpus. I notice that, while the egocentricity value for low divergence users in private is double of that in public, the high divergence users show nearly the same egocentricity ratio for both public and private channels.

Similarly, for the mention of time-related cues, we can observe from Table 2-6 and Figure 2-6 that 15.6% of all focal words mentioned in the all content corpus are time related. However, this fraction drops to 13.93% for the high divergence public corpus while increasing to 22.08% for low divergence public corpus. These findings can be explained by understanding the relationship between time and the concept of self-image. A recent study shows that thinking about time reminds people of their own lives and tends to make them less dishonest in their affairs (Gino and Mogilner 2013). A dominant view of the impression management literature has been that self-presentation behavior is associated with inaccuracy and exaggeration (i.e., we present idealized versions of ourselves) (Baumeister 1982; Gurevitch 1984; Schlenker 1986). This is consistent with my results showing that self-presenting individuals in public would use fewer time-related words as compared to people who are not as self-presenting.

The above analyses suggest that beyond sentiment divergence, semantic differentiation between high and low divergence users can be made on more nuanced levels using egocentricity measures, affect measures and time-related cues. Using these examples as illustrations, I show that divergence as measured based on sentiment of content also correlates to significant difference in the type of words generated, and their relative frequencies. I show that there exist semantic differences not just among channels but also among divergence types of users within each channel. Unlike prior studies on user-generated content, I characterize the textual and semantic differences in self-presenting behavior, showing that individuals who are more self-presenting (high-divergent) tend to have a distinct pattern of content production from individuals who are less self-presenting (low-divergent). Realizing this difference in pattern can help explain the economic value of these content producers, and subsequently be effectively used for targeting users on social media platforms.

Figure 2-6: Distribution of PNAR, egocentricity and temporal cues across corpuses



2.5.4 Autoregressive evolution of expenditure and self-presentation behavior

While investigating the interplay between online and offline behavior, it is not uncommon to expect linear temporal interdependencies between the offline and online variables, namely purchase and self-presentation. Specifically, there is a possibility that lagged values of customer expenditure might influence current values of self-presentation and vice versa. Therefore, to explore and possibly rule out such autoregressive behavior, I performed an additional robustness check, in which I allowed for lagged values of expenditure to affect future values of self-presentation, i.e., a SNS user’s self-presentation behavior is influenced by her prior purchase history. I estimated a panel-level vector autoregressive model (VAR) (Holtz-Eakin et al. 1988) with varying lag levels. The results from the VAR analysis are shown in Table 2-7 and illustrate that the impulse response estimates for the effect of the first and second order lagged expenditure, Expend (L1) and Expend (L2) on MeanDiv are not statistically significant. Thus, I conclude that reverse causality problems with regard to an individual's expenditure and her self-presentation behavior are not significant in the current context (i.e., the focal regressors are not correlated with the errors).

2.6 Research Contributions

This study makes five key contributions to the current marketing and information systems literature and practice. First, previous research has focused on the effects of user-generated content (UGC) on product adoption and economic outcomes of the content consumer. However, this is among the first studies to exploit content produced on social media to uncover motivations behind the content producer's online behavior (i.e., joining a brand page) as well as offline behavior (i.e., making a purchase). While prior research have overwhelmingly focused on the content consumer’s perspective, recent studies seem to indicate that content producers' perspective is equally, and perhaps more, valuable to the platform owners (K. Zhang et al. 2012). Understanding the content consumer’s motivations can also help with targeting and customer segmentation.

Table 2-7: Orthogonal impulse response function estimates

	Response Of			**Response Of**	
Response To	*Expend*	*MeanDiv*	**Response To**	*Expend*	*MeanDiv*
PubSent (L1)	-0.060 (0.038)	-0.00002 (0.001)	*PubSent (L2)*	-0.007 (0.025)	-0.00002 (0.001)
PriSent (L1)	0.004 (0.004)	0.0001 (0.0001)	*PriSent (L2)*	0.004 (0.003)	0.0001 (0.0001)
PubVol (L1)	-0.033 (0.020)	0.0001 (0.001)	*PubVol (L2)*	-0.0004 (0.013)	0.0001 (0.001)
PriVol (L1)	-0.006*** (0.002)	-0.0002*** (0.0001)	*PriVol (L2)*	-0.003** (0.001)	-0.0002*** (0.0001)
MeanDiv (L1)	-9.246*** (1.641)	-0.227*** (0.048)	***MeanDiv (L2)***	-6.908*** (0.931)	-0.227*** (0.048)
Expend (L1)	-0.168** (0.007)	**0.001** **(0.0004)**	***Expend (L2)***	0.009** (0.005)	**0.001** **(0.0004)**
NumFriends (L1)	0.386*** (0.084)	0.014*** (0.003)	*NumFriends (L2)*	-0.233*** (0.066)	0.014*** (0.003)
BrandPageJoin (L1)	-21.320*** (2.339)	-0.690*** (0.090)	*BrandPageJoin (L2)*	-4.909*** (0.815)	-0.690*** (0.090)

*** <0.01, ** <0.05, *<0.1

Second, this study uncovers the importance of investigating the economic value of different channels of communication in UGC at the individual level, an issue that existing studies seem to have not focused on. In particular, users may actively self-select into the social media sites for self-presentational motives. I show that such individual differences (i.e., high vs. low self-presentation) can be operationalized when we compute a sentiment divergence metric or egocentricity scores from the users' content. My results provide not only theoretically interesting insights about motivations on social network sites, but are also important for practitioners looking to target and channel customers from these online platforms to offline stores. I use purchase data from a brick-and-mortar store to show that self-presenting users tend to spend more offline when they join brand pages on social media platforms. However, there exists a dilemma around how marketers might have an indication of the amount of divergence similar to what I have computed in my models. While it may not be feasible for marketers, with notable exceptions[17], to have access to private user data of their customers, I show that even implicit cues in public UGC can be exploited to identify such self-presentational behavior, e.g., using egocentricity scores as discussed in the previous section. Further, I argue that advertisers should choose to invest more in building online communities in platforms with features that are fundamentally more appealing to self-presenting individuals, such as avatars, photo enhancing filters, etc. In addition, platform owners can design these features into their SNS that caters to certain individuals over others to increase the economic value to their advertisers.

Third, while recent research and media reports have emphasized social media brand communities as important site features that help marketers engage with their present and potential customers, few studies have empirically studied the level and type of engagement on these platforms at the individual-level. Moreover, there have been even fewer studies on the

[17] http://techcrunch.com/2012/06/22/we-are-not-afraid/

monetary value of joining these brand communities and on understanding how such communities benefit O2O commerce. The current study is among the first to study at an individual level the effects of online engagement and membership into brand communities on offline purchases. Furthermore, I offer some insights and evidence into possible mechanisms at play. In particular, I observe that users reduce their purchase expenditures and purchase fewer items upon joining the brand page. This is consistent with the notion that social media brand pages may serve as a leading source of discounts and promotions to brand's customers. Thus, these may attract customers who are largely utilitarian in their buying attitude. Such individuals are likely to be highly price-sensitive and might purchase products when there is an appropriate discount or promotion. However, some customers might have an added motivation for joining these brand communities such as for self-presentation motivations through brand attachment and endorsement. My findings show that while customers, on average, decrease spending upon joining the brand page, this reduction is attenuated for customers who are highly self-presenting. Thus, an important implication of my findings for marketers is to target and market their brand in a way that makes it easier and more attractive for the customers to identify with it. The more customers identify with a product, the more likely they are to signal attachment or endorsement on social media.

Fourth, and related to the previous two points, the analysis of sentiment divergence on the SNS provides opportunities for further theory development and its economic impacts. While the economic importance of offline self-presentation has been sporadically discussed in some prior work (Argo et al. 2005; Slama and Celuch 1995) , this empirical study is among the first to study cross-channel substitution effects of online self-presentation on offline consumer behavior.

Finally, this paper makes a methodological contribution by showing that using simple textual measures of sentiment divergence and egocentricity scores, we are able to explain offline purchases made by the content producers in a significant way. The text mining algorithms employed in this study can be computed in a space-time efficient fashion based on a hash-map computation with linear, O(M), dictionary lookup run-time complexity. Moreover, this approach is greatly scalable through parallelization due to the fact that the algorithms are naturally map-reducible using any conventional map-reduce application. As a result, this method can be applied to large computer clusters used in typical social network platforms. I also perform deeper text mining on the UGC to uncover implicit cues in the content produced by different individuals (Laserna et al. 2014). With tremendous amounts of content being generated on social media each day, it has become imperative to look at such non-experimental and computational approaches to identify and analyze user data online (Park et al. 2015).

While the results from the current set of analyses are promising and shows the value of text mining for behavioral targeting in analyzing value creation on the SNS and SNS brand pages, the present study has a number of limitations. First, my current data context includes user data from just one, albeit large and popular, SNS and its associated brand page for a fashion retailer. Moreover, this particular brand is representative of brands that advertise online but sell exclusively through brick and mortar stores. However, the results from my study might not generalize directly to firms that sell through multiple channels. In future work, it would be useful to investigate the behavior of self-presenting individuals who purchase through both channels. Second, I currently lack purchase data from competing fashion retail stores in the same region. As a result, it is hard for me to empirically explain whether individuals decrease their net expenditure on joining a brand page, or merely redistribute it across other retailers. Third, while I focus on fashion apparel in this study, it is possible that there might be certain product characteristics that are distinct from other non-conspicuous consumer goods. Thus, there could be unobserved taste shocks relevant to these types of products that are visible to the consumer, but not to the econometrician. Finally, I use content sentiment divergence as a simple proxy to measure the self-presentation behavior. However, I do not differentiate between specific modes of self-presentational behavior like, for instance, acquisitive or protective self-presentations (Slama and Celuch 1995; Slama et al. 1999). I believe that identifying these specific tactics from consumer data and correlating them with purchase information might provide a fruitful direction for future research.

2.7 Conclusions

In this study I investigate the effects of joining SNS brand pages on subsequent offline purchases made by the brand follower. I also analyze the moderating role of self-presentation on the purchase behavior. Empirically, I find that individuals, on average, spend less and purchase fewer quantities after joining brand pages but this reduction is attenuated for individuals who are more self-presenting than others. I propose a measure of self-presentation behavior that is captured by the amount of public-private sentiment divergence and the proportion of egocentric word usage, and show that these measures can explain the user's purchase expenditure and quantities. I contend that sentiment divergence and the egocentricity scores are indicators of self-presentation on social media and provide a deeper analysis using NLP approaches to validate this assertion. I also account for potential self-selection into the brand page using a PSM approach, and control for potential autoregressive interactions between the sentiment divergence construct and dependent variables of purchase behaviors, to show that the results are robust to such endogeneity concerns. In summary, this study offers illuminating theoretical insights on the offline economic impacts of online self-presentational behavior on social media. My results also provide insights to brand owners and digital marketers on how to effectively monitor and tap into the potentials of social media brand communities, and target valuable customers on these communities.

In the following study, I shift focus from external brands to understanding how the SNS users engage with each other on the SNS via content generation and friendship creation, to create value each other, as well as for the SNS providers.

SENSITIVITY ANALYSES - I

	Expend	*Quantity*	*PubSent*	*PriSent*	*PubVol*	*PriVol*	*MeanDiv*	*STDDiv*	*Fanpage Join*	*Age*	*SNSAge*	*LoyaltyAge*	*NumFriends*	*Promo*	*Ego*	*Gender*
Expend	1.000	0.944	0.033	0.005	-0.010	0.009	0.028	0.025	-0.074	0.027	-0.009	0.013	-0.019	0.122	-0.002	-0.021
Quantity		1.000	0.030	0.006	-0.007	0.007	0.027	0.025	-0.067	0.031	-0.001	-0.010	-0.026	0.123	0.0002	-0.026
PubSent			1.000	0.189	-0.015	0.007	0.160	0.062	-0.012	-0.076	0.012	-0.002	0.223	-0.011	0.041	0.035
PriSent				1.000	0.018	-0.030	0.074	0.074	-0.014	0.010	0.045	-0.003	0.045	-0.008	0.036	-0.013
PubVol					1.000	-0.762	-0.016	-0.012	-0.006	0.012	-0.010	0.005	0.050	0.006	-0.002	-0.013
PriVol						1.000	0.010	0.011	0.012	-0.023	0.00001	-0.001	-0.041	-0.006	-0.003	0.016
MeanDiv							1.000	0.278	0.0004	0.056	0.012	0.018	-0.056	-0.015	0.006	-0.006
STDDiv								1.000	0.021	0.030	0.025	0.017	-0.040	-0.024	0.033	-0.028
Fanpage Join									1.000	-0.079	0.039	0.148	0.049	0.049	0.054	0.010
Age										1.000	0.076	0.047	-0.195	0.035	-0.064	-0.030
SNSAge											1.000	0.126	0.012	-0.011	0.041	0.035
LoyaltyAge												1.000	-0.054	0.021	0.019	-0.056
NumFriends													1.000	0.013	0.029	0.138
Promo														1.000	-0.005	-0.025
Ego															1.000	-0.021
Gender																1.000

Table S1: Inter-correlation table for log-transformed and standardized variables

SENSITIVITY ANALYSES – II

Predictors	(a): Expend (FE)		(b): Quantity (FE)	
PubSent	-0.005	(0.010)	-0.006	(0.010)
PriSent	-0.002	(0.007)	-0.002	0.007)
PubVol	-0.001	(0.010)	0.005	(0.010)
PriVol	0.003	(0.011)	0.007	(0.010)
BrandPageJoin	-0.290***	(0.029)	-0.247***	(0.028)
Age		-		-
SNSAge		-		-
LoyaltyAge		-		-
NumFriends	0.137***	(0.051)	-0.0003	(0.001)
Promo	0.331***	(0.024)	0.145***	(0.050)
MeanDiv	-0.046**	(0.017)	-0.049***	(0.017)
STDDiv	0.003	(0.008)	0.0003	0.008)
BrandPageJoin * MeanDiv	**0.026*** **	**(0.001)**	**0.247*** **	**(0.028)**
Gender	0.331***	(0.024)	0.100***	(0.010)
Intercept	0.106***	(0.035)	0.061*	(0.035)
Month Dummies	**Present**		**Present**	
Brand Page Dummy	**Present**		**Present**	
Observations:	14,531		14,531	
Number of Consumers:	1,726		1,726	
R-squared	0.131		0.131	
Standard errors in parentheses	*** p<0.01, ** p<0.05, * p<0.1			

Table S1: Fixed-effects regression results using log-transformed and standardized variables

Chapter 3. IMPACT OF NETWORK EFFECTS ON CONTENT GENERATION ON THE SNS

3.1 Introduction

With the proliferation of Social Network Sites (SNS), platform owners are facing increasing challenges with regards to engaging users and, subsequently, generating revenue from advertisements (Edwards 2012; Hof 2011; Tucker 2012). Unlike other Internet-based services, the unique value of SNS lies in engaging interactions between two user roles – “content producers” who actively post, comment and share content with their friends, and “content consumers” who view and react to such content. Content producers in particular add considerable value by generating and sharing content through the network.

The content posting behavior of users as well as the users’ propensity to make new connections on social media is influenced partly by individual level factors (e.g. demographics, traits, etc.) and partly by their online social network characteristics such as the number of online friends, their network clustering, their network betweenness and so forth (Lu et al. 2013; Newman 2010). From previous research, it remains an empirical puzzle to estimate how a user’s social network, such as the number of friends or the extent of clustering in the user’s network, influences the user’s content generation behavior. The key challenge lies in that the user’s content generation behavior and social network co-evolve by influencing each other, i.e., behavior shapes the network at the same time that the network shapes behavior. From a classical social network perspective, solving this puzzle amounts to separating the effect of social influence (i.e. when network influences attitude/behavior) from homophilous interactions (i.e. when attitude/behavior influences network formation) and context effects (Borgatti and Foster 2003; McPherson et al. 2001; Shalizi and Thomas 2011).

There exist certain methodological limitations with previous approaches seeking to disentangle homophily from influence. Prior experimental approaches to addressing this problem usually estimate either influence or homophily while controlling for the other (Sacerdote 2001; Toubia and Stephen 2013). They also suffer from low ecological validity, e.g., it is hard to imagine real-world situations where friendship formation would be truly exogenous. Specifically, the process of holding either influence of homophily constant is highly stylized and never occurs naturally in the real world. Thus, research methods attempting to manipulate social ties using interventions tend to have lower ecological validity than observational methods. Non-experimental approaches have employed either a contingency table method (Fisher and Bauman 1988; Kandel 1978), an aggregated personal network based method (Kirke 2004; Yoganarasimhan 2012), or structural equation models (Iannotti et al. 1996; Krohn et al. 1996) to try and address this problem. However, most of these studies suffer from one or more of the following three limitations. First, they tend to ignore the network dependence of users, i.e., their dyadic independence assumption is often violated in real networks. Second, they fail to control for other competing mechanisms like that of shared contexts. Third, the studies do not take into account the possibility of errors introduced due to incomplete observations, as is common with discrete time models.

In addition, there exist a number of important theoretical gaps that remain unaddressed. For instance, earlier studies have investigated the presence of either homophily or social influence in separate contexts (Lazarsfeld et al. 1954; McPherson et al. 2001). The few that do look at their co-existence within a single context, tend to focus primarily on the relative strengths of homophily and influence (Borgatti and Foster 2003; Ennett and Bauman 1994; Kirke 2004). However, it is quite plausible that both homophily and influence play an important role in different temporal stages of the individual's evolution, and to varying extents. In addition to this temporal dependency of homophily and influence, there could also exist a dependency on the specific state of the behavior or preference in question, i.e., is an individual equally susceptible to a change in friendship or behavior at all levels of magnitude of the behavior in question? These are critical theoretical questions that have significant practical implications for platform owners and marketers. I seek to answer these through the current study.

In the present study, I improve on previous approaches by developing an actor-based and continuous-time co-evolution model that operates under a set of Markovian assumptions to jointly specify and estimate the evolution of the online social network and the evolution of online content generation behavior of the user. Generally, closed form solutions are not possible for the likelihood function in such models, making estimation methods such as maximum likelihood or Bayesian estimation inadequate. To overcome this hurdle, I resort to a simulation-based estimation framework based on Markov Chain Monte Carlo (MCMC) estimations (Snijders 2001; Steglich et al. 2010). Specifically, I use a MCMC based Method of Moments (MoM) estimator to estimate the co-evolution parameters in my model. While some prior studies have used computational simulations to model endogenous evolution of network ties and individual attributes (Carley 1991; Macy et al. 2003), the model I use allows for statistical inference testing, model fit assessments and counterfactual simulations. Moreover, it is flexible enough to allow for different forms of objective functions and operates under an acceptable set of assumptions (e.g., conditional independence etc.).

The idea of jointly modeling the co-evolution of networks and behavior using a statistically sound framework was introduced by Snijders et al (2007). For the purpose of this work, I build upon and extend Snijders et al. (2007) in certain key ways. Firstly, I model the coevolution in an online dynamic behavioral setting, where the behavioral traits are not limited to a dichotomous variable, as was the case with previous studies (e.g. smoking vs. no smoking, alcoholic vs. non-alcoholic). Instead, I discretize the number of posts made by the user into quantiles. This provides us with added information about posting behavior and increased flexibility in modeling changes in behavior over time. Secondly, to the best of my understanding, this is the first study that attempts to adapt the actor-driven approach beyond slow-moving and relatively stable traits and behaviors (e.g. music tastes, smoking habits etc.) to a dynamic and rapidly-changing behavioral setting (e.g. online posting and messaging behavior, photo uploads etc.). Third, prior applications of the co-evolution model have not modeled peer effects contingent on specific levels of the traits or behavior. However, I believe that individuals are likely to display varying extents of sensitivity towards peer effects depending on their current level of traits or behavior. By suitably specifying the co-evolution model, I uncover that homophily and peer influence, based on posting behavior, are sensitive to the current level of the users' posting behavior.

From the subsequent analyses, I make inferences about the nature and extent of peer influence as well as homophilous peer selection in content production on SNS. First, I find clear evidence for homophily based on similarity in content posting behavior, but not on the basis of individual covariates, like age or gender. Second, I observe the existence of peer influence, but in a direction opposite to that of homophilous interaction. This provides an interesting insight about the opposing roles of behavioral similarity at different stages of friendship formation. I show that individuals befriend others who are similar in content production during the friendship formation stage, but gradually diverge in content production behavior from these similar others over time. Third, and as mentioned earlier, I provide evidence that the strength of homophilous friend selection as well as social influence varies as a function of the specific level of the behavior in question i.e. content posting behavior in the current study. Specifically, I find that low content posters are more susceptible than heavy content posters to homophilous friend selection. However, once they make friends, low posters are more likely to diverge from their peers as compared to heavy posters.

In summary, this study offers a statistically disciplined approach to modeling the co-evolution of online social network structure and posting behavior on the SNS. Using a MCMC-based stochastic simulation model, I uncover insights about the mechanisms that drive peer effects and the behavioral dependency of these mechanisms on SNS. The findings from this study not only provide important theoretical implications, but can also be used to generate actionable recommendations for social network platform owners, social media marketers and advertisers.

In the following section, I present a summary of previous studies that discuss peer effects in social networks and a relatively newer set of studies that have used the co-evolution model in varying contexts. Next, I offer a brief summary of the coevolution model that I use in the empirical analyses. Following this, I discuss the empirical setting and demonstrate my findings. I conclude with a discussion of the key contributions of this study, the limitations, and a roadmap for future extensions.

3.2 Background and Related Work

3.2.1 Peer effects in social networks

Social science researchers have always been interested in understanding the interdependence between the behavior of group members and the group's structure, as reflected by the inter-member ties within the group. For instance, sociologists and psychologists have long discussed the effect of social cohesion among group members on norm compliance and deviance (Asch 1951; Durkheim 1884; Homans 1961). Researchers have also investigated the role of individual actions on emergent social outcomes and social structures (Emirbayer and Goodwin 1994; Homans 1961; Stokman and Doreian 1997).

More recently, researchers have observed that the preference and behavior of individuals tend to be more similar when they are connected in a relationship, than when they are not (Hollingshead 1949; Newcomb 1962). This phenomenon has been studied under various names, the most common of which are homogeneity bias (Fararo and Sunshine 1964) and network autocorrelation (Doreian 1989). The increased focus on understanding the mechanisms that lead to such network autocorrelation was initially driven by a need to understand the onset, and diffusion of addictive behaviors, including smoking, alcoholism, and substance-abuse among adolescents (Brook et al. 1983; Cohen 1977; Kandel 1978). However, over time, network autocorrelation has been also observed and studied extensively in the context of online social networks (Aral and Walker 2014; Aral et al. 2009, 2013; Lewis et al. 2012).

While some sociologists and social psychologists propose the idea of social influence or network-driven assimilation as a potential cause of such effects (Asch 1951; Friedkin 2001; Oetting and Donnermeyer 1998; Singh and Phelps 2013), others propose selection-based mechanisms like homophily to explain why such effects might occur (Aral et al. 2009, 2013; Lazarsfeld et al. 1954; McPherson et al. 2001; Nahon and Hemsley 2014). A third line of research challenges both influence and homophily based explanations and, instead, focuses on the role of a shared context between the networked individuals as a driving factor (Feld and Elmore 1982; Feld 1981, 1982). Borgatti and Foster (2003) described these competing perspectives in terms of the temporal ordering and causal validity of network or behavioral change (Borgatti and Foster 2003). Specifically, they suggest that if behavior is the consequence of network change, then this is explained by peer influence. If, however, the network is the consequence of behavior change, then this is explained by selection mechanisms such as homophily, but only if the temporal antecedence is causal. If it is merely correlational, then this can also be explained as a result of shared social contexts. Understanding what drives network autocorrelation in various contexts remains an open empirical puzzle, and several previous works have pointed out this underlying tension among the competing perspectives (Ennett and Bauman 1994; Kirke 2004; Michell and Pearson 2000; Pearson and West 2003).

It is a challenging exercise to disentangle these competing mechanisms in real-world contexts. While there have been several different approaches, there has been limited progress in identifying competing mechanisms through using observable data. Experimental approaches using lab and field studies have attempted to intervene with either the network or the behavior in order to identify the other (Aral and Walker 2014; Asch 1951; Herman et al. 2003; Sacerdote 2001). These approaches have been useful in uncovering causal relationships, but often at a price of reduced ecological validity. For instance, it is debatable whether friendship formation can truly be exogenized without losing realism. In addition, while experimental approaches are considered to be the holy grail of inference testing, some can be difficult to execute, while others face issues with ecological and external validity of the population (Berkowitz and Donnerstein 1982; Falk and Heckman 2009). Moreover, extensive longitudinal field studies are complicated to design, time consuming, loosely controlled, and potentially face human subject regulations. These challenges have limited the applicability of experiments in network research.

As a result, others have attempted to uncover the impact of peer effects using non-experimental methods. Such attempts can be largely classified under three major categories. The first is the contingency table approach (Billy and Udry 1985; Fisher and Bauman 1988; Kandel 1978), in which dyads of mutually selected friends are selected and cross-tabulated across subsequent periods. The observed measures on a behavioral attribute are similarly recorded. Estimate for influence is then obtained from pairs of individuals whose friendship is preserved over subsequent periods, but who show a change in behavior. In a similar fashion, estimates for selection are assessed from pairs of individuals whose friendship ties change over subsequent periods, but show identical behavior in both periods. A second approach is the aggregated personal network approach which follows a two-step strategy (Cohen 1977; Ennett and Bauman 1994; Kirke 2004; Pearson and West 2003). In the first step, the user's network characteristics (e.g. network structure measures) are collapsed into individual measures (e.g. user's transitivity, betweenness centrality, etc.). In the second step, these measures are used to predict user-level outcomes under an implicit assumption that such measures are independent across observations. Finally, the structural equation modeling approach attempts to model the cross-lagged panel of latent network-level constructs (Iannotti et al. 1996; Krohn et al. 1996). This particular approach is better than the previous two approaches by virtue of its ease of modeling and estimating selection and influence effects simultaneously using a system of equations.

There are three major shortcomings with all the above mentioned techniques for studying peer effects using longitudinal network data. First, most prior methods tend to ignore the network dependence of users. Thus, the assumption of independence across observations is clearly violated in such settings. Second, these methods tend to control or even ignore alternate mechanisms of network or behavior evolution such as the impact of shared social contexts. Lastly, these methods are problematic in the presence of incomplete observations, as is often the case with longitudinal discrete-time datasets where observations about the user and the network are only made at specific points in time, with little information about the inter-period dynamics. However, ignoring the evolutionary dynamics between discrete time periods can significantly affect our ability to make inferences about peer effects, as pointed out by Steglich et al. (2010).

3.2.2 Content production in online social networks

Social network sites (SNS) have been a subject of active research in several disciplines including information systems, marketing, social psychology and computer science. boyd and Ellison (2007) define SNS as "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system." While there have been extensive studies looking at how prolonged use of SNS influences the psychological well-being of the users and how this process of generating online social capital differs from offline users (Hargittai 2007; Steinfield et al. 2008; Valkenburg et al. 2006; Wellman et al. 2001), others have taken a more normative approach to discuss how user engagement increases on the SNS (Fogg and Eckles 2007) and whether engagement on SNS has a positive or negative impact on its users (Binder et al. 2009; Livingstone 2008). Finally, there have been some exemplary efforts probing how organizations use social media to engage more effectively with their target users both inside and outside the organization (Sinclaire and Vogus 2011; Steinfield et al. 2009; Waters et al. 2009).

The creation and spread of user generated content (UGC) as a means of online word-of-mouth (WOM) has interested social network researchers for several decades, and has been extensively used by brand marketers to understand and increase brand awareness, evaluation and sales (Chevalier and Mayzlin 2006; K. Goh et al. 2013; Reingen 1984). The common thread that emerges from the extant literature on WOM is the value of WOM behavior for both the users as well as the platform owners; Higher WOM levels essentially lead to higher engagement levels on the platform. I observe that high levels of self-disclosure on online platforms allow the sites to collect essential user data that can then be used in marketing implementations. Further, ensuring a persistent and critical mass of users on the platform enables advertisers to monetize by delivering more advertisements to the users in a targeted fashion (Goldfarb and Tucker 2011).

A number of studies have investigated the reasons why users develop a propensity to contribute public content on the SNS. The uses and gratifications theory has been successfully extended to understand the motivations behind Internet use (LaRose and Eastin 2004) as well as the motivations guiding the use of SNS groups (Park et al. 2009). These studies have found that SNS provide distinct gratifications through communication, entertainment, information, and status seeking.

Previous work using historical data, however, faces limitations by ignoring the interdependencies of the underlying network structure and the content production process. Consequently, there has been very little work that looks at objectively investigating and solving the network autocorrelation problem in online contexts (Backstrom et al. 2006; Crandall et al. 2008; Singla and Richardson 2008). The few attempts that exist focus primarily on establishing the presence of either influence or homophily and do not provide a flexible model that is geared towards performing stronger inference testing. A few exceptions to this are the recent studies by Aral et al. (2009) and Snijders et al. (2007). Both these models follow fundamentally different approaches for trying to separate homophily from influence. While Aral et al. (2009) use a matched sample estimation framework that hinges on the presence of several user-specific attributes and preferences to perform suitable matching, Snijders et al. (2007) use a more parsimonious random-graph based model in an offline setting with relatively stable behaviors like smoking and alcohol consumption. In the following section, I describe and extend on Snijder’s approach and illustrate the utility of this model in disentangling social effects for dynamic and non-stationary behavior in an online setting.

3.2.3 Stochastic actor-driven co-evolution model

The co-evolution models offer a continuous time scenario in which users simultaneously alter their network ties as well as their behavior at random instants in time, which may or may not be observed by the researcher. Similar Markovian models for longitudinal social network data have a rich history of use in the social networks literature (Holland and Leinhardt 1977; Wasserman 1977). Such continuous time models, in principle, provide greater flexibility and theoretical grounding than comparable discrete time models (Katz and Proctor 1959; Wasserman and Iacobucci 1988). However, some of the earlier continuous-time Markov chain models, like the reciprocity model (Wasserman 1977, 1980a), possess two main limitations. First, the models assume dyadic independence in the social network, which makes the analysis computationally convenient, but is untenable in most real-world contexts. Second, such models face restricted capability with parameter estimation and subsequent counterfactual analyses (Mayer 1984; Wasserman 1980b).

The above limitations were largely mitigated by the use of Monte Carlo Markov Chain (MCMC) based stochastic simulation models for sociometric data, as proposed by Snijders (1996) and later extended empirically in de Bunt et al. (1999). However, these models dealt with the issue of network evolution without focusing on any associated behavior. In a later work, Snijders et al. (2007) extended this actor-driven network evolution model to explicitly describe interrelationships between the network and the user's behaviors, and applied it to smoking and alcohol consumption behaviors. This new framework analyzed the network and set of user behaviors together in a joint state space and modeled how the network and behaviors evolved by influencing each other. The model accounted for network dependence of users and allowed researchers to investigate any number of alternative mechanisms of peer effects. A number of recent studies have used this co-evolution model to investigate the effects of selection and influence on social behaviors such as substance abuse among friends (Steglich et al. 2010), diffusion of innovation (Greenan 2015) as well as the evolution of self-reported music and movie tastes among adolescents (Lewis et al. 2012).

In this study, I develop a co-evolution model for large-scale observational data on the network and posting-behavior of SNS users. The model sets out to investigate how peer effects influence dynamic content production (e.g. posting public content) on the SNS. Unlike previous studies that have investigated co-evolution of network and behavior in an offline context with self-reported network and behavior data, the current research uses objective network and public content posting behavior data from one of the largest SNS in the world. Moreover, while previous studies have predominantly focused on stable behaviors like smoking and alcoholism, which do not change frequently over time, the current study focuses on dynamic behaviors, like content production which has a higher frequency of change. Furthermore, I extend the previous methods to model non-binary behaviors by discretizing online posting behavior based on several quantiles of intensity (e.g. ranging from levels 1 to 10). I posit that understanding the evolution of such online behavior is valuable to platform owners, marketers, and advertisers. However, due to the fast-changing nature of the behavior, it has been increasingly difficult for existing discrete-time models to accurately capture and predict these dynamics. In the following section, I describe the co-evolution model used in the current setting and illustrate how I perform parameter estimation of the network and behavior effects.

3.3 Empirical Analysis: A Co-evolution Model of Networks and Behavior

I develop an actor-based continuous-time model for the co-evolution of online network formation and content generation. The model builds upon and extends Snijders et al. (2007) and Steglich et al. (2010) in several key ways and is applied to a unique panel dataset obtained through collaboration with one of the largest online social networks in the world. This network-behavior co-evolution model draws upon past work on actor-oriented pure network evolution models (Snijders 2001).

3.3.1 Model specification

We observe a network with N users, for a total of T months, and model two main variables, namely, the state of the time-varying friendship network, a $N \times N$ matrix A_t and a $N \times 1$ time-varying integer-valued posting behavior vector P_t, which denotes the number of public posts contributed by users at time t.

3.3.1.1 Timing of decision

I assume that the evolution of both the network as well as the behavior follows a first-order Markov process, using very small time-increments, called “micro-steps” that occur at random instants in time. The network evolves in continuous-time but is observed at discrete moments. At a given micro-step, I constrain the network or the behavior to only allow a unit change, i.e., a tie forms or dissolves, or the posting volume increases or decreases by 1 unit. Using a Poisson process, I model these specific points in time when any given user i gets the opportunity to make a decision to change the vector of her outgoing tie variables $a_{ij} = [A]_{ij}$, $j = 1, \dots N-1$, or her behavior variable $p_i = [P]_i$.

Consequently, the rates at which the users make network decisions ($\lambda_i^{[A]}$) and behavioral decisions ($\lambda_i^{[P]}$) between time periods t and $t+1$are decided by rate functions as described in eq. (1) and (2) below.

$$\lambda_i^{[A]}(A_t, P_t) = \rho_m^{[A]} exp^{(h_i^{[A]}(\alpha^{[A]}, A_{t,} P_t))} \quad (network\ decisions) \tag{1}$$

$$\lambda_i^{[P]}(A_t, P_t) = \rho_m^{[P]} exp^{(h_i^{[P]}(\alpha^{[P]}, A_{t,} P_t))} \ (behavioral\ decisions), \tag{2}$$

where, the parameters $\rho_m^{[A]}$ and $\rho_m^{[P]}$ are dependent on the observed discrete time-period and capture periodic variations in either network or posting behavior[18], and the functions $h_i^{[A]}(.)$ and $h_i^{[P]}(.)$ model dependence on the current state of the network and the posting behavior. The exact functional forms of $h_i^{[A]}(.)$ and $h_i^{[P]}(.)$ depend on the network and behavioral effects that we choose to model in our context, and we specify these in detail in Sec. 3.3.2. However, in the current model specification, I assume that the rate functions are constant across the actors and are only dependent on the specific discrete observation periods *m*.

3.3.1.2 Objective function

While the rate functions model the timing of the users' decisions (i.e. to change network or behavior), the objective functions model the specific changes that are made. User *i* optimizes an objective function in the current time period over the set of feasible micro-steps she can take. This objective function is composed of three parts (Steglich et al. 2010): the evaluation functions $f_i^{[A]}$ and $f_i^{[P]}$, the endowment functions $g_i^{[A]}$ and $g_i^{[P]}$, and random disturbances $\epsilon_i^{[A]}$ and $\epsilon_i^{[P]}$, capturing residual noise.

The evaluation functions are parameterized by the vectors $\beta^{[A]}$ and $\beta^{[P]}$; the endowment functions are parameterized by the vectors $\gamma^{[A]}$ and $\gamma^{[P]}$, as shown in (3) and (4) below.

$$f_i^{[A]}(\beta^{[A]}, A_t, P_t) + g_i^{[A]}(\gamma^{[A]}, A_t, P_t | A_{t-1}, P_{t-1}) + \epsilon_i^{[A]}(A_t, P_t) \quad (network\ decisions) \tag{3}$$

$$f_i^{[P]}(\beta^{[P]}, A_t, P_t) + g_i^{[P]}(\gamma^{[P]}, A_t, P_t | A_{t-1}, P_{t-1}) + \epsilon_i^{[P]}(A_t, P_t) \quad (behavioral\ decisions) \tag{4}$$

The evaluation functions capture the utility obtained by a user i from her

---

[18] Estimating the rate functions $\lambda_i^{[A]}(.)$ and $\lambda_i^{[P]}(.)$ is similar to computing the ratio of network and behavior changes respectively in period m, to the total number of network and behavior changes across all m. However, the reason the parameters $\rho_m^{[A]}$ and $\rho_m^{[P]}$ are estimated from data and not just computed as a ratio is because this ignores the possibility of the actor not changing her network/behavior (or even reverting it). Consequently, the estimated rate functions will always be higher than the observed average number of changes.

network-behavior configuration. The functions $f_i^{[A]}(.)$ and $f_i^{[P]}(.)$ in (3) and (4) provide a measure of fitness of the state of the network and posting behavior, as perceived by the users. This implies that users constantly strive to make specific changes to their friendship network and posting behavior to maximize the value of this evaluation function.

The endowment functions $g_i^{[A]}(.)$ and $g_i^{[P]}(.)$, from (3) and (4) above, capture the part of utility that is lost when either the network ties or the posting behavior is changed by a single unit, but which was obtained without any "cost" when this unit was gained earlier. In other words, such endowment functions are useful to model situations where the creation and dissolution of ties, or an increase or decrease in posting behavior are asymmetric in terms of utility gained or lost. However, since in the context of this study, we do not model deletion of friends on the platform or the deletion of content, I do not include such endowment functions in my model.

3.3.1.3 Choice probabilities and intensity matrix

The final term in the objective function described in (3) and (4) above are the set of random and i.i.d. residuals $\epsilon_i^{[A]}$ and $\epsilon_i^{[P]}$. As is the case with random utility models, if we assume that these residuals follow type-1 extreme value distribution, it allows us to write the resulting choice probabilities for the network and posting micro-step decisions as a multinomial logit (Maddala 1986). For the network micro-step decision, the resulting choice probability is illustrated in (5) below.

$$Pr\left(a_{t+1} = a_t + \delta \middle| a_t, p_t, \beta^{[A]}\right) = \frac{\exp(f_i^{[A]}(\beta^{[A]}, a_{t+1} = a_t + \delta, p_t))}{\sum_{\varphi} \exp(f_i^{[A]}(\beta^{[A]}, a_{t+1} = a_t + \varphi, p_t))}, \qquad (5)$$

where, a_{t+1} is the resulting network at $t+1$ when a user i at micro-step t

alters the value of her tie variables by δ (or φ) where, $\delta, \varphi \in \{0,1\}$, i.e., user i either creates a new tie or makes no change to her network[19]. Similarly, for the posting micro-step decision, the resulting choice probability is illustrated in (6) below.

$$Pr\left(p_{t+1} = p_t + \delta \middle| a_t, p_t, \beta^{[P]}\right) = \frac{\exp(f_i^{[P]}(\beta^{[P]}, a_t, p_{t+1} = p_t + \delta))}{\sum_{\varphi} \exp(f_i^{[P]}(\beta^{[P]}, a_t, p_{t+1} = p_t + \varphi))}, \tag{6}$$

where, p_{t+1} denotes the resulting state of posting behavior in *t+1* when user i changes her posting volume at micro-step t by a factor of δ (or φ), where, $\delta, \varphi \in \{-1, 0,1\}$ i.e. the user increases her positing volume by 1 unit, decreases it by 1 unit or makes no new posts.

Once I have formulated the choice probabilities, the subsequent transition matrix Q, also called as the intensity matrix, models the transition from state (a_t, p_t) at micro-step *t* to a new state (a_{t+1}, p_{t+1}) at micro-step *t+1*, and can be specified by the following entries.

$$Q(a_{t+1}, p_{t+1}) \tag{7}$$
$$= \begin{cases} \lambda_i^{[A]} Pr(a_{t+1} = a_t + \delta | a_t, p_t), \text{ if } (a_{t+1}, p_{t+1}) = (a_t(i,\delta), p_t); \\ \lambda_i^{[P]} Pr(p_{t+1} = p_t + \delta | a_t, p_t) \text{ if } (a_{t+1}, p_{t+1}) = (a_t, p_t(i,\delta)); \\ -\sum_i \left\{ \sum_{\delta \in \{-1.1\}} Q(a_t(i,\delta), p_t) + \sum_{\delta \in \{-1.1\}} Q(a_t, p_t(i,\delta)) \right\}, \text{if } (a_{t+1}, p_{t+1}) = (a_t, p_t); \text{ and} \\ 0, \text{ otherwise.} \end{cases}$$

3.3.1.4 Model estimation

Due to the complexity of explicitly computing the likelihood function, I employ the use of simulation-based estimators. Specifically, I use a Markov Chain Monte Carlo (MCMC) based Method-of-Moments (MoM) estimator to recover the parameters of these rate and evaluation functions. The MoM

[19]There is no observed case of friendship dissolution (i.e. 1 to 0) in my data context.

estimator for my data and the parameters is based on the set of network and behavior related statistics that are specified in the following section. The MCMC implementation of the MoM estimator uses a stochastic approximation algorithm that is a variant of the Robbins-Monro (1951) algorithm (Robbins and Monro 1951) as detailed in Appendix 1.

The following section describes the empirical context for testing 1) the proposed co-evolution model to investigate the presence of peer effects and 2) the dependence of these peer effects on the state of the posting behavior.

3.3.2 Model parameterization in context of SNS

In the current context, the functions $h_i^{[A]}(.)$, $h_i^{[P]}(.)$, $f_i^{[A]}(.)$ and $f_i^{[P]}(.)$ from (1), (2), (3) and (4) can be modeled as a weighted sum of various network characteristics (e.g. degree, transitivity, homophily based on user covariates etc.) and behavioral characteristics (e.g. behavior trends, similarity measure, effect of user covariates on behavior etc.). I denote the matrix of network and behavior statistics computed in each time period t by $\mathrm{S}_\mathrm{t}^{[\mathrm{A}]}$ and $\mathrm{S}_\mathrm{t}^{[\mathrm{P}]}$, which are $\mathrm{N} \times \mathrm{K}_1$ and $\mathrm{N} \times \mathrm{K}_2$ matrices of K_1 network and K_2 behavioral characteristics, respectively. The functions $\mathrm{h}_\mathrm{i}^{[\mathrm{A}]}(.)$ and $\mathrm{h}_\mathrm{i}^{[\mathrm{P}]}(.)$ from the rate functions are specified as follows.

$$h_i^{[A]}\left(\alpha^{[A]}, A_t, P_t\right) = \sum_q \alpha_q^{[A]}\, s_{iqt}^{[A]}(A, P) \tag{8}$$

$$h_i^{[P]}\left(\alpha^{[P]}, A_t, P_t\right) = \sum_r \alpha_r^{[P]}\, s_{irt}^{[P]}(A, P) \tag{9}$$

Here, α_q indicates dependence on the statistics $s_{iqt}^{[A]}(A,P)$, and q $\subset$ K_1. Similarly, coefficient α_r indicates dependence on the statistics $s_{irt}^{[P]}(A,P)$, and r $\subset$ K_2 , where $s_{iqt}^{[A]}(A,P)$) and $s_{irt}^{[P]}(A,P)$) are vectors of one-dimensional statistics defined for each user i, and used to capture the rate dependence on the user's network characteristics (e.g. out-degree) and behavioral characteristics (e.g. SNS tenure) respectively. For the current set of analyses, however, I hold both sets of rate functions to be constant across all actors, and model only the dependence on the time period i.e. parameters $\rho_m^{[A]}$ and $\rho_m^{[A]}$ in (1) and (2).

Similarly, the functions $f_i^{[A]}(.)$ and $f_i^{[P]}(.)$ can be specified follows.

$$f_i^{[A]}\left(\beta^{[A]}, A_t, P_t\right) = \sum_{k_1} \beta_{k_1}^{[A]} s_{ik_1}^{[A]}(A_t, P_t) \quad (network\ evaluations) \tag{10}$$

$$f_i^{[P]}\left(\beta^{[P]}, A_t, P_t\right) = \sum_{k_2} \beta_{k_2}^{[P]} s_{ik_2}^{[P]}(A_t, P_t) \quad (behavior\ evaluations), \tag{11}$$

where, $s_{ik_1}^{[A]} = [S^{[A]}]_{ik_1}$ is the $k_1{}^{th}$ network statistic of user i, and, similarly, $s_{ik_2}^{[P]} = [S^{[P]}]_{ik_2}$ is the $k_2{}^{th}$ behavioral statistic of user i.

I parameterize the objective function based on the current research context, that of online posting behavior among a student population on a large and popular SNS. Specifically, I seek to investigate the presence of homophilous friendship formation based on similarities in posting behavior, as well as the role of peer influence in regulating content generation over time. Furthermore, I also analyze the dependency of peer effects on the specific state of the posting behavior to investigate whether active content posters react differently to peer effects as compared to less active posters.

3.3.2.1 The presence of homophily and peer influence

In this section, I define and specify estimation statistics for both the network as well as the public content posting behavior effects which I model in this study.

Social network effects:

The network effects from $S_t^{[A]}$ that I model are the user *i*'s out-degree ($s_{i1t}^{[A]}$), the transitivity ($s_{i2t}^{[A]}$), homophily effects based on posting behavior ($s_{i3t}^{[A]}$), and homophily based on the covariates, gender ($s_{i4t}^{[A]}$), age($s_{i5t}^{[A]}$), and SNS tenure ($s_{i6t}^{[A]}$). I also include effects that model the influence of individual covariates i.e., gender ($\mathrm{Gender}_i^{[A]}$), age $\left(\mathrm{Age}_i^{[A]}\right)$ and social network site (SNS) tenure ($\mathrm{SNS\ Tenure}_i^{[A]}$), on the propensity to form new friends. The mathematical illustrations are provided in (12) through (17).

(i) Degree ($s_{i1t}^{[A]}$) and Transitivity ($s_{i2t}^{[A]}$)

$$s_{i1t}^{[A]}(a) = \sum_{j} a_{ijt} \tag{12}$$

$$s_{i2t}^{[A]}(a) = \sum_{j,h} a_{ijt} * a_{jht} * a_{iht} \tag{13}$$

(ii) Homophily based on posting behavior and covariates(gender, age, SNS tenure)

$$s_{i3t}^{[A]}(a, p) = a_{i^+t}^{-1} \sum\nolimits_j a_{ijt} \left(1 - \frac{|p_{it} - p_{jt}|}{R_{pt}}\right), \tag{14}$$

where R_{pt} is the range of the posting variable *P* at step *t*. Variable $s_{i3t}^{[A]}$ represents the effect of homophily, based on posting behavior, such that $s_{i3t}^{[A]}$

takes a higher value for those users whose posting volume is closer to that of their peers (i.e. the value of $|p_{it} - p_{jt}|$ is small). Thus, a drive towards a higher value of $\mathrm{s}_{\mathrm{i3t}}^{[\mathrm{A}]}$ can be seen as an increased propensity towards creating homophilous friendships based on similarity in posting behavior.

For covariates $X = \{\mathrm{gender}, \mathrm{age}, \mathrm{SNStenure}\}$, we have similar expressions for s_{i4t}, s_{i5t} and s_{i6t} respectively.

(iii)Covariate (X_j) on the Degree effects (i.e., effect of user's gender, age, and SNS tenure on her Degree)

$$\mathrm{Gender}_{\mathrm{it}}^{[\mathrm{A}]}(\mathrm{a},\mathrm{x}) = \sum_{\mathrm{j}} \mathrm{a}_{\mathrm{ijt}} * \mathrm{x}_{1\mathrm{i}} \tag{15}$$

$$\mathrm{Age}_{\mathrm{it}}^{[\mathrm{A}]}(\mathrm{a},\mathrm{x}) = \sum_{\mathrm{j}} \mathrm{a}_{\mathrm{ijt}} * \mathrm{x}_{2\mathrm{it}} \tag{16}$$

$$\mathrm{SNSTenure}_{\mathrm{it}}^{[\mathrm{A}]}(\mathrm{a},\mathrm{x}) = \sum_{\mathrm{j}} \mathrm{a}_{\mathrm{ijt}} * \mathrm{x}_{3\mathrm{it}} \tag{17}$$

$\mathrm{Gender}_{\mathrm{it}}^{[\mathrm{A}]}$ represents the effect of the user i's gender (x_1) on her propensity to make new friends during step *t*, such that a positive and significant estimate on the statistic would imply that females *(Gender = 1)* make more friends than males *(Gender = 0)*, and vice versa. We have similar expressions for $\mathrm{Age}_{\mathrm{it}}^{[\mathrm{A}]}$ and $\mathrm{SNS\ Tenure}_{\mathrm{it}}^{[\mathrm{A}]}$ respectively. In all the above equations, $a_{\mathrm{ijt}} = 1$ if a tie exists between *i* and *j* in step *t*, and 0 otherwise.

Content posting behavior effects:

Next, I specify the rate and evaluation functions as defined for the content generation/posting behavior. In (11), the behavior effects that I model are the user's behavior tendency effect ($\mathrm{s}_{\mathrm{i1t}}^{[\mathrm{P}]}$), the peer influence effect i.e.

social influence $s_{i2t}^{[P]}$, and effects that capture the influence of individual covariates like gender ($\mathrm{Gender}_{it}^{[P]}$), age ($\mathrm{Age}_{it}^{[P]}$) and SNS tenure ($\mathrm{SNS\ tenure}_{it}^{[P]}$) on the posting behavior, *P*. I provide the mathematical illustrations in (18) through (22).

(i) Behavioral tendency effect (This captures the natural tendency of users to increase or decrease behavior over time)

$$s_{i1t}^{[P]}(a, p) = p_{it} \tag{18}$$

(ii) Peer influence effect (The propensity of users to assimilate in behavior towards their peers)

$$s_{i2t}^{[P]}(a, p) = a_{i+t}^{-1} \sum_j a_{ijt} \left(1 - \frac{|p_{it} - p_{jt}|}{R_{pt}}\right), \tag{19}$$

where, R_{pt} is the range of the posting variable P. $s_{i2t}^{[P]}$ represents the effect of peer influence, based on posting behavior, such that $s_{i2t}^{[P]}$ would have a higher value for those users whose posting volume is closer to that of their peers (i.e. the value of $|p_{it} - p_{jt}|$ is smaller). Thus, a positive and significant estimate on this statistic would indicate that users regulate their posting behavior to assimilate with their peers i.e. matching the posting rate of peers, and vice versa.

(iii) Influence of covariates (i.e. gender, age, SNS tenure) on behavior

$$\mathrm{Gender}_{it}^{[P]}(p, x) = p_{it} * x_{1i} \tag{20}$$

$$\mathrm{Age}_{it}^{[P]}(p, x) = p_{it} * x_{2it} \tag{21}$$

$$\text{SNS tenure}_{\text{it}}^{[P]}(p, \text{x}) = p_{it} * \text{x}_{3it} \tag{22}$$

Here, $\text{Gender}_{\text{it}}^{[P]}$ represents the effect of gender (x_{i1t}) on posting behavior such that a significant and positive estimate on this statistic would indicate that females *(Gender = 1)* post more than males *(Gender = 0)*. (21) and (22) denote similar expressions that represent the effects of age and SNS tenure on posting behavior respectively.

It is clear from the above formulation of effects, that the mathematical illustration for the network and behavior effects to compute homophily (14) and peer influence (19) are identical. This point lies at the core of the problem that is separating the effect of homophilous selection from peer influence. However, I exploit the longitudinal nature of the dataset to successfully identify temporal sequentiality across the periods. In other words, I use dyads of users who first become friends and then converge in behavior, to identify influence. Similarly, I use dyads of users who show similarity in behavior before becoming friends, to identify homophily. While there might be other latent confounds that I do not capture in this modeling, my approach makes an attempt at demonstrating a restricted form of causality. This view is consistent with several recent studies investigating related topics on homophily and influence among student populations (Lewis et al. 2012; Steglich et al. 2010).

3.3.2.2 Behavioral dependency of homophily and peer influence

While homophilous or assortative relationships among individuals have been reported extensively in previous research on the subject (Aral et al. 2009; McPherson et al. 2001; Park and Barabási 2007), what remains to be investigated is whether such homophilous selection effects vary in strength depending on the current state of the observable attribute or behavior. For instance, consider how an individual who smokes cigarettes is more likely to make friends with a fellow smoker (Christakis and Fowler 2008; Pearson and West 2003). However, would his affinity to make friends with a similar smoker be any higher or lower depending on how many cigarettes he smokes each day at the present moment? An analogous problem arises in studying influence. It has been widely observed that peer influence plays an important role in the onset and sustenance of various addictive behavior, including smoking (Christakis and Fowler 2008; Ennett and Bauman 1994). However, little is known about whether such peer influence effects are particularly stronger or weaker for different levels of the behavior itself.

In the current study, I investigate whether SNS users show varying strengths of selection bias due to homophily and susceptibility to peer influence depending on their current levels of posting behavior. To achieve this, I cluster all users depending on their levels of posting-behavior into three major categories. Based on the volume of content generated, I categorize the top 10 percentile of individuals in each time period as Most Active Posters (MAP), and categorize the bottom 10 percentile of individuals as Least Active Posters (LAP). All other users are categorized as Moderately Active Posters (MoAP). I introduce dummy variables for each of the first two groups in my model, keeping the middle group as the baseline. This is shown in (23) and (24) below. The estimates from the interaction between these dummy variables and the homophily and peer influence variables would help me address the question at hand.

$$\mathrm{s}_{\mathrm{i7t}}^{[\mathrm{A}]}(\mathrm{a},\mathrm{p}) = \mathrm{MAP}_{\mathrm{i}} * a_{i+t}^{-1} \sum_{\mathrm{j}} \mathrm{a}_{\mathrm{ijt}} \left(1 - \frac{|p_{it} - p_{jt}|}{R_{pt}}\right), \text{ and} \tag{23}$$

$$s_{i8t}^{[A]}(a,p) = LAP_i * a_{i+t}^{-1} \sum_j a_{ijt} \left(1 - \frac{|p_{it} - p_{jt}|}{R_{pt}}\right), \tag{24}$$

where, R_{pt} is the range of the variable P. In the above equations, the MAP_i and LAP_i dummy variables denote whether a user *i* is a heavy poster or low poster. The middle group ($MoAP_i$) is held as the baseline group for comparison of estimates. Similar effects are constructed for the interaction of these activity dummies and the behavioral homophily effect ($s_{i3t}^{[P]}$ and $s_{i4t}^{[P]}$).

3.4 Data Context

I obtained complete online network data through collaboration with a large online social network site (SNS) for 2507 undergraduate students attending a North American university for the months from September 2008 till February 2009. Additionally, I recorded the number of monthly public posts made by these users on the social media platform during the same period. The descriptive statistics of the key variables are illustrated in Table 1 in the following page.

P_{it} and D_{it} constitute the key variables for the co-evolution model, depicting the total number of monthly public posts and new friends added on the SNS respectively. For purpose of estimation, I perform a quantile split on P_{it} to categorize the posting variable into 10 levels (with 0 being the lowest posting rate and 10 being the highest posting rate). The covariates include $Gender_i$, the gender of the user (0,1 representing male and female respectively, and 2-4 representing rare gender types), Age_i, the biological age of the user, and $SNS\ tenure_i$, the total number of days spent by the user on the SNS at the time of recording the data.

Table 3-1: Descriptive summary of model variables

	Min	**Max**	**Mean**	**Std. Dev.**
Dependent Variable:				
Total Monthly Public Posts (P_{it})	0.000	15200.360	145.148	492.811
Independent Variables:				
Biological Age (Age_i) [years]	20.000	26.000	22.244	1.381
SNS Tenure, ($SNS\ tenure_i$) [days]	831.000	2591.000	1778.376	344.891
Gender ($Gender_i$)	0.000	4.000	1.471	0.543
Number of friends added on SNS over the 6 months (D_i)	1.000	98.000	6.270	7.055
Total Monthly Public Posts by Friends ($\sum_j P_{jt-1}$)	0.000	1293232.000	47006.730	79376.770

The network and behavior descriptive summaries are detailed in Appendices 2 and 3. Within our observation period, the students produced a substantial amount of content on the social media platform, and also established several new friendships. This provides us with sufficient variability in our data to test my proposed models.

3.5 Results

3.5.1 The evolution of homophily and peer influence

I estimate the rate and evaluation functions from the co-evolution model as specified in Sections 3.3.1 and 3.3.2.1 earlier using a Method of Moments (MoM) estimator and present the results in Table 3-2. The MoM estimator essentially tries to recover parameter estimates by matching the observed network data with the simulated network data. Appendices 4 and 5 provide details on the convergence descriptives for these simulations. Specifically, I provide information about the deviation of the simulated network and behavioral statistics from the observed data. Tables 3-2(a) and 3-2(b) highlight the estimation results for rate parameters $\rho_m^{[A]}$ and $\rho_m^{[P]}$, for a total of 5 months (i.e. one less than the total number of time periods since the first among six periods is conditioned upon during the estimation), and estimates for $\beta_p^{[A]}$ where p ranges from 1 to 9, and for $\beta_q^{[P]}$ where q ranges from 1 to 5.

3.5.1.1 Results on networks

For the network structure variables, as shown from the results in Tables 3-2(a) and 3-2(b), I observe that the estimate for the out-degree of the users is significantly negative (-9.536; p<0.01). Since, the evaluation function can be thought of as a measure of the “fitness” or “attractiveness” of the state of the network, this estimate indicates that users in our network show a lower propensity over time to establish new social connections. This can be attributed to the cost of forming social connections or constrained resources (Dunbar 1992; Phan and Airoldi 2015). Further, I observe that the estimate for network transitivity is positive (0.109; p<0.01).

Table 3-2 (a, b): Estimation results for network and behavior effects

Network Parameters	Estimate
Friendship rate (Period 1)	7.767*** (0.100)
Friendship rate (Period 2)	6.267*** (0.088)
Friendship rate (Period 3)	3.930*** (0.082)
Friendship rate (Period 4)	4.547*** (0.075)
Friendship rate (Period 5)	5.353*** (0.084)
Out-Degree	-9.536*** (0.011)
Transitivity	0.109*** (0.001)
Gender homophily	0.068 (0.057)
Gender on Degree	0.031 (0.020)
Agehomophily	0.024 (0.034)
Age on degree	0.011 (0.009)
Tenure homophily	0.003 (0.022)
Tenure on degree	-0.032** (0.016)
Posting homophily	**0.127*** (0.034)**

Behavior Parameters	Estimate*
Posting rate (Period 1)	4.337*** (0.145)
Posting rate (Period 2)	3.889*** (0.143)
Posting rate (Period 3)	5.229*** (0.205)
Posting rate (Period 4)	4.995*** (0.195)
Posting rate (Period 5)	3.681*** (0.119)
Posting Tendency (Linear Shape)	-0.196*** (0.007)
Influence	**-2.995*** (0.134)**
Gender on Posting	0.007 (0.012)
Age on Posting	0.003 (0.006)
Tenure on Posting	0.003 (0.010)

*** <0.01, ** <0.05, *<0.1

This indicates that there is an increased drive towards network closure in our observed network. For instance, if users i and j are friends, and users j and h are friends as well, then the user i has a stronger motivation to befriend user h over any other user in the network, as this increases the overall attractiveness of the new network state for i. I also find strong evidence for friendship formation among those with a similar level of posting behavior (0.127; $p<0.01$). Thus, the more active posters prefer to befriend other active posters, while the less active posters prefer other less active ones. Interestingly, none of the other covariates were found to contribute to homophilous friendship formation.

3.5.1.2 Results on behavior

Among the behavior variables, I observe that the estimate for the linear tendency parameter is significantly negative (-0.196; $p<0.01$). As mentioned earlier, the tendency effect represents a drive towards high posting volume. A zero value on this parameter indicates user's preference for the average posting volume. Since I obtain a negative estimate on this parameter, it indicates that as time goes by, users prefer to post less. I also find strong evidence of peer influence among the students, with a significantly negative parameter for the influence effect (-2.995; $p<0.01$). This implies that individuals tend to correct their posting behavior over time in a direction away from their peers. This could be a result of free-riding behavior in case the peers are contributing more, or could also be representative of an increased drive to behave in a non-conformist manner (e.g. "*If everyone else is posting more, I should do something different*").

While it is hard to uncover the specific reasons for the peer effects I find, interpreting the parameters for homophily and peer influence together leads to an increased understanding of the interplay between friendship formation and content production behavior in online networks. Taken together, the two parameters suggest that while students prefer to befriend other students who are similar to themselves in posting behavior, they tend to move apart over time after becoming friends. Thus, behavioral similarity could play the role of a facilitator during the early days of friendship formation, but act as a deterrent in the longer run. I contend that this insight is not only theoretically important to uncover but has very strong practical implications as well, which I shall discuss in Section 3.6.

3.5.2 Results on behavioral dependency of homophily and peer influence

In addition to the above, I also find strong evidence for the behavioral dependency of homophily and peer influence. Tables 3-3(a) and 3-3(b) illustrate the estimation results for rate parameters $\rho_m^{[A]}$ and $\rho_m^{[P]}$, for periods 2 to 6 (i.e. the first among six periods is conditioned upon during the estimation), and estimates for $\beta_p^{[A]}$ where p ranges from 1 to 11 and for $\beta_q^{[P]}$ where q ranges from 1 to 7. The results from the estimation show that the users in my sample demonstrate varying propensities to create homophilous relationships and varying susceptibility to peer influence, depending on the current state of their posting behavior. Specifically, compared to moderately active posters (MoAP), most active posters (MAP) were less likely to form friendships with other MAPs (-0.375; p<0.01), while least active posters (LAP) were more likely to form friendships with other LAPs (0.293; p<0.01). Further, compared to MoAPs, both MAPs and LAPs were found to be more susceptible to peer influence. However, while MAPs showed positive influence (i.e. converge in behavior with peers) (1.183; p<0.01), the LAPs showed negative influence (i.e. diverge in behavior from peers) (-6.437; p <0.01).

3.5.3 Comparative analysis of alternative modeling approaches

In this section, I present results from two baseline approaches. The first baseline approach models online content production using a fixed effect panel linear regression model and a fixed effect Poisson regression model, as illustrated by the model specifications (25) and (26) below.

Table 3-3 (a, b): Estimation results with behavioral dependency

Network Parameters	Estimate	Behavior Parameters	Estimate*
Friendship rate (Period 1)	7.750*** (0.100)	Posting rate (Period 1)	4.948*** (0.274)
Friendship rate (Period 2)	6.368*** (0.089)	Posting rate (Period 2)	4.137*** (0.166)
Friendship rate (Period 3)	3.997*** (0.075)	Posting rate (Period 3)	5.477*** (0.317)
Friendship rate (Period 4)	4.543*** (0.075)	Posting rate (Period 4)	5.879*** (0.255)
Friendship rate (Period 5)	5.431*** (0.084)	Posting rate (Period 5)	4.583*** (0.252)
Out-Degree	-9.562*** (0.011)	Posting Tendency (Linear Shape)	-0.182*** (0.011)
Transitivity	0.107*** (0.001)	**Influence**	**-2.951*** (0.268)**
Gender homophily	0.076 (0.058)	**High Posters Influence**	**1.183*** (0.096)**
Gender on Degree	0.023 (0.019)	**Low Posters Influence**	**-6.437*** (0.763)**
Year of birth homophily	0.023 (0.034)	Gender on Posting	0.009 (0.012)
Year of birth on degree	0.013 (0.009)	Age on Posting	0.004 (0.005)
Tenure homophily	-0.001 (0.022)	Tenure on Posting	0.004 (0.010)
Tenure on degree	-0.019 (0.016)	*** <0.01 ** <0.05 *<0.1	
Posting homophily	**0.112*** (0.042)**		
High Posters Homophily	**- 0.375*** (0.021)**		
Low Posters Homophily	**0.293*** (0.023)**		

Such aggregated personal networks have been commonly used in previous studies where an individual's social network is collapsed to a fixed number of sociometric variables, like the centrality measures (Kirke 2004; Yoganarasimhan 2012). These measures are then used as regressors, together with individual-level attributes, in a linear model to explain outcomes of individual-level behavior. This approach, however, ignores both homophilous friendship formation, as well as the continuous-time evolution of the network itself. The sociometric variable included is the out-degree D_{it-1}, which denotes the total number of friends added by the user i in time period *t-1* on the SNS

$$p_{it} = \gamma Z_{it-1} + \kappa_i + \tau_t + \epsilon_{it} \tag{25}$$

for ordinary least square linear regression, or,

$$\log p_{it} = \beta Z_{it-1} + \kappa_i + \tau_t + \epsilon_{it}$$

for Poisson regression, where,

$$\beta = \{\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$$

$$Z_{it-1} = \{1, D_{it-1}, \sum P_{jt-1}, Age_i, Gender_i, SNS\ tenure_i\}' \tag{26}$$

$\kappa_i = \{\kappa_1, \kappa_2, \dots, \kappa_n\}$, for *n* individuals,

$\tau_t = \{\tau_1, \tau_2, \dots, \tau_t\}$, for a total of *t* months,

$\epsilon_{it} = \{\epsilon_{11}, \epsilon_{12}, \dots, \epsilon_{nt}\}$, and

$\gamma = \{\gamma_0, \gamma_1, \dots, \gamma_5\}$

In the above model specifications, the coefficient β_2 provides an estimate of peer influence based on the posting behavior of the peers of a user i. I also control for the user i's biological age (Age_i), gender ($Gender_i$), as well as social network age ($SNS\ tenure_i$), which is the number of days spent by the user on the SNS at the time of recording the data. The descriptive statistics for the variables were provided earlier in Table 3-1.

The estimation results are illustrated in Table 3-4 below. The results show that using a discrete-time aggregated network approach such as this leads us to believe that the peer's posting behavior has a weakly positive effect on the individual’s posting behavior in the subsequent time period, after controlling for other covariates. However, as mentioned earlier, this method ignores any selection bias in friendship formation caused due to homophily and thus provides biased estimates of peer influence. The results from the co-evolution model from the previous section show that the effect of peer influence is actually the reverse (i.e. significantly negative), once I factor in homophilous friend selection into the model.

A second baseline model specifies and estimates the co-evolution of the network and behavior, but ignores both homophily based on dynamic content production, and the effect of peer's content posting behavior on the individual. Thus, I estimated a model that relies only on homophily based on stable attributes like age, gender etc., and ignores any role played by dynamic behaviors like content postings. The result from this model is illustrated in Table 3-5. The results from this model are consistent with the earlier results and reaffirm the belief that the students in my sample are not establishing friendships based on similarities in age, gender, or SNS tenure. Rather, they are forming new ties based on similarities in content posting behavior. Moreover, the results from this model prove that the students' content posting behavior is not influenced by their personal attributes such as age, gender or SNS tenure, but are instead influenced largely by the posting behavior of their peers, as was illustrated in Tables 3-2 and 3-3.

Table 3-4: Results from discrete-time aggregated network models

Variables	**(Random-effects Panel Linear Regression) Posts (P_{it})**	**(Fixed-effects Panel Linear Regression) Posts (P_{it})**	**(Fixed-effects Poisson Regression) Posts (P_{it})**
D_{it-1}		-6.381*** (0.814)	-0.007*** (0.0001)
$(\sum_j P_{jt-1})$		0.0001** (0.0001)	0.0000002*** (0.00000001)
Age_i	-77.831*** (8.245)	(omitted)	-0.789*** (0.043)
$SNS\ tenure_i$	0.150*** (0.033)	(omitted)	0.001*** (0.0002)
$Gender_i$ (=1)	15.072 (58.176)	(omitted)	-0.076 (0.289)
$Gender_i$ (=2)	-9.263 (58.144)	(omitted)	-0.196 (0.288)
$Gender_i$ (=4)	-120.111 (368.536)	(omitted)	-22.485 (3763.516)
Time dummies	Present	Present	Present
Sample size	2030	2012	2012
R-squared	0.040	0.052	

*** <0.01, ** <0.05, *<0.1

Table 3-5: Estimation results for network and behavior based on covariates alone

Network Parameters	Estimate	Behavior Parameters	Estimate
Friendship rate (Period 1)	7.558*** (0.098)	Posting rate (Period 1)	3.064*** (0.100)
Friendship rate (Period 2)	6.226*** (0.097)	Posting rate (Period 2)	2.804*** (0.114)
Friendship rate (Period 3)	3.977*** (0.076)	Posting rate (Period 3)	3.694*** (0.149)
Friendship rate (Period 4)	4.727*** (0.080)	Posting rate (Period 4)	3.510*** (0.117)
Friendship rate (Period 5)	5.608*** (0.087)	Posting rate (Period 5)	2.679*** (0.100)
Out-Degree	-9.070*** (0.010)	Posting Tendency (Linear Shape)	-0.116*** (0.008)
Transitivity	0.056*** (0.001)	Gender on Posting	-0.003 (0.013)
Gender homophily	0.007 (0.056)	Age on Posting	0.006 (0.012)
Gender on Degree	0.007 (0.019)	Tenure on Posting	-0.002 (0.011)
Year of birth homophily	-0.006 (0.023)		
Year of birth on degree	0.006 (0.016)		
Tenure homophily	-0.007 (0.021)		
Tenure on degree	-0.002 (0.016)		

*** <0.01, ** <0.05, *<0.1

3.5.4 Sensitivity to latent homophily

While our analysis conditions on observable behavioral (e.g. posting) and individual-level covariates, (e.g. age and gender), there is a possibility that the network formation might be driven by homophily based on latent factors, such as personality traits and similarity in tastes or preferences. The presence of such latent homophily has been cited as an important confound in the estimation of social influence (Shalizi and Thomas 2011). We look to test the sensitivity of our modeling approach to the presence of such latent homophily using a latent space modeling approach, similar to what has been described in (Davin et al. 2014). Latent space models are well known in social networks literature and have been traditionally employed in identifying and visualizing communities within networks. For our analysis, we use 2-dimensional latent space positions as proxy variables to control for potential latent homophily. The intuition behind this approach is that if two actors are close to each other in a latent social space, then this similarity is driven by both observed as well as unobserved factors. Thus, adding latent space coordinates as model covariates would serve to reduce the bias associated with influence estimate by controlling for some latent homophily. There have been some prior work that have used latent space models to address similar questions in economics and marketing (Ansari et al. 2011; Braun and Bonfrer 2011). A summary of how the latent space models for our current context were specified and estimated has been illustrated in Appendix 6.

We estimate the rate and evaluation functions from the co-evolution model as specified in Sections 3.3 earlier, using the latent space positions as coordinates, and present the results in Table 3-6. We find that the results for both homophily based on posting behavior as well as peer influence are consistent with our previous results. As expected, after controlling for homophily based on latent space coordinates, the estimate for posting homophily (0.104; $p<0.01$) reduces in strength, but continues to be statistically significant. This shows that there does exist evidence of homophily based on latent factors beyond the observable factors of age, gender and SNS tenure. However, our proposed effect of posting homophily exists even after controlling for possible latent confounders. Similarly, the estimate for peer influence is weaker (-0.015; $p<0.01$) than our earlier models that do not account for latent homophily. In summary, we leverage latent space positions of actors in our network to account for possible latent homophily, and show that our results for homophily and peer-influence based on posting behavior are valid even after controlling for these latent positions.

Table 3-6. Latent homophily corrected results for network and behavior

Network	Estimate	Behavior	Esti
Friendship rate (Period	7.529*** (0.098)	Posting rate (Period	3.496*** (0.182)
Friendship rate (Period	6.204*** (0.141)	Posting rate (Period	3.590*** (0.132)
Friendship rate (Period	3.949*** (0.075)	Posting rate (Period	4.223*** (0.114)
Friendship rate (Period	4.586*** (0.085)	Posting rate (Period	4.603*** (0.122)
Friendship rate (Period	5.487*** (0.091)	Posting rate (Period	3.104*** (0.104)
Out-Degree	-9.913*** (0.014)	Posting Tendency	-0.191*** (0.008)
Transitivity	0.098*** (0.001)	Influence	-0.015*** (0.001)
Gender homophily	0.048 (0.077)	Gender on Posting	0.007 (0.014)
Gender on Degree	0.011 (0.021)	Age on Posting	0.008 (0.012)
Age homophily	0.011 (0.032)	Tenure on Posting	-0.002 (0.012)
Age on degree	0.005 (0.024)		
Tenure homophily	-0.010 (0.030)		
Tenure on degree	-0.012 (0.018)		
Posting homophily	0.104*** (0.047)		
Latent Pos. (X) homophily	1.030*** (0.126)		
Latent Pos. (Y) homophily	1.110*** (0.112)		

*** <0.01,

3.6 Discussion

In the current study, I develop and estimate a model for analyzing the coevolution of content production and social network structure using real world data from a large social network site. My results demonstrate the role of social network structure and user-characteristics in influencing content production on SNS. I adopt an actor-driven and co-evolution based MCMC modeling approach to jointly estimate the evolution of the user's social network and behavior. I contend that this approach is more statistically disciplined than several previous methods, which tend to violate some key assumptions of network-based modeling. Furthermore, I depart from previous instances of the actor-driven models whose applicability is restricted to stable dichotomous behaviors, like smoking, and substance abuse. In the current study, I adapt the co-evolution model to a dynamic behavior (i.e. online public posts) which often changes rapidly over successive time periods. I avoid convergence related difficulties with MCMC estimations of such continuous behavioral variables by discretizing the behavioral variable into several quantiles to represent the intensity of behavior. I contend that by using this quantile-based binning strategy, I was able to achieve high convergence in estimations without much loss of information. The results from my analyses uncover important insights about how users make friends on SNS, and how the network, in turn, influences their content production behavior. Specifically, I show that users are more likely to make friends with users who show a similar level of posting behavior, as observed by the number of public posts. However, this homophilous behavior is short-lived and the users are found to diverge in their content production rates from their peers over time.

Furthermore, my analyses shows that the propensity to form friendships based on homophily, and the susceptibility to peer influence after forming the friendships, are dependent on the current state of the behavior. Thus, users who are very active contributors on SNSs show very different peer effects as compared to users who are less active on the SNS.

3.6.1 Theoretical implications

Using the co-evolution perspective, I address the following two theoretical gaps in the existing research on the evolution of online social networks and social behavior.

First, I show that homophilous peer selection and peer-influence might have varying strengths depending on the stage of network evolution. I find strong evidence of selection bias on the basis of homophily in content production, i.e., the students make friends with others who are similar in their content production behavior. Once they become friends, however, my findings show that they exhibit a negative influence effect. This means that the students actively try to distinguish themselves from their friends in terms of their content production behavior. This is an interesting phenomenon, which demonstrates that dynamic behaviors such as content production can influence network evolution in competing ways.

Second, I uncover a behavioral dependency of these network effects, such that homophilous selection and peer-influence increase or decrease in strength as a function of the current magnitude of an individual’s behavior. I find that students who are very active content producers (i.e. MAP users) are qualitatively very different from students who are highly inactive producers (i.e. LAP users) and students who are moderately active content producers (i.e. MoAP users). These three groups of students displayed different degrees of inclination towards homophilous peer selection and different degrees of susceptibility towards peer-influence. Taken together, these results reveal an interesting pattern of how online social networks co-evolve with the content produced on these platforms.

3.6.2 Practical contributions

Understanding the nature of peer effects on SNS has clear practical implications for several stakeholders. Firstly, and most importantly, I offer a framework within which online user contributions can be studied as a function of the underlying network. While it is common for researchers and practitioners to use predictive and explanatory models of social media content production, they often tend to ignore the underlying social network that connects the content producers. I offer a robust statistical model to help explain content production while being conscious of the evolution in the underlying network structure. This would help platform owners and marketers derive more reliable insights about their users.

Secondly, my results provide intelligence to marketers to identify and better target valuable users on SNS. Understanding what drives content production on online platforms, and the impact of peers on the user's propensity to produce content is key to devising better strategies to enable and sustain content production on the platform. Moreover, by understanding how friendships are created and altered over time, platforms like Facebook and Twitter can help improve friend recommendations and personalized content through customized "newsfeeds". Specifically, my results suggest that it might not be a good idea to recommend heavy content posters as friends to other heavy posters, as such friendships tend to be detrimental to the content production of either of the friends, i.e., high posters prefer other high posters in making friends, but reduce their posting rate over time after the friendship is created. Moreover, I also show that this tendency to alter behavior in response to peers is strongest for heavy posters and weakest for low posters. Thus, the findings from this study can guide platform owners on better managing their active content producers.

Thirdly, the model developed in this study also allows for predictive analysis of posting behavior on these platforms, such that managers and researchers can effectively seed content, and forecast the diffusion of this content through social networks. Such predictive models for user behavior on dynamic networks can be invaluable not just to the platform owners, but also to advertisers and third-party marketers who wish to leverage social media for their own businesses. Thus, I believe that the specific findings from this study and the methodology in general can increase content-creation and user retention in such SNS platforms.

3.7 Limitations and Conclusion

As an initial attempt to model and analyze the co-evolution of network structure and user behavior in online social networks, this study is prone to several limitations that offer opportunities for future research. Firstly, and as mentioned earlier, the current paper focuses on providing a statistically sound method to uncover the dynamic peer effects in a university social network. However, additional analyses are required to further separate out the specific rationale behind why individuals show such effects. Moreover, the absence of any data on the offline social interactions among users makes this process of interpretation particularly difficult. In future extensions of this work, I would seek to identify potential proxy variables to control for the offline interaction of the users in my sample. Secondly, the current modeling approach requires computational resources to simulate the networks in each stage of the estimation procedure. This might be a concern for extremely large networks of users, and networks with high sparsity. In such cases, I might have to resort to bootstrapping approaches which introduce concerns about network-based sampling, a non-trivial area of active research in its own right. My model imposes a standard Markovian assumption on the data, which is reasonable in most cases. However, this assumption implies that there are no latent confounding factors that might influence the social network or the user behavior. Even though I have controlled for common covariates that have been used in recent social network studies (e.g. age, gender and experience), I cannot completely rule out the possibility of unobserved confounds that might play a role. In addition, since the data on content generation comes mainly from public posts, we cannot observe the associated changes in private conversations which can also likely influence some of the findings of this study. Lastly, I consider all friendships to be bi-directional or symmetric ties. While this is not a limitation in the present study, it could be useful to identify the directionality of friendship i.e. separate out in-degree from out-degree. While in-degree can be considered to be a measure of popularity, out-degree provides a better indication of SNS activity. Thus, by separating out the two effects, we will be able to investigate more complex social constructs in future studies.

In the following section, I investigate the case of introduction of new privacy controls on a large SNS, to analyze its impact on the value created for the SNS provider, in the form or increased or reduced content generation.

Chapter 4. IMPACT OF NEW PRIVACY CONTROLS BY SNS PROVIDERS ON USERS' CONTENT GENERATION BEHAVIOR

4.1 Introduction

As mentioned in previous chapters, social network sites (SNS) have rapidly grown in popularity over the years, and offer a unique platform for its users to self-disclose and self-present, in an effort to communicate and manage their online self-identities better (boyd and Ellison 2007; Marwick and Boyd 2011). A recent research reports that 74% of all online adults use at least one SNS in their daily lives ("Social Networking Fact Sheet | Pew Research Center's Internet & American Life Project" 2014). Facebook, the largest SNS, hosts 1.39 billion monthly active users who generate an average of 293,000 status updates and 510 comments every minute (Noyes 2015). The content produced by individuals on these platforms offer a medium for them to effectively communicate and engage with their audiences (boyd and Ellison 2007). In absence of any face-to-face interaction, this user-generated-content (UGC) in the form of textual updates, uploaded photos and private communications offers audiences a window into the personality and attitudes of the content producer (Hogan 2010; Lazer et al. 2009; Marwick and Boyd 2011; Wilson et al. 2012). For the platform owners too, this offers an incredible opportunity to understand more about the tastes and preferences of their users, which helps them design better content delivery (Bakshy et al. 2015) and friend recommendation systems (Backstrom and Leskovec 2011; Backstrom et al. 2006). Further, UGC from these platforms are effectively exploited by brand owners and online marketers to identify profitable customers (K. Y. Goh et al.

2013; Lee et al. 2014), and to perform better targeting of product based ads. While there are plenty of social and economic value to sharing information on SNS as explained in Chapter 1, there are significant privacy concerns surrounding publicly available content (Goldfarb and Tucker 2012, 2011). Responding to the consumers' concerns over data privacy, governments including the United States, European Union, and in Asian countries have introduced increasingly strict policies and laws[20] to regulate corporations such as SNS owners who record and share user data.

Over the years, SNSs have developed several features on their platforms to enable their users to create more value by producing better content (e.g. Facebook reactions) and by forging new social ties (e.g. "friends you may know" feature on Facebook). They have also introduced a number of audience filtering features, aimed at providing users with enhanced control over the intended audience of their content (Cheng 2009; Sanghvi 2009). Platform owners claim that the introduction of such privacy controls is essentially in the best interest of the users in providing them with increased control over who gets to see their SNS profiles and public updates. Privacy activists, however, have voiced concerns that such feature changes might be part of a hidden agenda on part of the SNS to increase the amount of content produced on their platforms by increasing a sense of perceived trust among their user base (Bankston 2009; Kincaid 2009; Opsahl 2010). However, to date there has been no quantitative investigation into understanding how such privacy controls increase or decrease the disclosure behavior of users on the SNS. As discussed in earlier chapters, since such disclosures are a key source

[20] http://www.informationshield.com/intprivacylaws.html

of value proposition for the user herself, the SNS provider as well as external businesses, it becomes an important business imperative to analyze the implementation of such feature changes by the SNS provider.

In the current study, I question the role of privacy changes enacted by a large and popular SNS in fostering public content generation among its users. The SNS made a major change in its privacy policy in December 2009, when it announced to its users that it has revamped its privacy features to provide users better control over information they share on the site. This change made it possible for users to apply a greater range of privacy controls to determine access permissions for the audience for each post (Cheng 2009; Sanghvi 2009). Specifically, I examine the impact of the introduction of enhanced privacy controls by a large SNS on the public and private content generation patterns of its users. Public content generated on the SNS include content that are broadcasted by the SNS users and can generally be viewed by any registered user on the platform (e.g. public Tweets on Twitter or public status updates on Facebook). In contrast, private content include directed conversations that users hold with other users on the platform, and are only visible to the participants of the conversation (e.g. Facebook private messages, group chats etc.). The empirical analysis I perform is, however, complicated by the fact that unlike other social media platforms (e.g. blogs, e-commerce sites), SNSs are characterized by an underlying network of relationships. These relationships are often formed out of an increased sense of preferential attachment towards certain individuals, a term that is referred to as homophily in recent social networks literature (McPherson et al. 2001). As a result, the effect of the privacy intervention on the content production behavior of users is confounded by the presence of network effects i.e. we need to estimate the effect of the treatment after controlling for all changes that can be attributed to the accompanying change in the underlying friendship network. I resolve this empirical challenge by leveraging the network-behavior co-evolution network (Snijders et al. 2007), that I have proposed and explained in detail in Chapter 3.

Through collaboration with a popular SNS, I was able to obtain data on the volume of content produced and relationships made on the SNS by students of an American University, who have registered accounts on the SNS. The time span for my dataset spans several weeks before and after the privacy control introduction, allowing me to study both the change in content production as well as the change in friendship networks after the privacy controls were introduced by the SNS.

The results from the analysis show that the privacy intervention had no significant effect on the volume of public posts produced in the week following the intervention. However, and interestingly, the intervention had a significant and negative impact on the volume of private messages exchanged in this week. These results were found to be robust even after controlling for the underlying network change. Moreover, I compared and contrasted the results against a panel regression model to show that a regression model, which ignores the network dependence of users, gave inaccurate estimates and poorer model fit than my proposed co-evolution model. I provide additional evidence on the identification of the intervention effects, by exploiting a quasi-experimental setting. I show that the privacy intervention has a weaker effect on users who are more privacy conscious, than others.

The results from this study can potentially make significant contributions to the theory of how individuals manage their privacy calculus in the presence of policy regulations or feature changes. On the managerial front, the insights from the analyses can also help SNS providers better understand the net value created or diminished as a result of their interventions, and also to advertisers who often leverage privacy-sensitive data in targeting their customers (Goldfarb and Tucker 2011; Tucker 2014).

4.2 Background and Related Work

In this section, I review some related work on self-disclosure in online platforms, factors influencing an individual's privacy attitudes, and the effect of external privacy-related factors on an individual's behavior.

4.2.1 Individual's information privacy

Information privacy has been an active area of interest among Information Systems researchers and practitioners. I draw primarily from two extensive meta-analyses by Bélanger and Crossler (2011) and Smith et al. (2011) to discuss and guide my review of the current discourse in this popular domain. In their analysis, Smith et al. note that most of the previous research in information privacy spanning Economics, Marketing, Law, Philosophy and Information Systems disciplines have attempted to answer one of the following three questions about privacy: (i) *What is (and is not) privacy and how is it different from the notion of security?* (ii) *What is the relationship between privacy and other related constructs?* (iii) *To what extent does context matter in the relationship between privacy and other constructs?* (i.e. how generalizable are privacy related findings across industries and environments?).

Several studies have attempted to discuss the first from philosophical, psychological, sociological and legal perspectives, with limited consensus (Solove 2006; Westin 1968). This has led to a stark increase in several competing theoretical frameworks, with often conflicting empirical evidence (Bélanger and Crossler 2011; Siponen 2005). At the heart of these discussions on information privacy lies an ongoing debate between the idea of privacy as a general right (Bennett 2012; Rosen 2012; Warren and Brandeis 1890) and as a commodity (Campbell and Carlson 2002; Davies 1997; Laudon 1996).

In addition, a related stream of literature illustrates the idea of a *privacy calculus* by assuming that individuals face a trade-off between the costs and benefits of privacy disclosure, and that this trade-off is salient in guiding the user's behavior in privacy decisions (Chellappa and Sin 2005; Hui et al. 2006; Klopfer and Rubenstein 1977; Laufer and Wolfe 1977; Posner 1981; Stone and Stone 1990; Xu et al. 2009). In other words, an individual's decision to reveal personal information depends on the outcome of a rational cost-benefit analysis of disclosing this information (Dinev et al. 2006; Krasnova and Veltri 2010). More recent studies have pointed out that while higher privacy is clearly desired by end-users (Goldfarb and Tucker 2012), it might reduce the quality of services provided to them e.g. poor targeting of online ads, and thus adversely affect their preferences towards the service (Goldfarb and Tucker 2011). However, this reduction might be countered by an increase in the willingness of the users to use the service, due to the added privacy guarantees (Tucker 2014). Evidently, researchers have proposed certain information-theoretic frameworks to better quantify these risks and benefits of data disclosure (Brickell and Shmatikov 2008; Li and Li 2009; Rastogi et al.

2007; Sankar et al. 2013).

Empirical research, on information privacy has mostly focused on the individual (Dinev and Hart 2006; Hui et al. 2007; Smith et al. 2011), with some work also focused towards the organizational and societal levels (Siponen 2005; Walczuch and Steeghs 2001). Several studies have looked at the influence of information technology advances on individual perceptions of security and subsequent behavior. The effects of direct marketing efforts (Blattberg and Deighton 1991; Campbell 1997), Internet commerce (Acquisti and Varian 2005; Dhillon and Moores 2003; Malhotra et al. 2004), increased surveillance (Allen et al. 2007), and social networks (Acquisti and Gross 2006; boyd and Ellison 2007; Boyd 2008) have been reported in recent research. While there exist an abundance of individual-centric studies, few seek to understand how organizations can achieve privacy and security with their user data or how strategic policies and interventions influence individual level perceptions of privacy. This is a gap that I seek to address with the current study.

4.2.2 Self-disclosure on SNS

SNS offer users a convenient medium to communicate with others, and to self-present by means of public posts, private messages, tags etc. While different SNS have different designs and communication strategies, appealing to different crowds with unique interests, the profile pages (of SNS users) are the common denominator among all such sites. The SNS profile is “a representation of their [selves] (and, often, of their own social networks) - to others to peruse, with the intention of contacting or being contacted by others”

(Gross and Acquisti 2005). The elements of profile data offer a window into the user's self-reported tastes and preferences and range from relatively innocuous, such as favorite music or book, fields to potentially sensitive ones, such as sexual orientation or political affinity. Early studies on self-disclosure and self-presentation in online platforms focused largely on online dating sites and personal homepages and investigated questions related to information disclosure patterns, cultural and gender differences and differences in site usage behavior among users (Gross and Acquisti 2005; Kim and Papacharissi 2003).

Given that Facebook and Myspace were the most prominent SNS during the formative years of online social networks, particularly among college students, early privacy studies on SNSs almost exclusively focused on the users of these sites. Facebook was often referred to as a "walled garden" (Tufekci 2007) due to the sharp demarcation between what was publicly visible from one's profile, versus what was only visible to insiders from one's friendship network. Conversely, Myspace was open to everyone by default, and therefore regarded as less private. Among early studies in these contexts, Jones and Soltren (2005) found that more than half of the students disclosed information about their favorite books, music, and interests, but much less (17.1%) disclosed their phone numbers. Stutzman (2006) concluded that students overwhelmingly disclosed their birthday, relationship status, and political view, while disclosure of cell phone number was limited to 16.4%. Further, Gross and Acquisti found that only a small set of users adjust the default (permissive) privacy settings to restrict the visibility of their profiles and that even the highly concerned users revealed extensive personal information on their profiles (Acquisti and Gross 2006; Gross and Acquisti 2005). These findings were consistent with previous findings by Lampe et al. (2006) and Tufekci (2007). Departing from these studies that focused exclusively on static profile attributes, Lewis et al. (2008) examined relational data in the form of friendship and roommate ties as factors that contributed to a student's privacy preference, measured by a choice to have a private vs. a public profile on the SNS. Their results showed that a student was more likely to have a private profile if the student's friends, and especially roommates, have private profiles or the student was active on Facebook. In a related work, Stutzman and Kramer-Duffield (2010) found that having a friends-only profile was more likely for users with a large friend network, implying that there might be a potential inflection point in the number of friends beyond which users transform their profiles from open to friends only. Prior research has also identified gender and racial differences in influencing self-disclosure, but the results are not consistent across studies (Acquisti and Gross 2006; Lewis et al. 2008; Tufekci 2007).

4.2.3 Malleability of privacy preferences

While individuals are often unaware of the factors that influence their concerns about privacy in a given context, organizations which benefit from heightened information disclosure have developed tools to analyze and encourage information revelation. These include SNS providers who wish to encourage their users to engage more frequently with the platform and post more content, and digital advertisers who wish to perform better ad targeting based on personalized information (Goldfarb and Tucker 2011; Lazer et al.

2009). Such entities often exploit what is called the malleability of privacy concerns (Acquisti et al. 2015), a term used to refer to the situation where certain factors, subliminal or otherwise, can be used to activate or suppress our privacy concerns, which in turn can impact our observable behaviors.

Websites, particularly SNS, often make use of default settings to influence information disclosure. Extant research have shown that the default choice option often disproportionately impacts decisions as varied as organ donation and retirement savings (Amir et al. 2005; Johnson and Goldstein 2003). Default settings are sometimes not salient, and at other times offer a convenient option for individuals to persist with, in absence of a better choice. Popular examples of such default settings include features on SNS to choose profile visibility (Lewis et al. 2008), and opt-in or opt-out privacy policies on websites (Johnson et al. 2002). In addition to default settings, websites can often create features that frustrate or confuse users into revealing personal information (Conti and Sobiesk 2010). For example, websites often have privacy policy mentioned on the platform, but this is seldom read by the users. In a survey, 62% of respondents incorrectly believed that the existence of a privacy policy ensured that the site could not share their personal data without their permissions (Marx 2001).

The final aspect that researchers have looked at, and is also the focus of the current study, is about privacy controls. Control over personal information and audience of this information (i.e. what I disclose, and who views it) is critical to most SNSs today (Marwick and Boyd 2011). For instance, while studying the reaction of Facebook users to the introduction of the “News Feed” feature, Hoadley et al. (2010) found that users expressed a higher concern for privacy, stemming from a perceived loss of control over personal information. In their study on location-based services, Xu (2007) showed that perceived controls can potentially mitigate such privacy concerns. In principle, privacy controls empower users with a heightened sense of control over their self-disclosure, which in turn reduces their privacy concerns. However, this often has unintended consequences, as pointed out by privacy proponents in the past. For instance, a past research finds that individuals who are provided with the option of deciding whether and how much of their personal information could be used by the researchers for publication, ended up disclosing more information with a public audience (Brandimarte et al. 2012), which is the exact opposite intention of introducing such controls in the first place.

There are two major gaps with the previous research that looks at the impact of privacy controls on observable behavior. First, the studies do not consider or control for any accompanying change in the social network structure that might result from the privacy-related intervention. In absence of this, the effect of the intervention on behavior might be under- or over-estimated. Second, the studies only focus on public channels for observing behavior. However, it remains to be observed how such privacy interventions influence behavior in private channels too. Most social media platforms today are characterized by a public as well as private channel, and hence, it is important to understand the impact of such interventions on, and the potential spill-over effects between, both these channels.

4.3 Hypotheses Development

Based on the review of previous work in this area, I propose three competing hypothesis to test the effect of privacy controls on online content generation behavior at a large SNS.

Existing theories on psychology, particularly the Theory of Planned Behavior (Ajzen 1991) argue that an individual’s intention to perform a behavior is determined partly by her perceived behavioral controls. Controllability, which is an individual’s perceived sense of control over a specific action, has been shown to positively affect actual behavior (Ajzen 2002; Terry and O’Leary 1995). This is partly mediated by lowering of perceived risks involved in performing the actions. For example, previous studies show that the effect of risk perception on risk acceptance is moderated by a perceived level of control, such that individuals tend to evaluate risks as less severe when they perceive higher control over the action (Fischhoff et al. 1978). Consequently, when individuals perceive a lower risk, they tend to participate or engage more with the associated action. Thus, in our context, a lowering of perceived risk of sharing personal information might prompt users to participate on the SNS more actively. Therefore, I hypothesize that the introduction of privacy controls positively influences the volume of content produced publicly on SNS.

H1: Introduction of privacy controls increases the volume of public content generated on SNS, after controlling for any change in the underlying friendship network.

A second possibility I explore in the current study is that of a substitution effect across channels. We know from the previous discussion about the privacy calculus and also from related theories like the Communication Privacy Management (Petronio 2012), that individuals have personalized information boundaries and they constantly balance their need to maintain secrecy with an urge to increase openness. Increased self-disclosure on SNS has been linked to an increase in self-esteem and psychological well-being (Gonzales and Hancock 2011; Kim and Lee 2011; Valkenburg et al. 2006). However, recent studies have also linked high self-esteem with a lowering of self-control and the subsequent indulgence in risky activities like, for example, revealing more personal information online that what is required (Wilcox and Stephen 2013). Since, most individuals on SNS manage both public and private channels, they try to compartmentalize their disclosure based on the directedness and secrecy, but often fail to do so successfully, a phenomenon termed as context collapse by sociologists (boyd and Ellison 2007; Marwick and Boyd 2011). Thus, I predict that in response to privacy-related changes, individuals might substitute content on their public channel (e.g. public updates), with more content from their private channel (e.g. private messages). This should manifest by a decrease in the overall volume of private messages sent by the individual. Thus, I propose the second competing hypothesis:

H2: Introduction of privacy controls decreases the volume of private content generated on SNS, after controlling for any change in the underlying friendship network.

4.4 Data Context

Through collaboration with a large and popular SNS, I first identified SNS users who were American students and attended college anytime during

2007-2011, and signed up on the SNS before the privacy change event. This search resulted in more than 1.3 million active users. For this study, I restricted my sample to users who attended the same university as identifiable from their self-reported profile information. The resulting user set consisted of 2,696 active users. Next, I gathered profile information such as gender, biological age and the date of registration for each user in the user set. These are used as control variables and are the same as the ones used in the previous chapter, namely gender, age and SNS tenure (hereafter, referred to as SNS Age). Additionally, I also obtained data on the dynamic network of number of friends added in each week for each user[21]. Finally, I computed the volume of public postings and private conversations for each user. The total number of messages and posts produced by each user was aggregated at a weekly level from Oct 6, 2009 (week 110) till Feb 15, 2010 (week 130). The descriptive summary of the key variables are provided in Table 4-1 in the following page.

The privacy control was introduced by the SNS on December 9, 2009 (week 119) and was widely publicized by the SNS as well as by the popular press. Since the policy change was purely exogenous to the users of the SNS, it provides an opportunistic setting to study the impact of giving users more control over the information they generate and share on the SNS. Specifically, I seek to investigate if this introduction of finer privacy controls that provided users with greater control over what they shared, resulted in more open disclosure, as supposedly intended by the SNS and worried by privacy advocates. Figures 4-1 and 4-2 show the change in content generated and friendships formed before and after the introduction of the privacy change.

[21] The network data showed no evidence of any friend deletions

The Figures 4-1, 4-2 and Table 4-2 offer a descriptive view of the impact of the privacy control on the number of friends added as well as the public and private content produced on the platform. A visual inspection of the plots as well as the summary in Table 4-2 suggests that the treatment had no effect on the volume of public posts, but reduced the volume of private messages. Also, Figure 4-1 shows that the number of new friendships formed on the SNS after introduction of the privacy control saw a significant reduction before picking up after about a month. While I do not hypothesize whether or not the reduction in network growth can be attributed to the privacy change, this significant decrease creates a confound for us in analyzing the effect of the privacy treatment. I address this by leveraging the network-behavior co-evolution model that was introduced in the previous chapter. The next section specifies this model for the current context and presents the estimation results.

Table 4-1: Descriptive statistics of model variables

	Mean	**Std. Dev.**	**Min**	**Max**
Number of Users (Undergraduate students from a large US University): 2696 **Period of observation:** Oct 6, 2009 - Feb 15, 2010 No. of observations: 56,616				
Dependent Variables:				
Number of Public Posts *(PubVol)*	0.435	0.504	0.000	11.000
Number of Private Messages *(PriVol)*	0.355	0.541	0.000	13.000
Independent Variables:				
Privacy Treatment *(Privacy Treatment)*	0.571	0.494	0	1
Out-degree *(Degree)*	104.032	82.173	0	677
SNS Age (in days)	1776.502	381.540	712	2592
Biological Age *(Age)* (in years)	22.350	1.541	20	26
Gender	1.458	0.563	0	4

Figure 4-1: The weekly average number of friends (above) and weekly "new" friends added (below) 10 weeks before the introduction of privacy controls during the week of December 7 (red vertical line), and 10 weeks after

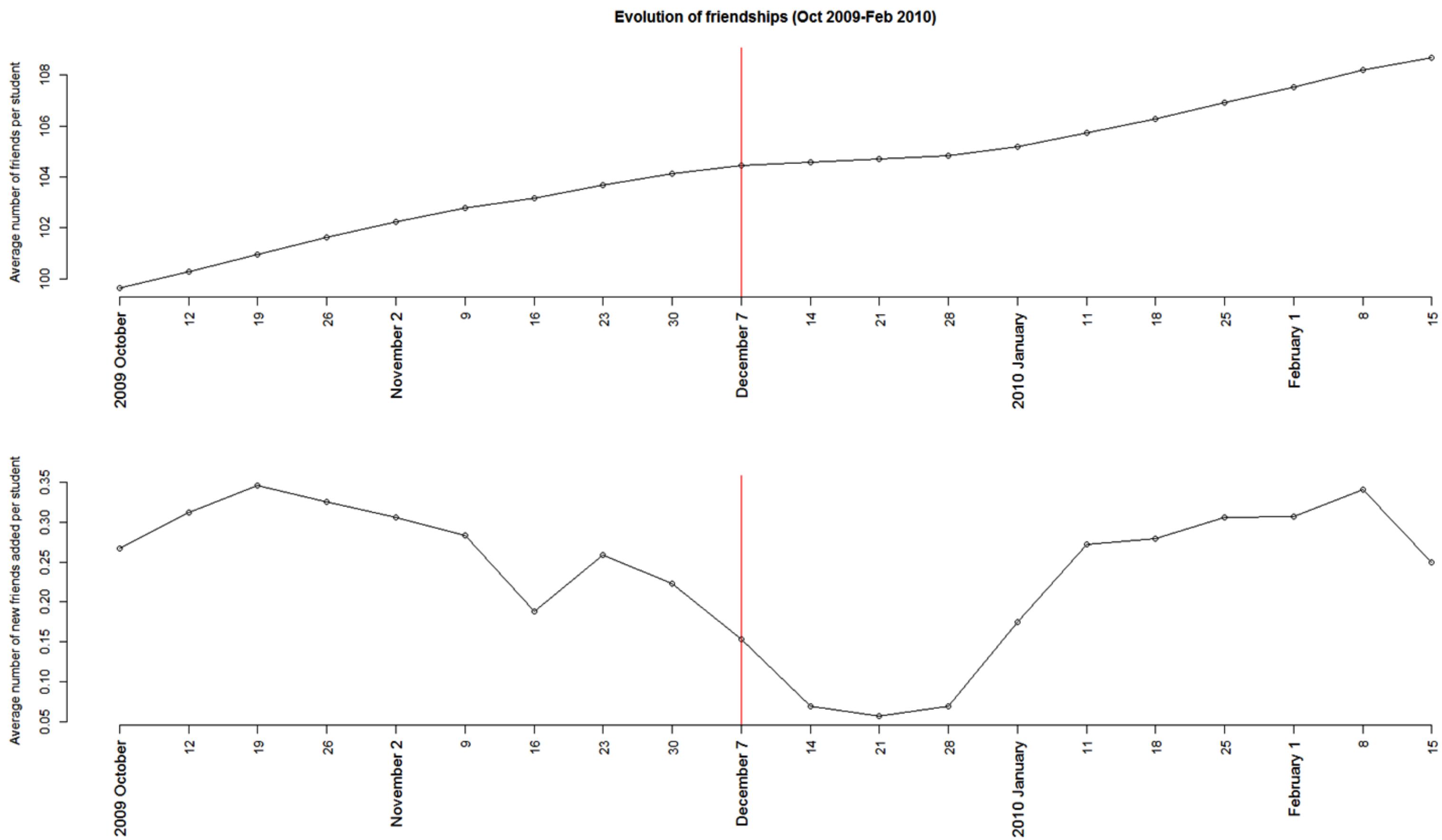


Figure 4-2: The average number of posts (above) and messages (below) generated per user per week between October, 2009, and February, 2010. The privacy controls were implemented during the week of December 7 (red vertical line)

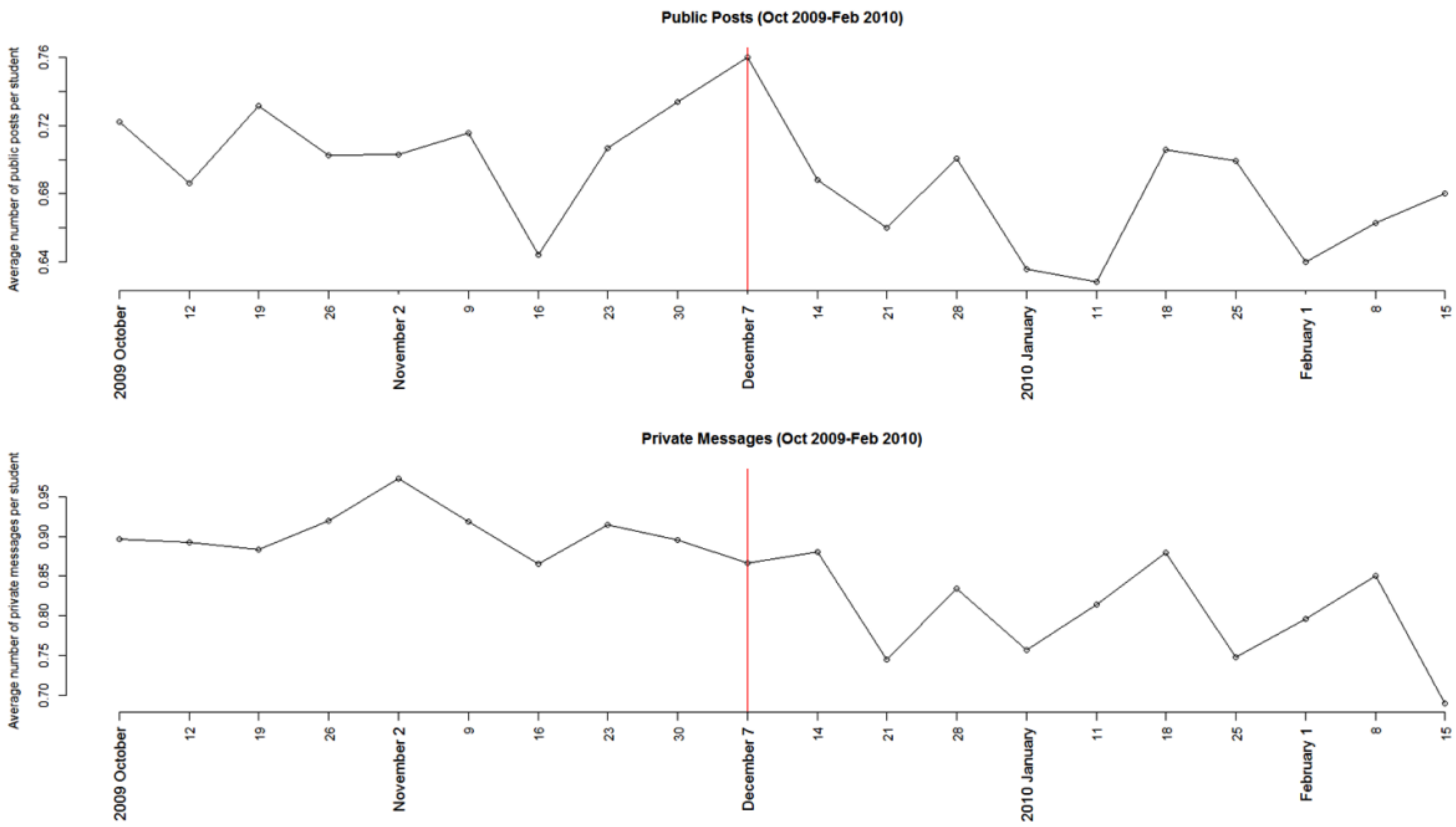


Table 4-2: Descriptive summary of treatment effect on public and private content generation

	Pre-treatment		Post-treatment	
	Mean	Std.Div.	Mean	Std.Div.
Public Posts	0.435	0.507	0.436	0.502
Private Messages	0.370	0.572	0.342	0.514

4.5 Empirical Analysis: The Co-evolution Model

As mentioned earlier, the empirical analysis in the current context to model the effects of the privacy control intervention is complicated by the presence of a confounding network i.e. the social network and the content production co-evolve by influencing each other. In other words, the effect of the natural treatment on one of the outcomes can potentially be confounded by the presence of the other. The details of the co-evolution model (Lewis et al. 2012; Snijders et al. 2007; Steglich et al. 2010) have already been discussed in detail in the previous chapter, and I do not reiterate the model construction details here. The objective functions (also referred to as evaluation functions) that I estimate in the model are specified as follows.

$$f_i^{[A]}\left(\beta^{[A]}, A_t, P_t\right) = \sum_{k_1} \beta_{k_1}^{[A]} s_{ik_1}^{[A]}(A_t, P_t) \quad (\mathit{network\ evaluations}) \tag{1}$$

$$f_i^{[P]}\left(\beta^{[P]}, A_t, P_t\right) = \sum_{k_2} \beta_{k_2}^{[P]} s_{ik_2}^{[P]}(A_t, P_t) \quad (\mathit{behavior\ evaluations}) \tag{2}$$

where, $s_{ik_1}^{[A]} = [S^{[A]}]_{ik_1}$ is the ${k_1}^{th}$ network statistic of user i, and, similarly, $s_{ik_2}^{[P]} = [S^{[P]}]_{ik_2}$ is the ${k_2}^{th}$ behavioral statistic of user i. In the following two sections I detail the specification of the network statistics ($\mathrm{s}_{\mathrm{t}}^{[\mathrm{A}]}$) and behavior statistics ($\mathrm{s}_{\mathrm{t}}^{[\mathrm{P}]}$) that I model in the current context.

4.5.1 Specification for network statistics for the co-evolution model

The network statistics that I model in this study are the user's out-degree ($\mathrm{s}_{\mathrm{i1t}}^{[\mathrm{A}]}$), transitivity ($\mathrm{s}_{\mathrm{i2t}}^{[\mathrm{A}]}$), homophilous friendship formation based on posting behavior[22] ($\mathrm{s}_{\mathrm{i3t}}^{[\mathrm{A}]}$), homophilous friendship formation based on the covariates i.e. Gender ($\mathrm{s}_{\mathrm{i4t}}^{[\mathrm{A}]}$), Age ($\mathrm{s}_{\mathrm{i5t}}^{[\mathrm{A}]}$), and SNS Age ($\mathrm{s}_{\mathrm{i6t}}^{[\mathrm{A}]}$). The complete specifications for these statistics are as follows.

- Out-degree ($\mathrm{s}_{\mathrm{i1t}}^{[\mathrm{A}]}$)= $\sum_{\mathrm{j}} \mathrm{a}_{\mathrm{ijt}}$
- Transitivity ($\mathrm{s}_{\mathrm{i2t}}^{[\mathrm{A}]}$) $= \sum_{\mathrm{j,h}} \mathrm{a}_{\mathrm{ijt}} * \mathrm{a}_{\mathrm{jht}} * \mathrm{a}_{\mathrm{iht}}$
- Homophily based on posting behavior ($\mathrm{s}_{\mathrm{i3t}}^{[\mathrm{A}]}$) =

$$\mathrm{s}_{\mathrm{i3t}}^{[\mathrm{A}]}(\mathrm{a},\mathrm{p}) = \mathrm{a}_{i^+t}^{-1} \sum_{\mathrm{j}} \mathrm{a}_{\mathrm{ijt}} \left(1 - \frac{|p_{it} - p_{jt}|}{R_{pt}}\right)$$

, where R_{pt} represents the range of the posting variable at step t. $\mathrm{s}_{\mathrm{i3t}}^{[\mathrm{A}]}$ represents the effect of homophily based on posting behavior such that $\mathrm{s}_{\mathrm{i3t}}^{[\mathrm{A}]}$ takes higher value for those users who have a similar posting volume as that of their peers (i.e. reflected by a smaller value of $|p_{it} - p_{jt}|$). For homophily based on covariates like the gender, age and SNS age, we have similar expressions for $\mathrm{s}_{\mathrm{i4t}}^{[\mathrm{A}]}$, $\mathrm{s}_{\mathrm{i5t}}^{[\mathrm{A}]}$ and $\mathrm{s}_{\mathrm{i6t}}^{[\mathrm{A}]}$.

I model the effect of the privacy treatment on the network out-degree using the following specification.

---

[22] The phrase “content posting” or “posting behavior” is used synonymously with “content generation” or “content production” in this study, as refers to the generation of publicly viewable posts on the SNS.

- $Privacy\ Treatment_{it}^{[A]} = \sum_j \mathrm{a}_{ijt} * \mathrm{privacy}_{it}$

where, a_{ijt} indexes the adjacency matrix and takes the value 1 if a network tie exists between i and j at period t and 0 otherwise, and $\mathrm{privacy}_{it}$ is a dummy variable that takes the value of 1 if the user i has been exposed to the privacy control intervention in period t, and 0 otherwise.

In addition to the network statistics and the effect of the privacy intervention, I also explicitly control for the effect of actor-level covariates on the network evolution (e.g. out-degree). The following are the specifications of the effect of covariates on the network out-degree.

- $\mathrm{Gender}_{it}^{[A]} = \sum_j \mathrm{a}_{ijt} * \mathrm{x}_{1it}$
- $\mathrm{Age}_{it}^{[A]} = \sum_j \mathrm{a}_{ijt} * \mathrm{x}_{2it}$
- $\mathrm{SNS\ Age}_{it}^{[A]} = \sum_j \mathrm{a}_{ijt} * \mathrm{x}_{3it}$

where, $\mathrm{Gender}_{it}^{[A]}$ represents the effect of the user i's gender x_{1it} on her propensity to make new friends during step t, such that a positive and significant estimate on this statistic would imply that females (Gender = 1) have a higher propensity to make friends than males (Gender = 0) and vice versa. Similarly, for $\mathrm{Age}_{it}^{[A]}$ and $\mathrm{SNS\ Age}_{it}^{[A]}$. In all the above equations, a_{ijt} is 1 if a network tie exists between i and j at period t and 0 otherwise.

4.5.2 Specification for behavioral statistics for the co-evolution model

The behavioral statistics that I model in this study include the content generation tendency ($\mathrm{s}_{i1t}^{[P]}$), peer influence on content posting ($\mathrm{s}_{i2t}^{[P]}$)[23], the

[23] The phrase "content posting" is used synonymously with "content generation" or "content

effect of the privacy treatment on the content posting behavior ($Privacy\ Treatment_{\mathrm{it}}^{[\mathrm{P}]}$), and the effect of individual specific covariates like gender ($\mathrm{Gender}_{\mathrm{it}}^{[\mathrm{P}]}$), biological age ($\mathrm{Age}_{\mathrm{it}}^{[\mathrm{P}]}$) and SNS Age ($\mathrm{SNS\ Age}_{\mathrm{it}}^{[\mathrm{P}]}$) on the posting behavior. The complete specifications for these effects are as follows.

- Posting tendency ($\mathrm{s}_{\mathrm{i1t}}^{[\mathrm{P}]}) = \mathrm{p}_{\mathrm{it}}$, captures the natural tendency of the actors to increase or decrease content generation behavior over time[24].
- Peer influence on posting behavior ($\mathrm{s}_{\mathrm{i2t}}^{[\mathrm{P}]}) = \mathrm{s}_{\mathrm{i2t}}^{[\mathrm{P}]}(\mathrm{a}, \mathrm{p}) =$

  $a_{i^+t}^{-1} \sum_{\mathrm{j}} \mathrm{a}_{\mathrm{ijt}} \left(1 - \frac{|p_{it} - p_{jt}|}{R_{pt}}\right)$

  where, R_{pt} represents the range of the posting variable at Step t. $\mathrm{s}_{\mathrm{i2t}}^{[\mathrm{P}]}$ represents the effect of peer influence on posting behavior such that $\mathrm{s}_{\mathrm{i2t}}^{[\mathrm{P}]}$ takes higher value for those users whose posting volume is closer to that of their peers (i.e. reflected by a smaller value of $|\mathrm{p}_{\mathrm{it}} - \mathrm{p}_{\mathrm{jt}}|$). Thus, a positive and significant estimate on this statistic would indicate that users regulate their posting behavior to assimilate with their peers i.e. matching the posting rate of peers, and vice versa.

I model the effect of the privacy treatment on the posting behavior using the following specification.

- $\mathrm{Privacy\ Treatment}_{\mathrm{it}}^{[\mathrm{P}]} = \mathrm{p}_{\mathrm{it}} * \mathrm{privacy}_{\mathrm{it}}$

where, p_{it} is the number of posts made by the user i in period t,

production" in this study, as refers to the generation of publicly viewable posts on the SNS.

[24] One can think of this effect as similar to the intercept term in a regression model

$\text{privacy}_{\text{it}}$ is a dummy variable that takes the value of 1 if the user i has been exposed to the privacy control intervention in period t, and 0 otherwise.

In addition to the network statistics and the effect of the privacy intervention, I also explicitly control for the effect of actor-level covariates on the content generation behavior (e.g. volume of public posts and private messages generated). The following are the specifications of the effect of covariates on content generation.

- $\text{Gender}_{\text{it}}^{[\text{P}]} = p_{\text{it}} * \text{x}_{\text{i1t}}$
- $\text{Age}_{\text{it}}^{[P]} = p_{\text{it}} * \text{x}_{12\text{t}}$
- $\text{SNS Age}_{\text{it}}^{[\text{P}]} = p_{\text{it}} * \text{x}_{13\text{t}}$

where, $\text{Gender}_{\text{it}}^{[\text{A}]}$ represents the effect of the user i's gender on her propensity to produce new content during step t, such that a positive and significant estimate on this statistic would imply that females (Gender = 1) have a higher propensity to make friends than males (Gender = 0) and vice versa. Similarly, for $\text{Age}_{\text{it}}^{[P]}$ and $\text{SNS Age}_{\text{it}}^{[P]}$. In all the above equations, p_{it} denotes the volume of public posts or private messages generated on the SNS, depending on the choice of the dependent variable.

The following section details and discusses some of the results from the model estimation.

4.6 Results

Similar to the previous study, and due to the complexity of explicitly computing the likelihood function, I employ the use of simulation-based estimators. Specifically, I use a Markov Chain Monte Carlo (MCMC) based

Method-of-Moments (MoM) estimator to recover the parameters of these rate and evaluation functions. The MCMC implementation of the MoM estimator uses a stochastic approximation algorithm that is a variant of the Robbins-Monro (1951) algorithm (Robbins and Monro 1951).

I consider the following empirical analyses in this section:

1) The effects of the privacy controls on public posts generated in the week following the intervention

2) The effects of the privacy controls on private messages generated in the week following the intervention

In a later section, I compare results from my Markovian-based model to traditional regression models as a baseline.

4.6.1 Effect on public posts and private messages

The estimation results are presented in Table 4-3 and illustrate the parameter estimates for the effect of privacy controls on public posting and private messaging volume in the week after the intervention, while controlling for the network evolution as a potential confounder. Four model results are presented: (i) the first column shows the effects on public posting behavior, without modeling the network, (ii) the second column shows the effects on public posting, after controlling for the underlying network, (iii) the third column presents the effects on private messaging behavior without modeling the network, and finally (iv) the fourth column shows the effects on private messaging, after controlling for the underlying network. I find that after controlling for the interrelationships between the network and posting behavior, as well as other covariates, the privacy intervention has insignificant effect on the volume of public posts generated in the week following the privacy intervention ($\beta = -0.049$; $p>0.1$), but a significant negative effect on the volume of private messages generated in the same period ($\beta = -0.157$; $p<0.01$). Also, it is worth noting that intervention does have a significant and negative impact on the propensity of actors to make new friends across the post model ($\beta = -9.639$; $p < 0.01$), as well as the messages model ($\beta = -10.646$; $p < 0.01$). This provides further impetus for using a co-evolution model that controls for this confounding network evolution while estimating the effect of the privacy intervention on content generation.

Table 4-3 also highlights the potential bias introduced if we do not control for network evolution. For instance, in model (III), I find that in absence of the network evolution, the effect of the privacy treatment on volume of private messages generated is significant and negative ($\beta = -0.060$; $p < 0.01$). However, in model (IV), I model the underlying network evolution, and show the effect of the privacy treatment on private messages becomes stronger by a factor of over 2 ($\beta = -0.157$; $p < 0.01$).

In the next section, I exploit a quasi-experimental design to provide further identification evidence that the change in content generation behavior is caused by the privacy intervention.

4.6.2 Quasi-experimental analysis

In the previous section, I showed that following the introduction of new privacy controls, users on the SNS significantly reduced their generation of private messages, while showing no change in the volume of public messages. In this section, I provide further evidence that this change in content generation behavior is driven purely by the privacy intervention and not any other confounding factors, unobserved to the researcher. Specifically, I exploit a quasi-experimental design (Cook et al. 1990) where I classify users based on their privacy consciousness prior to the intervention. The intuition behind this identification strategy is that if the change in content generation behavior is indeed driven by the privacy intervention, then individuals with varying prior beliefs about privacy should react in different ways.

Table 4-3: Estimation results for co-evolution model

Parameter	**I. Posts**	**II. Posts** **(with Networks)**	**III. Messages**	**IV. Messages** **(with Networks)**
Privacy Treatment	-0.029 (0.021)	-0.049 (0.040)	-0.060*** (0.022)	-0.157*** (0.038)
Posting Tendency	-0.812*** (0.011)	-0.527*** (0.016)	-1.032*** (0.012)	-1.028*** (0.015)
Gender	-0.236*** (0.020)	-0.135*** (0.029)	-0.220*** (0.019)	-0.214*** (0.025)
Age	-0.247*** (0.010)	-0.098*** (0.013)	-0.207*** (0.009)	-0.209*** (0.012)
SNS Age	0.0004*** (0.00001)	0.00001 (0.0001)	0.0004*** (0.00001)	0.0004*** (0.0001)
Out-degree	-	-9.639*** (0.016)	-	-10.646*** (0.035)
Transitivity	-	0.066*** (0.001)	-	0.081*** (0.001)

*** <0.01, ** <0.05, *<0.1

To this effect, I construct an individual-level measure, *Privacy Index (PI)* to be the ratio of the volume of private messages to the volume of public posts generated by each user from Oct 5th, 2009 till Dec 8th, 2009 (i.e. till the day before the privacy intervention). I contend that individual with a very high score of PI (e.g. top 20%ile) are individuals with high privacy consciousness and are much less likely to be affected by the privacy intervention on the SNS. This is because such individuals already have a strong prior awareness about the privacy sensitivity of the content they produce on the SNS, as evident by their disproportionate use of private channel over the public one. Thus, any new privacy-related intervention is unlikely to significantly influence the way they generate content or disclose information. These individuals, thus, form a quasi-control for us following the privacy intervention. In contrast, individuals with a lower PI value (e.g. bottom 80%ile) are those with low privacy consciousness and are thus, more likely to become aware of privacy concerns surrounding their content generation as a result of the intervention. These users would therefore regulate their content generation behavior to a greater extent when made aware about privacy-related issues and features on the platform. These individuals, thus, form the quasi-treatment group for us following the privacy intervention.

Based on the 80-20 split in the PI value[25], I was able to classify 809 users as high privacy conscious and 1887 as low privacy conscious users in my sample. Tables 4-4 and 4-5 present the descriptive summary of covariates across the two groups, and results from a t-test on these observable covariates, namely distribution of age, SNS age and number of males vs. females across the quasi-treatment vs. quasi-control group. The t-test results show that there exist no significant differences across the two groups on any observable

[25] I also perform sensitivity analysis using 70-30 and 90-10 privacy splits. The effect is stronger for the 90-10 split and weaker for the 70-30 split. The complete sensitivity analysis results have not been presented here due to page limitations.

covariates (i.e. the only observable difference is the exposure or non-exposure to privacy treatment).

Next, I re-estimated the co-evolution model with an additional network and behavioral statistic to capture the group membership. The additional network statistic included in the model is as follows:

- $Privacy\ Treatment\ High\ PI_{it}^{[A]} = \sum_j \mathrm{a}_{ijt} * \ \mathrm{privacy}_{it} * \mathrm{group}_i$

where, a_{ijt} indexes the adjacency matrix and takes the value 1 if a network tie exists between i and j at period t and 0 otherwise, $\mathrm{privacy}_{it}$ is a dummy variable that takes the value of 1 if the user i has been exposed to the privacy control intervention in period t, and 0 otherwise.

Table 4-4: Descriptive summary of quasi-experiment groups

	High Privacy Conscious (N = 809)			Low Privacy Conscious (N = 1887)		
	Gender	Age	SNS Age	Gender	Age	SNS Age
Min	0	20	712	0	20	719
Max	4	26	2592	3	26	2591
Mean	1.497	22.38	1791	1.441	22.34	1770
SD	0.552	1.521	386.964	0.566	1.55	379.214

Table 4-5: Mean comparison of covariates across groups

	Age	**SNS Age**	**Gender**
t stat	0.597	1.31	-1.64
p	0.551	0.19	0.101

$\mathrm{group_i}$ is a dummy variable that takes the value of 1 if the user i is within the top 20%ile of the PI value, and 0 otherwise. Similarly, the additional behavioral statistic included in the model is as follows:

- $\text{Privacy Treatment High PI}_{it}^{[P]} = p_{it} * \text{privacy}_{it} * \text{group}_i$

where, p_{it} is the number of posts made by the user i in period t, privacy_{it} is a dummy variable that takes the value of 1 if the user i has been exposed to the privacy control intervention in period t, and 0 otherwise, group_i is a dummy variable that takes the value of 1 if the user i is within the top 20%ile of the PI value, and 0 otherwise. The estimation results are illustrated in Table 4-6 below.

The results from the estimation support the hypothesis that it is indeed the privacy intervention that alters the content generation behavior of the users. Specifically, and as I predicted, while the volume of posts no significant difference across the two groups of users, the volume of private messages showed a significant decrease on average (β = -0.163; $p < 0.01$), but a significantly attenuated decrease for the high privacy conscious use group (β = 0.146; $p < 0.05$). This shows that high privacy conscious users are less influenced by the privacy treatment, and the overall decrease in the volume of private messages was mainly driven by the large number of SNS users who had a low prior state of privacy consciousness.

In the next section, I compare the current results with that from a regression model and highlight the superiority of my approach in terms of explanation and model fit.

4.6.3 Comparison with regression models

I perform a baseline analysis using a discrete-time regression model (fixed and random effects panel regression) that regress the volume of posts, and the volume of messages on the network degree, as well as the associated covariates of gender, age and SNS age i.e. Controls in the model specification below:

$$Posts_{it} = \beta_0 + \beta_1 * Degree_{it} + \beta_2 * Privacy_t + \sum_{k=1}^{3} \beta_k * Controls_k + \alpha_i + \gamma_t + \epsilon_{it} \quad (3)$$

$$Messages_{it} = \beta_0 + \beta_1 * Degree_{it} + \beta_2 * Privacy_t + \sum_{k=1}^{3} \beta_k * Controls_k + \alpha_i + \gamma_t + \epsilon_{it} \quad (4)$$

Table 4-6: Sub-group estimation results for co-evolution model

Parameter	**Posts** **(with Networks)**	**Messages** **(with Networks)**
Privacy Treatment	-0.041 (0.051)	-0.163*** (0.054)
Privacy Treatment High PI	0.004 (0.080)	0.146** (0.074)
Posting Tendency	-0.587*** (0.023)	-0.838*** (0.022)
Gender	-0.151*** (0.035)	-0.141*** (0.039)
Age	-0.113*** (0.018)	-0.086*** (0.018)
SNS Age	0.00001 (0.0001)	0.00001 (0.0001)
Out-degree	-9.788*** (0.021)	-9.755*** (0.021)
Transitivity	0.069*** (0.001)	0.070*** (0.001)

*** <0.01, ** <0.05, *<0.1

Table 4-7 below illustrates the estimation results from the fixed effects and random effects panel regression model[26]. A Hausman Test revealed that the fixed effects estimator was the only consistent estimator.

Table 4-7: Estimation results for fixed- and random-effects regression

Parameter	**I. Posts (FE)**	**II. Messages (FE)**	**III. Posts (RE)**	**IV. Messages (RE)**
Out-degree (*Degree*)	0.003* (0.002)	-0.005* (0.003)	0.002*** (0.0001)	0.002*** (0.0001)
Privacy Treatment (*Privacy*)	0.003 (0.011)	-0.069*** (0.003)	-0.015 (0.011)	-0.136*** (0.012)
Gender	-	-	-0.112*** (0.010)	-0.087*** (0.009)
Age	-	-	-0.081*** (0.005)	-0.052*** (0.005)
SNS Age	-	-	0.0001 (0.00002)	0.0001 (0.00002)
Intercept	0.093 (0.189)	1.007*** (0.278)	2.025*** (0.089)	1.425*** (0.094)
Week dummies R^sq N	Present 0.121 2696	Present 0.075 2696	Present 0.181 2696	Present 0.105 2696

*** <0.01, ** <0.05, *<0.1

On comparing the results from Table 4-7, column II, with that of the earlier co-evolution model, I find that the fixed-effects regression model underestimates the effect of the privacy intervention on volume of messages exchanged by a factor of over 2 ($\beta = -0.069$; $p < 0.01$). Moreover, and as I show in the following section, the regression model also provides a

[26] A Hausman Test was performed and it was ascertained that the Fixed effects model was the only consistent model of the two.

substantially poorer model fit when compared to the co-evolution model.

4.6.3.1 Goodness of Fit Analysis

I highlight the goodness of fit of the analysis through two complementary approaches. (i) Comparing fit of in-model statistics (i.e. functions that were directly fitted by my model) and (ii) Comparing fit of auxiliary statistics (i.e. functions that were not directly fitted by my model).

For the former, Figures 4-3 and 4-4 compare the Mean Absolute Error (MAE) score obtained after fitting the out-degree statistic across the co-evolution model using the Method-of-Moments (MoM) estimator and the regression model (FE estimator). The MAE score obtained for 1000 simulation runs of the co- evolution model were consistently lower than the MAE scores for the regression model. Second, to avoid possible concerns surrounding over-fitting of the co-evolution model, I perform a goodness of fit analysis using out-of-model auxiliary statistics. I compute the ratio of the fitted sample-mean for the weekly posts to the fitted sample-mean for the weekly messages, as my auxiliary statistic. Since this statistic was not explicitly included in the co-evolution model specification, the fit of the co-evolution model would be not as good as with the previous approach, thereby reducing potential concerns of model over-fitting. However, as the Figure 4-5 demonstrates below, even for the auxiliary statistic, the fitted ratio generated from the co-evolution model over 1000 simulated runs is closer to the actual observed ratio for most of the runs, when compared to the fitted ratio from the regression model. This highlights the fit-superiority of the co-evolution model over conventional discrete-time regression models.

Figure 4-3: Comparison of Mean Absolute Error (MAE) from 1000 draws of the co-evolution posts model and FE regression model (red)

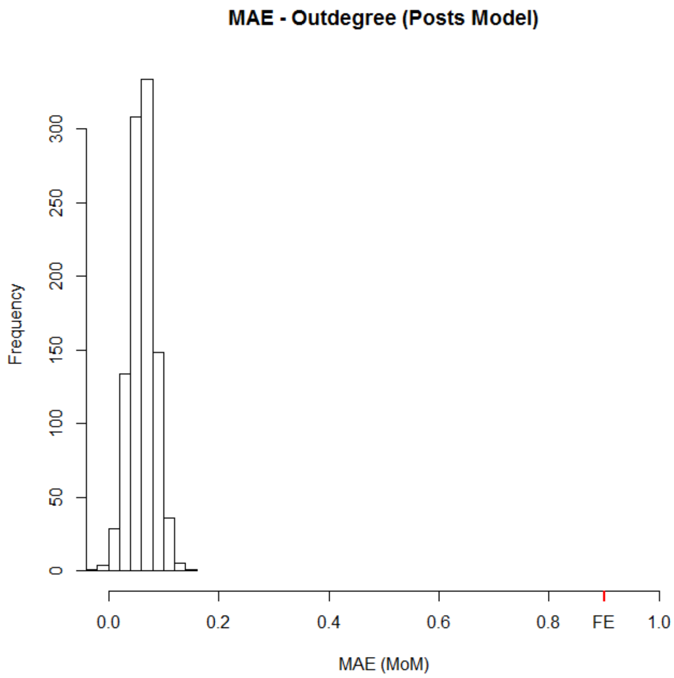


Figure 4-4: Comparison of Mean Absolute Error (MAE) from 1000 draws of the co-evolution messages model and FE regression model (red)

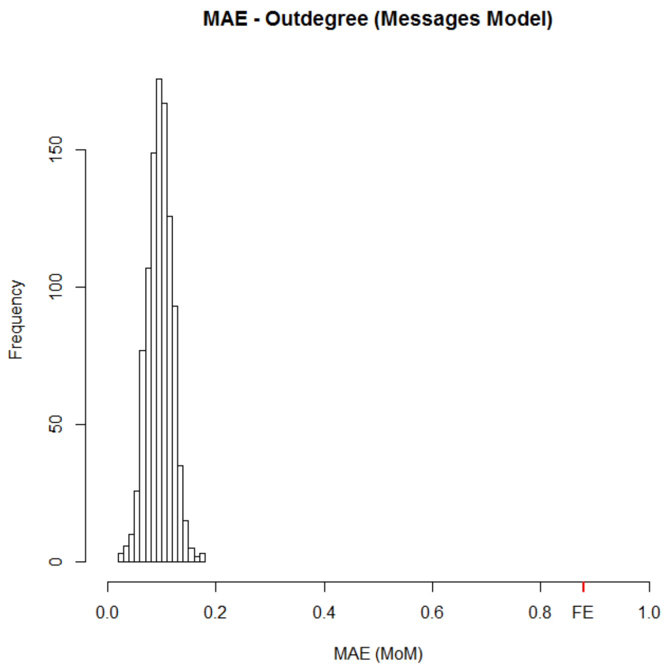
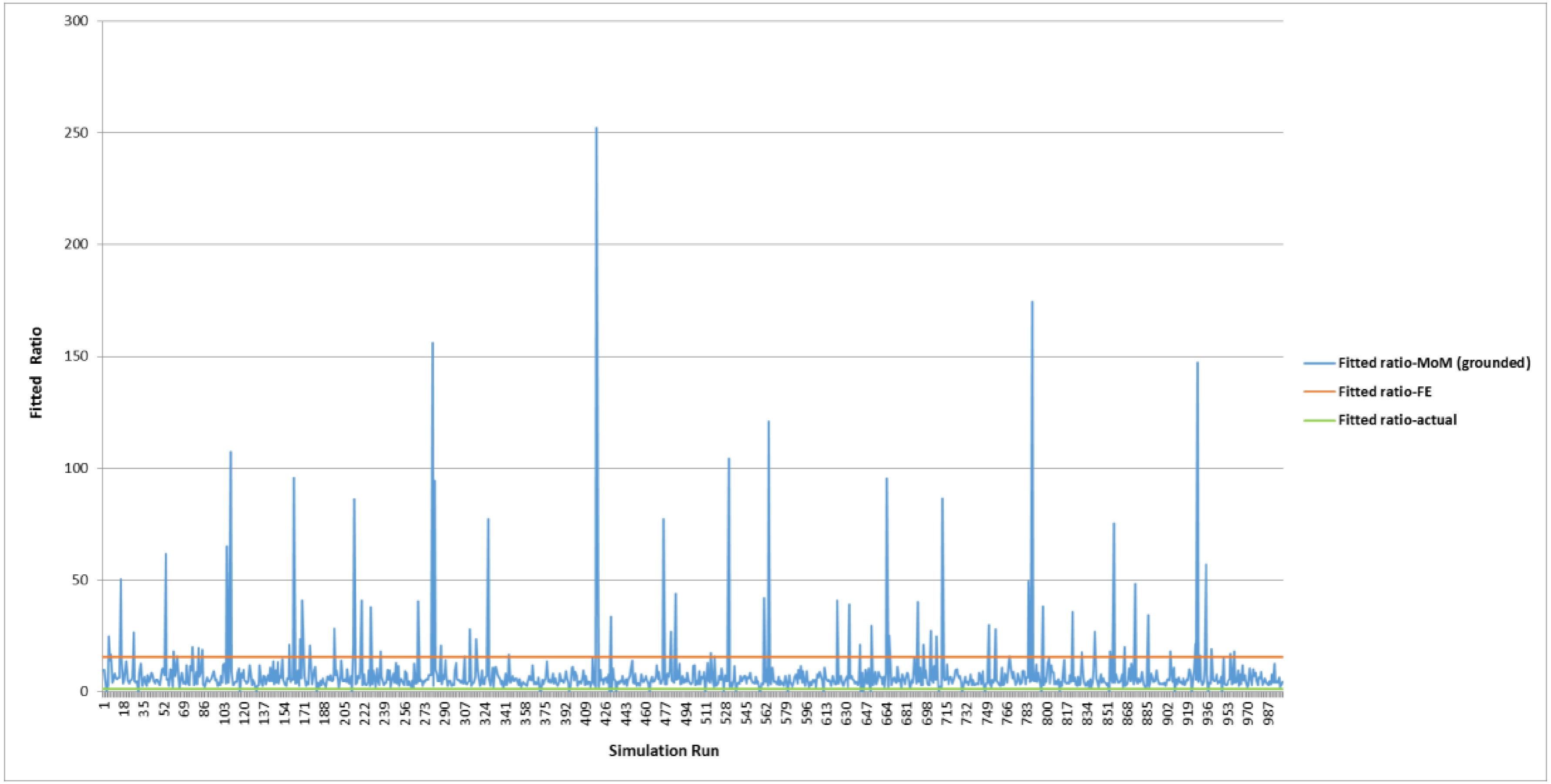


Figure 4-5: Comparison of auxiliary fit statistic across co-evolution and FE regression model



4.7 Discussion and Conclusion

The current study provides a good real world application of the co-evolution model as demonstrated in the previous chapter, and also illuminates our understanding of individual decision-making in response to privacy-related changes. While previous studies have discussed the behavioral malleability of privacy behavior in individuals, and illustrated how default settings on websites, or the existence of opt-in and opt-out policies affect individual's privacy behavior (Acquisti et al. 2015; Johnson et al. 2002; Lewis et al. 2008), little is understood about how privacy controls influence content production on large online social networks. The findings from this study, thus, contribute to the theories of information disclosure, as well as to the privacy calculus theory by offering two key improvements – (i) illustrating the effect of privacy-related interventions on user behavior across public vs. private channel, and (ii) highlighting the confounding role of the network evolution in estimating the true effect of privacy interventions in such online contexts. I illustrate each of these points in detail next.

For instance, and as predicted by hypothesis 2(b), I show that while privacy interventions have no significant impact on the volume of public posts generated, the volume of private messages goes down significantly in the week following the introduction of the privacy controls. However, there existed no such change in the volume of posts or messages generated for the same weeks in the preceding year in our dataset, thereby indicating that the reduction might be a direct result of the introduction of the privacy controls. While I cannot conclusively infer the true reason for this, the current analysis indicates that privacy controls seem to be having their intended effect, in encouraging individuals to substitute more and more of their public posts with private messages i.e. users transfer content from their private to their public channel owing to higher confidence in the SNS's capability in preserving their privacy. Due to this volume substitution effect, the volume of public posts remains constant, but the volume of private messages significantly decreases, following the privacy intervention. Moreover, using a pair of quasi-treatment and quasi-control groups constructed based on the users' prior propensity to participate in private vs. public channels, I show that this effect of privacy controls is less pronounced for highly privacy conscious individuals, as compared to other individuals. This provides further evidence and helps to identify the effect of the privacy control on the content generation.

The other key finding that I illustrate through my analyses is the confounding role of the evolving network. I show that, by ignoring any underlying network change, as has been the case in several past studies in the area, we underestimate the effect of the privacy treatment by a factor of over 2. Moreover, by comparing my results from the network-behavior co-evolution model with a more conventional fixed- and random-effects panel regression models, I show that the regression models underestimate the impact of the privacy treatment. Further, my analysis of goodness of fit of the models shows that the regression models offer poorer model fit as compared to co-evolution model, on both, in-model statistics (i.e. variables that were explicitly fitted by the model) as well as auxiliary statistics (i.e. variables that were not explicitly fitted by the model). These set of analyses emphasize the applicability and superiority of the co-evolution model in the current research context.

The findings from this research can potentially benefit several stakeholders in the SNS ecosystem. For instance, SNS providers would be very interested in understanding how privacy-related feature changes would impact their businesses through increased or decreased user engagement across public and private channels. Privacy-related changes by SNS have become commonplace with an increase in attention on data-protection and confidentiality related issues worldwide. However, very little is objectively understood about how these privacy controls influence either friendship formation, or content production on these online platforms. In addition to platform owners, digital marketers and advertisers can also draw insights from my results to better understand the implication of privacy-related actions on their target consumer base. Recent studies have shown that advertisers should be increasingly conscious about such implications (Goldfarb and Tucker 2011; Tucker 2014). Finally, the SNS users themselves can glean more insights from my results on how such feature changes on the platform can subliminally affect their friendship formation and content generation behavior on the platform (Acquisti et al. 2015). This is extremely important not just for the individual users, but also for policy makers worldwide since several past work in the area have highlighted that individuals tend to disclose a lot of personally identifiable information on public platforms without being fully aware of their value or its consequences.

Chapter 5. **FUTURE PLAN**

In the current dissertation, I have attempted to shed some light on how value is created on large online social network sites (SNS), by the participating stakeholders like the individual user, the SNS provider, as well as advertisers. I have illustrated a set of three empirical studies that look at the role of value creation for each, and possibly more, of these three stakeholders. However, there are a number of important extensions that I plan to execute for each of these studies, as well as a couple of newer research questions that my dissertation spawns. I conclude this dissertation with a discussion of these future projects.

First, while the first study looks at the role of self-presenting users on the SNS, it would be extremely important to dive deeper into understanding the various self-presentation tactics used by users on the SNS (e.g. exaggeration, self-promotion, sympathy-seeking etc.), and how these tactics might influence other users on the SNS as well as on brand communities. Moreover, individuals who participate in these different forms of self-presentation might also display varying purchase trajectories.

Second, for both, the second and third studies, further work is required to fully explain the underlying mechanism that drives (a) the influence based on public posting behavior in Chapter 3, and (b) the reaction to privacy interventions in Chapter 4. While I do my best to offer potential explanations based on what I can glean from observational data, I plan to do more follow up studies to rule out possible confounds and conclusively infer the underlying mechanisms at play.

Third, a related question to what has been studied in Chapter 4 is the role of privacy treatments on the friendship formation behavior on the platform. This is, in a way, the reverse of what has been analyzed and presented in Chapter 4 in the sense that the network formation would now become the focal outcome variable, while the content generation would be the confounding factor. However, understanding how privacy interventions influence the formation and dissolution of social ties over time is an extremely important theoretical exercise. Moreover, it has strong practical implications for SNS providers as well as policy makers.

Fourth, while my first study (Chapter 2) discusses the role of self-presentation on the SNS, and the second and third studies (Chapters 3, 4) illustrates the role of evolving user behaviors on the SNS, an interesting question that my dissertation spawns is about whether self-presentation as a trait can evolve over time? If yes, how does it evolve as a function of the SNS participation as well as the focal user's network structure? While traits, by definition, are considered to be largely stable over time, it might be extremely useful to understand how the traits or the manifestations of these traits online vary over the course of the users' SNS tenure.

Fifth, and lastly, the current dissertation sheds light on value creation on the SNS. I foresee two key extensions that might be useful in fully deconstructing this process of value creation. First, while the current dissertation conceptualizes "value" separately for the three studies, as (i) economic value for external brands who advertise on the SNS, (ii) subjective well-being for the users, and (iii) value for the SNS providers, it might be useful to come up with a single multi-dimensional construct that fully captures all these facets of value creation on the SNS. This would not only be a strong theoretical contribution to both the social networks as well as the organizational value creation literature, but also be of great practical relevance to businesses in understanding how to quantify the value encapsulated in these online platforms. Second, in a future work, it might also be useful to adopt an innovation perspective to looking at value creation on the SNS. A number of SNS features like brand communities (e.g. Facebook groups and pages etc.), reactive buttons (e.g. Facebook reactions, Twitter love buttons etc.), and location-based tools (e.g. WeChat “find-near-you”) have strong potentials to spawn innovative social, business and political applications. In a future work, I plan to study the role of SNS features in driving new innovation in related domains (e.g. e-commerce) as well as unrelated domains (e.g. politics, social outcomes).

APPENDIX 1: STOCHASTIC APPROXIMATION

In this study, I employ a Method of Moments (MoM) estimation procedure for the model specified in Sec. 3.1.3 (Bowman and Shenton 1985). The MoM estimator for my data (A, P) and parameter sets $\theta^{[A]}$ and $\theta^{[P]}$ is based on a set of network and behavioral statistics $S_t^{[A]}$ and $S_t^{[P]}$, and is defined as the parameter value set for which the following conditions are satisfied.

$$E_{\theta^{[A]}}\left(S^{[A]}\right) = s^{[A]}(a, p) \tag{i}$$

$$E_{\theta^{[P]}}\left(S^{[P]}\right) = s^{[P]}(a, p) \tag{ii}$$

i.e., the expected values and the observed values of the statistics are the same.

The choice of network and behavior statistics have been discussed in Sec. 3.2. In the general case, conditional expectations from the moment equations (i and ii) cannot be computed explicitly. Thus, I use a stochastic approximation method (Robbins and Monro 1951) to solve these moment equations. The method used to solve (i) and (ii) involves iteratively generating a parameter sequence $\hat{\theta}$ according to the following iteration steps.

$$\hat{\theta}_{t+1}^{[A]} = \hat{\theta}_t^{[A]} - \sigma_t D_0^{-1}(S_t^{[A]} - s^{[A]}) \tag{iii}$$

$$\hat{\theta}_{t+1}^{[P]} = \hat{\theta}_t^{[P]} - \sigma_t D_0^{-1}(S_t^{[P]} - s^{[P]}) \tag{iv}$$

where, $S_t^{[A]}$ and $S_t^{[P]}$ are generated according to the distributions defined by $\hat{\theta}_t^{[A]}$ and $\hat{\theta}_t^{[P]}$ respectively. The step size σ_t needs to be a sequence that converges to zero. The sequence $\sigma_t = \frac{a}{b+t}$ for any two integers a and b satisfies this constraint. D_0^{-1} is an identity matrix.

Snijders (2001) shows that the convergence properties of this algorithm hold asymptotically for t ->∞(Polyak 1990; Ruppert 1988; Yin 1991).

APPENDIX 2: DESCRIPTIVE SUMMARY FOR SOCIAL NETWORK DATA

(a) :Descriptive summary for social network data

	Time Period					
Observation	**1**	**2**	**3**	**4**	**5**	**6**
Density	0.025	0.027	0.028	0.029	0.030	0.031
Average Degree *	63.276	67.165	70.377	72.42	74.844	77.747
Number of Ties	79317	84191	88217	90778	93817	97456
Missing Fraction	0	0	0	0	0	0

*** Average degree across all periods = 70.971**

(b) :Social network evolution summary

	Change in Ties					
Period	**0 => 0**	**0 => 1**	**1 => 0**	**1 => 1**	**Jaccard ***	**Missing**
1==>2	3057080	4874	0	79317	0.942	0 (0%)
2==>3	3053054	4026	0	84191	0.954	0 (0%)
3==>4	3050493	2561	0	88217	0.972	0 (0%)
4==>5	3047454	3039	0	90778	0.968	0 (0%)
5==>6	3043815	3639	0	93817	0.963	0 (0%)

* Jaccard Index = $\frac{N_{11}}{N_{01}+N_{10}+N_{11}}$, where N_{hk} is the number of tie variables with value h in one wave, or observation from my dataset, and the value k in the next wave.

APPENDIX 3: DESCRIPTIVE SUMMARY FOR BEHAVIOR DATA

2(a) Descriptive summary for behavior data

	Time Period					
Posting quantile	**1**	**2**	**3**	**4**	**5**	**6**
1 (lowest)	630	763	787	644	806	774
2	945	1027	1019	903	1015	1009
3	364	317	324	326	325	336
4	193	177	157	223	147	147
5	122	86	76	118	85	88
6	77	47	41	84	53	54
7	33	24	37	50	18	30
8 (highest)	26	18	19	45	18	20

Note: The figures in the cells indicate the number of users who have posted in that time period. Row 1 indicates the total number of first-quantile posters (i.e. low posters) in each of the 6 time periods. Similarly, Column 1 indicates the number of posters in each of the 8 posting quantiles for the first time period.

2(b) Behavior evolution summary

	Number of users			
Period	**Decrease Posting Behavior**	**Increase Posting Behavior**	**Constant**	**Missing**
1 => 2	1009	427	1071	0
2 => 3	674	653	1180	0
3 => 4	378	1057	1072	0
4 => 5	1066	367	1074	0
5 => 6	625	711	1171	0

APPENDIX 4: CONVERGENCE ASSESSMENT FOR NETWORK VARIABLES

Convergence assessment for network variables

Network Variables	Observed Value for Target Statistics	Av. Deviation of simulated statistic from target statistic (SD)
Friendship rate (Period 1)	9748.000	-370.044 (139.914)
Friendship rate (Period 2)	8052.000	-141.393 (126.777)
Friendship rate (Period 3)	5122.000	24.940 (97.883)
Friendship rate (Period 4)	6078.000	63.709 (108.249)
Friendship rate (Period 5)	7278.000	254.976 (120.326)
Out-Degree	454459.000	-83.906 (131.380)
Transitivity (No. of triads)	4445064.000	-2548.047 (3834.512)
Gender on Degree	11547.242	-76.155 (108.814)
Gender homophily	-2109.296	-37.071 (37.992)
Age on degree	-29955.212	-209.697 (276.408)
Age homophily	-6381.205	-45.223 (68.114)
Tenure on degree	19104.656	186.366 (157.918)
Tenure homophily	-1968.574	12.749 (98.977)
Posting homophily	22210.253	-199.796 (80.153)

APPENDIX 5: CONVERGENCE ASSESSMENT FOR BEHAVIOR VARIABLES

Convergence Assessment for Behavior Variables

Behavior Variables	Observed Value for	Av. Deviation of (SD Deviation)
Posting rate (Period 1)	2150.000	-33.206 (47.848)
Posting rate (Period 2)	1663.000	-25.970 (46.780)
Posting rate (Period 3)	2359.000	-89.171 (51.167)
Posting rate (Period 4)	2310.000	-67.194 (50.710)
Posting rate (Period 5)	1614.000	-77.020 (46.518)
Posting Tendency (Linear Shape)	1801.000	-5.009 (129.923)
Influence	1798.000	7.562 (15.943)
Gender on Posting	1825.000	3.782 (78.867)
Age on Posting	1755.000	-8.259 (220.323)
Tenure on Posting	1900.000	-8.098 (119.231)

APPENDIX 6: LATENT SPACE MODELS

Following past work on statistical network models on exponential random graph models (Frank and Strauss 1986; Wasserman and Pattison 1996), the homogenous monadic Markov model (Frank and Strauss 1986), the stochastic and mixed membership block-models (Airoldi et al. 2008; Wang and Wong 1987), and the latent class membership models (Nowicki and Snijders 2001), Hoff et al. proposed a statistical approach to represent network actors as points on a latent social space (Hoff et al. 2002). The actors' positions on this Euclidean space are a result of the actors' observed as well as unobserved characteristics, and hence, the distance between these points is reflective of any underlying latent homophily based on these unobserved factors. The latent space model emphasizes conditional independence of the relational ties such that, conditional on the positions of the actors in the latent space, the probabilities of the tie formation are independent of each other. The latent space model is specified as follows:

$$\Pr(A|Z, X, \theta) = \prod_{i \neq j} P(a_{ij}|z_i, z_j, x_{i,j}, \theta)$$

where, $x_{i,j}$ is a vector of the observed covariates comprising similarity based on age, gender and SNS tenure, and z_i captures the latent space positions of actor *i*. The latent position vector Z and the parameter set θ are both estimated from the model. Now, a convenient specification for the tie-formation probability $P(a_{ij}|z_i, z_j, x_{i,j}, \theta)$ is the logistic regression model as follows:

$$\eta_{ij} = \log odds\ (y_{ij} = 1\ |z_i, z_j, x_{i,j}, \alpha, \beta) = \alpha + \beta' x_{i,j} - |z_i - z_j|$$

We follow (Hoff et al. 2002) and assume that the $z_i's$ are independent draws from a spherical multivariate normal distribution as follows:

$$z_1, z_2 \ldots z_N \sim MVN_k(0, \sigma_Z^2 I_k)$$

where, N is the sample size, k is the dimension of the latent space,

The log-likelihood for the above latent space model is then constructed as follows:

$$log\ Pr(A|\eta) = \sum_{i \neq j}\{\eta_{ij} a_{ij} - \log(1 + e^{\eta_{ij}})\}$$

where, η_{ij} is the log odds of tie formation and given as $\alpha + \beta' x_{i,j} - |z_i - z_j|$. As is clear from the formulation of the log likelihood, the computation of this function requires a sum over N(N-1) terms, which leads to a run-time complexity of O(N^2). This makes the direct MLE estimation infeasible for large-sample networks datasets, such as ours. We perform the likelihood based inference by following an approximation strategy proposed in (Raftery et al. 2012) which reduces the computational cost from O(N^2) to O(N). The approximation uses a case-controlled approach as popularized by Breslow (1996) and Breslow et al. (1980) but with a stratified sampler, to represent the likelihood as a sum of case likelihood (for $a_{ij} = 1$) and control likelihood (for $a_{ij} = 0$). We then estimate the approximate likelihood using a MCMC estimator.

Bibliography

Acquisti, A., Brandimarte, L., and Loewenstein, G. 2015. "Privacy and human behavior in the age of information," *Science* (347:6221), American Association for the Advancement of Science, pp. 509–514.

Acquisti, A., and Gross, R. 2006. "Imagined communities: Awareness, information sharing, and privacy on the Facebook," in *Privacy enhancing technologies*, pp. 36–58.

Acquisti, A., and Varian, H. R. 2005. "Conditioning prices on purchase history," *Marketing Science* (24:3), INFORMS, pp. 367–381.

Ajzen, I. 1991. "The theory of planned behavior," *Organizational behavior and human decision processes* (50:2), Elsevier, pp. 179–211.

Ajzen, I. 2002. "Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior1," *Journal of Applied Social Psychology* (32:4), pp. 665–683 (doi: 10.1111/j.1559-1816.2002.tb00236.x).

Allen, M. W., Coopman, S. J., Hart, J. L., and Walker, K. L. 2007. "Workplace surveillance and managing privacy boundaries," *Management Communication Quarterly* (21:2), Sage Publications, pp. 172–200.

Allport, G. W. 1955. *Becoming; basic considerations for a psychology of personality* (Vol. 20), Yale University Press.

Amir, O., Ariely, D., Cooke, A., Dunning, D., Epley, N., Gneezy, U., Koszegi, B., Lichtenstein, D., Mazar, N., Mullainathan, S., Prelec, D., Shafir, E., and Silva, J. 2005. "Psychology, Behavioral Economics, and Public Policy," *Marketing Letters* (16:3-4), pp. 443–454 (doi: 10.1007/s11002-005-5904-2).

Andrade, E. B., and Ho, T.-H. 2009a. "Gaming Emotions in Social Interactions," *Journal of Consumer Research* (36:4), The Oxford University Press, pp. 539–552 (doi: 10.1086/599221).

Andrade, E. B., and Ho, T.-H. 2009b. "Gaming Emotions in Social Interactions," *Journal of Consumer Research* (36:4), The Oxford University Press, pp. 539–552 (doi: 10.1086/599221).

Angelova, V. 2013. "7 Reasons Why Social Media is an Absolute Must for Business Branding — likeable media," (available at http://www.likeable.com/blog/2013/09/7-reasons-why-social-media-is-an-absolute-must-for-business-branding; retrieved April 28, 2015).

Ansari, A., Koenigsberg, O., and Stahl, F. 2011. "Modeling multiple relationships in social networks," *Journal of Marketing Research* (48:4), American Marketing Association, pp. 713–728.

Aral, S., Muchnik, L., and Sundararajan, A. 2009. "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," *Proceedings of the National Academy of Sciences* (106:51), National Acad Sciences, pp. 21544–21549.

Aral, S., Muchnik, L., and Sundararajan, A. 2013. "Engineering social contagions: Optimal network seeding in the presence of homophily," *Network Science* (1:02), Cambridge Univ Press, pp. 125–153.

Aral, S., and Walker, D. 2012. "Identifying influential and susceptible members of social networks," *Science* (337:6092), American Association for the Advancement of Science, pp. 337–341.

Aral, S., and Walker, D. 2014. "Tie strength, embeddedness, and social influence: A large-scale networked experiment," *Management Science* (60:6), INFORMS, pp. 1352–1370.

Archak, N., Ghose, A., and Ipeirotis, P. G. 2011. "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science* (57:8), pp. 1485–1509 (doi: 10.1287/mnsc.1110.1370).

Argo, J. J., Dahl, D. W., and Manchanda, R. V. 2005. "The Influence of a Mere Social Presence in a Retail Context," *Journal of Consumer Research*, pp. 207–212 (doi: 10.1086/432230).

Argote, L., and Ingram, P. 2000. "Knowledge Transfer: A Basis for Competitive Advantage in Firms," *Organizational Behavior and Human Decision Processes* (82:1), Academic Press, pp. 150–169 (doi: 10.1006/obhd.2000.2893).

Asch, S. 1951. "Effects of group pressure upon the modification and distortion of judgments," *Groups, Leadership, and Men.*, pp. 222–236 (available at http://books.google.com.sg/books?hl=en&lr=&id=SOk8TFBYWHIC&oi=fnd&pg=PA222&dq=Effects+of+group+pressure+on+the+modification+and+distortion+of+judgments&ots=f3Pxv5F6C0&sig=ULiruWgOhguLsLWm1iskp_Zm1Ms).

Asur, S., and Huberman, B. A. 2010. "Predicting the Future with Social Media," in *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, IEEE, August 31, pp. 492–499 (doi: 10.1109/WI-IAT.2010.63).

Backstrom, L., Huttenlocher, D., Kleinberg, J., and Lan, X. 2006. "Group formation in large social networks," in *Proceedings of the 12th ACM SIGKDD international conference on*

Knowledge discovery and data mining - KDD '06, New York, New York, USA: ACM Press, August 20, p. 44 (doi: 10.1145/1150402.1150412).

Backstrom, L., and Leskovec, J. 2011. "Supervised random walks," in *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11*, New York, New York, USA: ACM Press, February 9, p. 635 (doi: 10.1145/1935826.1935914).

Bakshy, E., Messing, S., and Adamic, L. 2015. "Exposure to ideologically diverse news and opinion on Facebook," *Science* (forthcomin) (doi: 10.1126/science.aaa1160).

Banerjee, A. V. 1992. "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics* (107:3), Oxford University Press, pp. 797–817 (doi: 10.2307/2118364).

Bankston, K. 2009. "Facebook's New Privacy Changes: The Good, The Bad, and The Ugly | Electronic Frontier Foundation," *Electronic Frontier Foundation* (available at https://www.eff.org/deeplinks/2009/12/facebooks-new-privacy-changes-good-bad-and-ugly; retrieved July 4, 2015).

Barney, J. 1991. "Firm Resources and Sustained Competitive Advantage," *Journal of Management* (17:1), SAGE Publications, pp. 99–120 (doi: 10.1177/014920639101700108).

Barry, F. 2012. "Nonprofits Value a Facebook Like at $214.81 | npENGAGE," *npEngage.com* (available at http://npengage.com/nonprofit-fundraising/nonprofit-value-facebook-like/; retrieved April 28, 2015).

Baumeister, R. F. 1982. "A self-presentational view of social phenomena.," *Psychological Bulletin* (91:1), pp. 3–26 (doi: 10.1037//0033-2909.91.1.3).

Baumeister, R. F., and Jones, E. E. 1978. "When self-presentation is constrained by the target's knowledge: Consistency and compensation.," *Journal of Personality and Social Psychology* (36:6), p. 608.

Baumeister, R. F., Tice, D. M., and Hutton, D. G. 1989. "Self-presentational motivations and personality differences in self-esteem," *Journal of personality* (57:3), Wiley Online Library, pp. 547–579.

Baumeister, R., Hutton, D., and Tice, D. 1989. "Cognitive processes during deliberate self-presentation: How self-presenters alter and misinterpret the behavior of their interaction partners," *Journal of Experimental Social Psychology* (78), pp. 59–78 (available at http://www.sciencedirect.com/science/article/pii/0022103189900395).

Bélanger, F., and Crossler, R. E. 2011. “Privacy in the digital age: a review of information privacy research in information systems,” *MIS Quarterly* (35:4), Society for Information Management and The Management Information Systems Research Center, pp. 1017–1042 (available at http://dl.acm.org/citation.cfm?id=2208940.2208951).

Bellezza, S., Gino, F., and Keinan, A. 2014. “The red sneakers effect: Inferring status and competence from signals of nonconformity,” *Journal of Consumer Research* (41:1), The Oxford University Press, pp. 35–54.

Bennett, S. C. 2012. “Right to Be Forgotten: Reconciling EU and US Perspectives, The,” *Berkeley Journal of International Law* (30) (available at http://heinonline.org/HOL/Page?handle=hein.journals/berkjintlw30&id=163&div=&collection=).

Berkowitz, L., and Donnerstein, E. 1982. “External validity is more than skin deep: Some answers to criticisms of laboratory experiments.,” *American psychologist* (37:3), American Psychological Association, p. 245.

Bikhchandani, S., Hirshleifer, D., and Welch, I. 1992. “A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades,” *Journal of Political Economy* (100:5), The University of Chicago Press, pp. 992–1026 (doi: 10.2307/2138632).

Billy, J. O. G., and Udry, J. R. 1985. “Patterns of adolescent friendship and effects on sexual behavior,” *Social Psychology Quarterly*, JSTOR, pp. 27–41.

Binder, J., Howes, A., and Sutcliffe, A. 2009. “The problem of conflicting social spheres,” in *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, New York, New York, USA: ACM Press, April 4, p. 965 (doi: 10.1145/1518701.1518849).

Blaine, B., and Crocker, J. 1993. “Self-esteem and self-serving biases in reactions to positive and negative events: An integrative review,” in *Self-esteem*, Springer, pp. 55–85.

Blattberg, R. C., and Deighton, J. 1991. “Interactive marketing: exploiting the age of addressability,” *Sloan Management Review* (33).

Borgatti, S. P., and Foster, P. C. 2003. “The network paradigm in organizational research: A review and typology,” *Journal of management* (29:6), Sage Publications, pp. 991–1013.

Bowman, K. O., and Shenton, L. R. 1985. “Method of moments,” *Encyclopedia of statistical sciences* (5), Wiley New York, pp. 467–473.

boyd, danah m., and Ellison, N. B. 2007. "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication* (13:1), pp. 210–230 (doi: 10.1111/j.1083-6101.2007.00393.x).

boyd, danah michele. 2004. "Friendster and publicly articulated social networking," in *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04*, New York, New York, USA: ACM Press, April 24, p. 1279 (doi: 10.1145/985921.986043).

Boyd, D. 2008. "Facebook's Privacy Trainwreck," *Convergence: The International Journal of Research into New Media Technologies* (14:1), pp. 13–20.

Brandimarte, L., Acquisti, A., and Loewenstein, G. 2012. "Misplaced Confidences: Privacy and the Control Paradox," *Social Psychological and Personality Science* (4:3), pp. 340–347 (doi: 10.1177/1948550612455931).

Braun, M., and Bonfrer, A. 2011. "Scalable inference of customer similarities from interactions data using Dirichlet processes," *Marketing Science* (30:3), INFORMS, pp. 513–531.

Brickell, J., and Shmatikov, V. 2008. "The cost of privacy," in *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, New York, New York, USA: ACM Press, August 24, p. 70 (doi: 10.1145/1401890.1401904).

Brook, J. S., Whiteman, M., and Gordon, A. S. 1983. "Stages of drug abuse in adolescence: Personality, peer, and family correlates.," *Developmental Psychology* (19:2), American Psychological Association, p. 269.

Bughin, J. 2007. "How companies can make the most of user-generated content," *McKinsey Quarterly* (august) (available at http://www.inovacaoedesign.com.br/artigos_cientificos/hoca07.pdf).

Bullas, J. 2015. "33 Social Media Facts and Statistics You Should Know in 2015," *jeffbullas.com* (available at http://www.jeffbullas.com/2015/04/08/33-social-media-facts-and-statistics-you-should-know-in-2015/).

De Bunt, G. G., Van Duijn, M. A. J., and Snijders, T. A. B. 1999. "Friendship networks through time: An actor-oriented dynamic statistical network model," *Computational & Mathematical Organization Theory* (5:2), Springer, pp. 167–192.

Campbell, A. J. 1997. "Relationship marketing in consumer markets: A comparison of managerial and consumer attitudes about information privacy," *Journal of Direct*

Marketing (11:3), pp. 44–57 (doi: 10.1002/(SICI)1522-7138(199722)11:3<44::AID-DIR7>3.0.CO;2-X).

Campbell, J. E., and Carlson, M. 2002. “Panopticon. com: Online surveillance and the commodification of privacy,” *Journal of Broadcasting & Electronic Media* (46:4), Taylor & Francis, pp. 586–606.

Carley, K. 1991. “A theory of group stability,” *American Sociological Review* (56:3), JSTOR, pp. 331–354.

Carpenter, C. J. 2012. “Narcissism on Facebook: Self-promotional and anti-social behavior,” *Personality and Individual Differences* (52:4), pp. 482–486 (available at http://www.sciencedirect.com/science/article/pii/S0191886911005332).

Chellappa, R. K., and Sin, R. G. 2005. “Personalization versus privacy: An empirical examination of the online consumer’s dilemma,” *Information Technology and Management* (6:2-3), Springer, pp. 181–202.

Chen, H., Chiang, R., and Storey, V. 2012. “Business Intelligence and Analytics: From Big Data to Big Impact.,” *MIS Quarterly* (36:4), pp. 1165–1188 (available at https://noppa.aalto.fi/noppa/kurssi/37e44000/materiaali/37E44000_3_generations_of_business_analytics_overview.pdf).

Chen, Y., Wang, Q., and Xie, J. 2011. “Online Social Interactions: A Natural Experiment on Word of Mouth Versus Observational Learning,” *Journal of Marketing Research* (48:2), American Marketing Association, pp. 238–254 (doi: 10.1509/jmkr.48.2.238).

Cheng, J. 2009. “An updated guide to Facebook privacy: December 2009 edition | Ars Technica,” *arstechnica* (available at http://arstechnica.com/business/2009/12/an-updated-guide-to-facebook-privacy-december-2009-edition/; retrieved July 4, 2015).

Chevalier, J. A., and Mayzlin, D. 2006. “The Effect of Word of Mouth on Sales: Online Book Reviews,” *Journal of Marketing Research* (43:3), American Marketing Association, pp. 345–354 (doi: 10.1509/jmkr.43.3.345).

Chintagunta, P. K., Gopinath, S., and Venkataraman, S. 2010. “The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets,” *Marketing Science* (29:5), INFORMS, pp. 944–957 (doi: 10.1287/mksc.1100.0572).

Christakis, N. A., and Fowler, J. H. 2008. “The collective dynamics of smoking in a large social network,” *New England journal of medicine* (358:21), Mass Medical Soc, pp. 2249–2258.

Cialdini, R. B., and Goldstein, N. J. 2004. “Social influence: compliance and conformity.,” *Annual review of psychology* (55:1), pp. 591–621 (doi: 10.1146/annurev.psych.55.090902.142015).

Cioffi-Revilla, C. 2010. “Computational social science,” *Wiley Interdisciplinary Reviews: Computational Statistics* (2:3), pp. 259–271 (doi: 10.1002/wics.95).

Cohen, J. M. 1977. “Sources of peer group homogeneity,” *Sociology of Education* (50:4), JSTOR, pp. 227–241.

Conti, G., and Sobiesk, E. 2010. “Malicious interface design,” in *Proceedings of the 19th international conference on World wide web - WWW ’10*, New York, New York, USA: ACM Press, April 26, p. 271 (doi: 10.1145/1772690.1772719).

Cook, T. D., Campbell, D. T., and Peracchio, L. 1990. “Quasi experimentation.,” Consulting Psychologists Press.

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., and Suri, S. 2008. “Feedback effects between similarity and social influence in online communities,” in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 160–168.

Creamer, M. 2012. “Study: Only 1% of Facebook ‘Fans’ Engage With Brands | Digital - Advertising Age,” (available at http://adage.com/article/digital/study-1-facebook-fans-engage-brands/232351/; retrieved April 28, 2015).

Davies, S. G. 1997. “Re-engineering the right to privacy: how privacy has been transformed from a right to a commodity,” in *Technology and privacy*, pp. 143–165.

Davin, J. P., Davin, J. P., Gupta, S., and Piskorski, M. J. 2014. “Separating homophily and peer influence with latent space Separating homophily and peer influence with latent space.,”

Dekel, E. 2013. “Facebook Pages Are a Bad Investment for Small Businesses,” *Forbes* (available at http://www.forbes.com/sites/elandekel/2013/01/22/facebook-pages-are-a-bad-investment-for-small-businesses/#5723b1bc4259).

Dellarocas, C. 2003. “The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms,” *Management Science* (49:10), INFORMS, pp. 1407–1424 (available at http://pubsonline.informs.org/doi/abs/10.1287/mnsc.49.10.1407.17308).

Dhar, V., and Chang, E. A. 2009. "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing* (23:4), pp. 300–307 (available at http://www.sciencedirect.com/science/article/pii/S1094996809000723).

Dhillon, G. S., and Moores, T. T. 2003. "Internet privacy: Interpreting key issues," *Advanced topics in information resources management*, IGI Publishing, pp. 52–61.

Diener, E., and Diener, C. 1996. "Most people are happy," *Psychological science* (7:3), SAGE Publications, pp. 181–185.

Dinev, T., Bellotto, M., Hart, P., Russo, V., Serra, I., and Colautti, C. 2006. "Privacy calculus model in e-commerce--a study of Italy and the United States," *European Journal of Information Systems* (15:4), Nature Publishing Group, pp. 389–402.

Dinev, T., and Hart, P. 2006. "An extended privacy calculus model for e-commerce transactions," *Information Systems Research* (17:1), INFORMS, pp. 61–80.

Donath, J., and Boyd, D. 2004. "Public Displays of Connection," *BT Technology Journal* (22:4), Kluwer Academic Publishers, pp. 71–82 (doi: 10.1023/B:BTTJ.0000047585.06264.cc).

Doreian, P. 1989. "Network Autocorrelation Models: Problems and Prospects," *Spatial Statistics: Past, Present, Future*, pp. 369–389.

Dunbar, R. I. M. 1992. "Neocortex size as a constraint on group size in primates," *Journal of Human Evolution* (22:6), pp. 469–493 (doi: 10.1016/0047-2484(92)90081-J).

Durkheim, E. 1884. "The division of labor in society," *Journal des Economistes*, p. 211.

Edwards, J. 2012. "Meet The 30 Biggest Social Media Advertisers Of 2012 [Ranked]," (available at http://www.businessinsider.com/the-30-biggest-social-media-advertisers-of-2012-2012-9?op=1#ixzz2S8ayIUHf; retrieved April 29, 2015).

Eisenmann, T., Parker, G., and Van Alstyne, M. W. 2006. "Strategies for two-sided markets," *Harvard business review* (84:10), p. 92.

Emirbayer, M., and Goodwin, J. 1994. "Network analysis, culture, and the problem of agency," *American journal of sociology*, JSTOR, pp. 1411–1454.

Ennett, S. T., and Bauman, K. E. 1994. "The contribution of influence and selection to adolescent peer group homogeneity: the case of adolescent cigarette smoking.," *Journal of personality and social psychology* (67:4), American Psychological Association, p. 653.

Epstein, S. 1973. “The self-concept revisited: Or a theory of a theory.,” *American psychologist* (28:5), American Psychological Association, p. 404.

Falk, A., and Heckman, J. J. 2009. “Lab experiments are a major source of knowledge in the social sciences,” *science* (326:5952), American Association for the Advancement of Science, pp. 535–538.

Fararo, T. J., and Sunshine, M. H. 1964. *A study of a biased friendship net*, Youth Development Center, Syracuse University.

Feld, S. L. 1981. “The focused organization of social ties,” *American journal of sociology* (86:5), JSTOR, pp. 1015–1035.

Feld, S. L. 1982. “Social structural determinants of similarity among associates,” *American Sociological Review* (47:6), JSTOR, pp. 797–801.

Feld, S. L., and Elmore, R. 1982. “Patterns of sociometric choices: transitivity reconsidered,” *Social Psychology Quarterly* (45:2), JSTOR, pp. 77–85.

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., and Combs, B. 1978. “How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits,” *Policy Sciences* (9:2), pp. 127–152 (doi: 10.1007/BF00143739).

Fisher, L. A., and Bauman, K. E. 1988. “Influence and Selection In the Friend-adolescent Relationship: Findings from Studies of Adolescent Smoking and Drinking1,” *Journal of Applied Social Psychology* (18:4), pp. 289–314 (doi: 10.1111/j.1559-1816.1988.tb00018.x).

Fogg, B. J., and Eckles, D. 2007. “The behavior chain for online participation: how successful web services structure persuasion,” Springer-Verlag, pp. 199–209 (available at http://dl.acm.org/citation.cfm?id=1780402.1780435).

Friedkin, N. E. 2001. “Norm formation in social influence networks,” *Social networks* (23:3), Elsevier, pp. 167–189.

Gerber, S. 2014. “9 Ways Social Media Marketing Will Change in 2014,” (available at http://mashable.com/2014/01/27/social-media-marketing-2014/; retrieved April 28, 2015).

Ghose, A., and Ipeirotis, P. G. 2011. “Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics,” *Knowledge and Data Engineering, IEEE Transactions on* (23:10), IEEE, pp. 1498–1512 (doi: 10.1109/TKDE.2010.188).

Ghose, A., Ipeirotis, P. G., and Li, B. 2012. “Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content,” *Marketing Science* (31:3), pp. 493–520 (doi: 10.1287/mksc.1110.0700).

Gino, F., and Mogilner, C. 2013. “Time, Money, and Morality.,” *Psychological science*, p. 0956797613506438– (doi: 10.1177/0956797613506438).

Gobry, P.-E. 2012. “Think Facebook’s $100 Billion Valuation Is Nuts? Read This,” *Business Insider* (available at http://www.businessinsider.com/facebook-valuation-2012-1?IR=T&r=US&IR=T).

Godes, D., and Mayzlin, D. 2004. “Using Online Conversations to Study Word-of-Mouth Communication,” *Marketing Science* (23:4), pp. 545–560 (doi: 10.1287/mksc.1040.0071).

Goffman, E. 1959a. *The Presentation of Self in Everyday Life*, Anchor (available at http://www.amazon.com/The-Presentation-Self-Everyday-Life/dp/0385094027).

Goffman, E. 1959b. *The presentation of self in everyday life*, Doubleday (available at http://books.google.com.sg/books/about/The_presentation_of_self_in_everyday_lif.html?id=Sdt-cDkV8pQC&pgis=1).

Goh, K., Heng, C., and Lin, Z. 2013. “Social Media Brand Community and Consumer Behavior : Quantifying the Relative Impact of User- and Marketer- Generated Content Social Media Brand Community and Consumer Behavior : Quantifying the Relative Impact of User- and Marketer-Generated Content,” (December).

Goh, K. Y., Heng, C. S., and Lin, Z. 2013. “Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content,” *Information Systems Research* (24:1), pp. 88–107 (doi: 10.1287/isre.1120.0469).

Goldfarb, A., and Tucker, C. 2012. “Shifts in privacy concerns,” *The American Economic Review* (102:3), American Economic Association, pp. 349–353.

Goldfarb, A., and Tucker, C. E. 2011. “Privacy Regulation and Online Advertising,” *Management Science* (57:1), pp. 57–71 (doi: 10.1287/mnsc.1100.1246).

Gonzales, A. L., and Hancock, J. T. 2011. “Mirror, mirror on my Facebook wall: Effects of exposure to Facebook on self-esteem,” *Cyberpsychology, Behavior, and Social Networking* (14:1-2), Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, pp. 79–83.

Greenan, C. C. 2015. “Diffusion of innovations in dynamic networks,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* (178:1), Wiley Online Library, pp. 147–166.

Greenberg, J., and Pyszczynski, T. 1985. “Compensatory self-inflation: A response to the threat to self-regard of public failure.,” *Journal of Personality and Social Psychology* (49:1), American Psychological Association, p. 273.

Greenwald, A. G. 1980. “The totalitarian ego: Fabrication and revision of personal history.,” *American psychologist* (35:7), American Psychological Association, p. 603.

Gross, R., and Acquisti, A. 2005. “stutz,” in *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pp. 71–80.

Gurevitch, Z. 1984. “Impression formation during tactical self-presentation,” *Social Psychology Quarterly* (47:3), pp. 262–270 (available at http://www.jstor.org/stable/10.2307/3033823).

Hargittai, E. 2007. “Whose Space? Differences Among Users and Non-Users of Social Network Sites,” *Journal of Computer-Mediated Communication* (13:1), pp. 276–297 (doi: 10.1111/j.1083-6101.2007.00396.x).

Heine, S. J., Lehman, D. R., Markus, H. R., and Kitayama, S. 1999. “Is there a universal need for positive self-regard?,” *Psychological review* (106:4), American Psychological Association, p. 766.

Herman, C. P., Roth, D. A., and Polivy, J. 2003. “Effects of the presence of others on food intake: a normative interpretation.,” *Psychological bulletin* (129:6), American Psychological Association, p. 873.

Hoadley, C. M., Xu, H., Lee, J. J., and Rosson, M. B. 2010. “Privacy as information access and illusory control: The case of the Facebook News Feed privacy outcry,” *Electronic Commerce Research and Applications* (9:1), pp. 50–60 (doi: 10.1016/j.elerap.2009.05.001).

Hof, R. D. 2011. “Advertisers Flock to Social Networks | MIT Technology Review,” (available at http://www.technologyreview.com/article/424409/advertisers-flock-to-social-networks/; retrieved April 29, 2015).

Hoffman, D. L., and Fodor, M. 2010. “Can you measure the ROI of your social media marketing?,” *MIT Sloan Management Review* (52:1), Massachusetts Institute of Technology, p. 41.

Hogan, B. 2010. “The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online,” *Bulletin of Science, Technology & Society* (30:6), pp. 377–386 (doi: 10.1177/0270467610385893).

Holland, P. W., and Leinhardt, S. 1977. “A dynamic model for social networks†,” *Journal of Mathematical Sociology* (5:1), Taylor & Francis, pp. 5–20.

Hollingshead, A. de B. 1949. *Elmtown’s Youth: the Impact of Social Classes on Adolescent*, New York: Wiley.

Holtz-Eakin, D., Newey, W., and Rosen, H. S. 1988. “Estimating vector autoregressions with panel data,” *Econometrica: Journal of the Econometric Society*, JSTOR, pp. 1371–1395.

Homans, G. C. 1961. “Social behavior: Its elementary forms.,” Harcourt, Brace.

Hui, K.-L., Tan, B. C. Y., and Goh, C.-Y. 2006. “Online information disclosure: Motivators and measurements,” *ACM Transactions on Internet Technology (TOIT)* (6:4), ACM, pp. 415–441.

Hui, K.-L., Teo, H. H., and Lee, S.-Y. T. 2007. “The value of privacy assurance: an exploratory field experiment,” *Mis Quarterly*, JSTOR, pp. 19–33.

Hyman, H. H. 1942. “The psychology of status,” *Archives of Psychology* (269).

Iannotti, R. J., Bush, P. J., and Weinfurt, K. P. 1996. “Perception of friends’ use of alcohol, cigarettes, and marijuana among urban schoolchildren: A longitudinal analysis,” *Addictive Behaviors* (21:5), pp. 615–632 (doi: 10.1016/0306-4603(95)00086-0).

Ismail, W. 2011. “15 Must Have Features on an E-Commerce Site,” *searchenginejournal.com* (available at https://www.searchenginejournal.com/15-must-have-features-ecommerce-site/27233/).

James, W. 2013. *The principles of psychology*, Read Books Ltd.

Johnson, E. J., Bellman, S., and Lohse, G. L. 2002. “Defaults, Framing and Privacy: Why Opting In-Opting Out,” *Marketing Letters* (13:1), Kluwer Academic Publishers, pp. 5–15 (doi: 10.1023/A:1015044207315).

Johnson, E. J., and Goldstein, D. 2003. “Medicine. Do defaults save lives?,” *Science (New York, N.Y.)* (302:5649), pp. 1338–9 (doi: 10.1126/science.1091721).

Jones, E. E., and Wortman, C. B. 1973. *Ingratiation: An Attributional Approach*, General Learning Press (available at http://books.google.com.sg/books/about/Ingratiation.html?id=Nf-EGQAACAAJ&pgis=1).

Jones, H., and Soltren, J. H. 2005. "Facebook: Threats to privacy," *Project MAC: MIT Project on Mathematics and Computing* (1), pp. 1–76.

Kalehoff, M. 2013. "The Reasons Consumers Become Facebook Brand Fans | Syncapse," *Syncapse.com* (available at http://www.syncapse.com/why-consumers-become-facebook-brand-fans/#.VVa-wJNCjnf; retrieved May 16, 2015).

Kandel, D. B. 1978. "Homophily, Selection, and Socialization in Adolescent Friendships.," *American Journal of Sociology* (84:2), pp. 427–436 (available at http://eric.ed.gov/?id=EJ187683).

Kang, S. C., Morris, S. S., and Snell, S. A. 2007. "Relational archetypes, organizational learning, and value creation: Extending the human resource architecture," *Academy of Management Review*, Academy of Management, pp. 236–256 (doi: 10.5465/AMR.2007.23464060).

Katz, L., and Proctor, C. H. 1959. "The concept of configuration of interpersonal relations in a group as a time-dependent stochastic process," *Psychometrika* (24:4), Springer, pp. 317–327.

Kim, H., and Papacharissi, Z. 2003. "Cross-cultural differences in online self-presentation: A content analysis of personal Korean and US home pages," *Asian Journal of Communication* (13:1), pp. 100–119 (doi: 10.1080/01292980309364833).

Kim, H. W., Chan, H. C., and Kankanhalli, A. 2012. "What motivates people to purchase Digital Items on virtual community websites? The desire for online self-presentation," *Information Systems Research* (23:4), pp. 1232–1245 (doi: 10.1287/isre.1110.0411).

Kim, J., and Lee, J.-E. R. 2011. "The Facebook paths to happiness: effects of the number of Facebook friends and self-presentation on subjective well-being.," *Cyberpsychology, behavior and social networking* (14:6), Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, pp. 359–64 (doi: 10.1089/cyber.2010.0374).

Kincaid, J. 2009. "The Facebook Privacy Fiasco Begins," *Techcrunch* (available at http://techcrunch.com/2009/12/09/facebook-privacy/; retrieved July 4, 2015).

Kirke, D. M. 2004. "Chain reactions in adolescents' cigarette, alcohol and drug use: similarity through peer influence or the patterning of ties in peer networks?," *Social Networks* (26:1), pp. 3–28 (doi: 10.1016/j.socnet.2003.12.001).

Klopfer, P. H., and Rubenstein, D. I. 1977. “The concept privacy and its biological basis,” *Journal of social Issues* (33:3), Wiley Online Library, pp. 52–65.

Kouchaki, M., and Smith, I. H. 2013. “The morning morality effect the influence of time of day on unethical behavior,” *Psychological science*, Sage Publications, p. 0956797613498099.

Krasnova, H., and Veltri, N. F. 2010. “Privacy Calculus on Social Networking Sites: Explorative Evidence from Germany and USA,” in *2010 43rd Hawaii International Conference on System Sciences*, IEEE, pp. 1–10 (doi: 10.1109/HICSS.2010.307).

Krohn, M. D., Lizotte, A. J., Thornberry, T. P., Smith, C., and McDowall, D. 1996. “Reciprocal causal relationships among drug use, peers, and beliefs: A five-wave panel model,” *Journal of Drug Issues* (26), pp. 405–428.

Kumar, V., Vikram, B., Rohan, M., and Milap, S. 2013. “Practice Prize Winner — Creating a Measurable Social Intangibles and Tangibles for Hokey Pokey Creating a Measurable Social Media Marketing Strategy : Increasing the Value and ROI of Intangibles and Tangibles for Hokey Pokey,” *Marketing Science* (32:April 2015), pp. 194–212 (doi: 10.1287/mksc.1120.0768).

Lampe, C., Ellison, N., and Steinfield, C. 2006. “A Face (book) in the crowd: Social searching vs. social browsing,” in *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pp. 167–170.

Laroche, M., Habibi, M. R., and Richard, M.-O. 2013. “To be or not to be in social media: How brand loyalty is affected by social media?,” *International Journal of Information Management* (33:1), pp. 76–82 (doi: 10.1016/j.ijinfomgt.2012.07.003).

LaRose, R., and Eastin, M. S. 2004. “A Social Cognitive Theory of Internet Uses and Gratifications: Toward a New Model of Media Attendance,” *Journal of Broadcasting & Electronic Media* (48:3), Routledge, pp. 358–377 (doi: 10.1207/s15506878jobem4803_2).

Laserna, C. M., Seih, Y.-T., and Pennebaker, J. W. 2014. “Um... Who Like Says You Know Filler Word Use as a Function of Age, Gender, and Personality,” *Journal of Language and Social Psychology*, SAGE Publications, p. 0261927X14526993.

Laudon, K. C. 1996. “Markets and privacy,” *Communications of the ACM* (39:9), ACM, pp. 92–104.

Laufer, R. S., and Wolfe, M. 1977. “Privacy as a concept and a social issue: A multidimensional developmental theory,” *Journal of Social Issues* (33:3), Wiley Online Library, pp. 22–42.

Lazarsfeld, P. F., Merton, R. K., and others. 1954. “Friendship as a social process: A substantive and methodological analysis,” *Freedom and control in modern society* (18:1), pp. 18–66.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. 2014. “Big data. The parable of Google Flu: traps in big data analysis.,” *Science (New York, N.Y.)* (343:6176), pp. 1203–5 (doi: 10.1126/science.1248506).

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., and van Alstyne, M. 2009. “Computational social science,” *Science* (323:February), pp. 721–723 (doi: 10.1126/science.1167742).

Leary, M., and Kowalski, R. 1990. “Impression management: A literature review and two-component model.,” *Psychological bulletin* (107:I), pp. 34–47 (available at http://psycnet.apa.org/journals/bul/107/1/34/).

Leary, M. R. 1996. *Self-presentation: Impression Management And Interpersonal Behavior (Social Psychology Series)*, Westview Press (available at http://www.amazon.com/Self-presentation-Impression-Management-Interpersonal-Psychology/dp/0813330041).

Lee, D., Hosanagar, K., and Nair, H. S. 2014. “The E ff ect of Social Media Marketing Content on Consumer Engagement : Evidence from Facebook,” *Working Paper* (Summer 2013).

Lepak, D. P., Smith, K. G., and Taylor, M. S. 2007. “Value creation and value capture: A multilevel perspective,” *Academy of Management Review*, Academy of Management, pp. 180–194 (doi: 10.5465/AMR.2007.23464011).

Levine, E. L., Xu, X., Yang, L.-Q., Ispas, D., Pitariu, H. D., Bian, R., Ding, D., Capotescu, R., Che, H., and Musat, S. 2011. “Cross-National Explorations of the Impact of Affect at Work Using the State-Trait Emotion Measure: A Coordinated Series of Studies in the United States, China, and Romania,” *Human Performance* (24:5), Routledge, pp. 405–442 (doi: 10.1080/08959285.2011.614302).

Lewis, K., Gonzalez, M., and Kaufman, J. 2012. “Social selection and peer influence in an online social network.,” *Proceedings of the National Academy of Sciences of the United States of America* (109:1), pp. 68–72 (doi: 10.1073/pnas.1109739109).

Lewis, K., Kaufman, J., and Christakis, N. 2008. “The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network,” *Journal of Computer-Mediated Communication* (14:1), pp. 79–100 (doi: 10.1111/j.1083-6101.2008.01432.x).

Li, N., and Wu, D. D. 2010. “Using text mining and sentiment analysis for online forums hotspot detection and forecast,” *Decision Support Systems* (48:2), pp. 354–368 (doi: 10.1016/j.dss.2009.09.003).

Li, T., and Li, N. 2009. “On the tradeoff between privacy and utility in data publishing,” in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’09*, New York, New York, USA: ACM Press, June 28, p. 517 (doi: 10.1145/1557019.1557079).

Liu, Y. 2006. “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *Journal of Marketing* (70:3), American Marketing Association, pp. 74–89 (doi: 10.1509/jmkg.70.3.74).

Liu, Y., Chen, Y., Lusch, R. F., Chen, H., Zimbra, D., and Zeng, S. 2010. “User-Generated content on social media: Predicting market success with online word-of-mouth,” *IEEE Intelligent Systems* (25:6), IEEE COMPUTER SOC 10662 LOS VAQUEROS CIRCLE, PO BOX 3014, LOS ALAMITOS, CA 90720-1314 USA, pp. 8–12.

Livingstone, S. 2008. “Taking risky opportunities in youthful content creation: teenagers’ use of social networking sites for intimacy, privacy and self-expression,” *New Media & Society* (10:3), pp. 393–411 (doi: 10.1177/1461444808089415).

Loss, J., Lindacher, V., and Curbach, J. 2013. “Do Social Networking Sites Enhance the Attractiveness of Risky Health Behavior? Impression Management in Adolescents’ Communication on Facebook and its Ethical Implications,” *Public Health Ethics* (doi: 10.1093/phe/pht028).

Lu, Y., Jerath, K., and Singh, P. V. 2013. “The emergence of opinion leaders in a networked online community: A dyadic model with time dynamics and a heuristic for fast estimation,” *Management Science* (59:8), INFORMS, pp. 1783–1799.

Luo, X., Zhang, J., and Duan, W. 2012. “Social Media and Firm Equity Value,” *Information Systems Research* (24:1), pp. 146–163 (doi: 10.1287/isre.1120.0462).

Ma, M., and Agarwal, R. 2007. “Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities,” *Information Systems Research* (18:1), pp. 42–67.

Macy, M. W., Kitts, J. A., Flache, A., and Benard, S. 2003. “Polarization in dynamic networks: A Hopfield model of emergent structure,” *Dynamic social network modeling and analysis*, National Academies Press (Washington, 2003), pp. 162–173.

Maddala, G. S. 1986. *Limited-dependent and qualitative variables in econometrics*, Cambridge university press.

Malhotra, N. K., Kim, S. S., and Agarwal, J. 2004. “Internet users’ information privacy concerns (IUIPC): the construct, the scale, and a causal model,” *Information Systems Research* (15:4), Informs, pp. 336–355.

Manski, C. F. 1993. “Identification of Endogenous Social Effects: The Reflection Problem,” *The Review of Economic Studies* (60:3), p. 531 (doi: 10.2307/2298123).

Marwick, A. E. 2005. “‘I’m a Lot More Interesting than a Friendster Profile’: Identity Presentation, Authenticity and Power in Social Networking Services,” (available at http://papers.ssrn.com/abstract=1884356).

Marwick, A. E., and Boyd, D. 2011. “I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience,” *New Media & Society* (13:1), pp. 114–133 (doi: 10.1177/1461444810365313).

Marx, G. T. 2001. “Murky conceptual waters: The public and the private,” *Ethics and Information Technology* (3:3), Kluwer Academic Publishers, pp. 157–169 (doi: 10.1023/A:1012456832336).

Maslow, A. H. 1943. “A theory of human motivation.,” *Psychological review* (50:4), American Psychological Association, p. 370.

May, K. 2011. “Why do consumers follow or like travel brands in social media? For the discounts, alas - Tnooz,” *Tnooz* (available at http://www.tnooz.com/article/why-do-consumers-follow-or-like-travel-brands-in-social-media-for-the-discounts-alas/; retrieved July 4, 2015).

Mayer, T. F. 1984. “Parties and networks: stochastic models for relationship networks,” *Journal of Mathematical Sociology* (10:1), Taylor & Francis, pp. 51–103.

McPherson, M., Smith-Lovin, L., and Cook, J. M. 2001. “Birds of a feather: Homophily in social networks,” *Annual review of sociology* (27), JSTOR, pp. 415–444.

Mehdizadeh, S. 2010. “Self-presentation 2.0: narcissism and self-esteem on Facebook.,” *Cyberpsychology, behavior and social networking* (13:4), pp. 357–64 (doi: 10.1089/cyber.2009.0257).

Metaxas, P. T., and Mustafaraj, E. 2012. “Social Media and the Elections,” *Science* (338:6106), American Association for The Advancement of Science, Washington, USA, pp. 472–473.

Michell, L., and Pearson, M. 2000. “Smoke rings: social network analysis of friendship groups, smoking and drug-taking,” *Drugs: Education, Prevention, and Policy* (7:1), Informa UK Ltd UK, pp. 21–37.

La Monica, P. R. 2016. "Why Facebook could one day be worth $1 trillion," *CNN Money* (available at http://money.cnn.com/2016/04/28/investing/facebook-trillion-dollar-market-value/).

Moorman, C. 2013. "Survey Results - August 2013 | The CMO Survey," *CMO Survey* (available at http://cmosurvey.org/results/survey-results-august-2013/; retrieved July 3, 2015).

Moran, P., and Ghoshal, S. 1996. "Value Creation by Firms," *Academy of Management Proceedings* (1996:1), Academy of Management, pp. 41–45 (doi: 10.5465/AMBPP.1996.4978175).

Muntinga, D. G., Moorman, M., and Smit, E. G. 2015a. "Introducing COBRAs," *International Journal of Advertising* (30:1), Routledge, pp. 13–46.

Muntinga, D. G., Moorman, M., and Smit, E. G. 2015b. "Introducing COBRAs," *International Journal of Advertising* (30:1), Routledge, pp. 13–46 (available at http://www.tandfonline.com/doi/abs/10.2501/IJA-30-1-013-046#.VcHL4_nrCjw).

Nadkarni, A., and Hofmann, S. G. 2012. "Why Do People Use Facebook?," *Personality and individual differences* (52:3), Elsevier Ltd, pp. 243–249 (doi: 10.1016/j.paid.2011.11.007).

Nahon, K., and Hemsley, J. 2014. "Homophily in the Guise of Cross-Linking Political Blogs and Content," *American Behavioral Scientist* (58:10), SAGE Publications, pp. 1294–1313.

Naylor, R. W., Lamberton, C. P., and West, P. M. 2012. "Beyond the 'like' button: the impact of mere virtual presence on brand evaluations and purchase intentions in social media settings," *Journal of Marketing* (76:6), American Marketing Association, pp. 105–120.

Newcomb, T. M. 1962. "Student Peer-Group Influence.," John Wiley & Sons Inc.

Newman, M. 2010. *Networks: An Introduction*, Oxford University Press (available at http://books.google.com/books?hl=en&lr=&id=LrFaU4XCsUoC&pgis=1).

Noyes, D. 2015. "Top 20 Facebook Statistics - Updated February 2015," *The Top 20 Valuable Facebook Statistics* (available at https://zephoria.com/top-15-valuable-facebook-statistics/; retrieved May 4, 2015).

O'Connor, B., Balasubramanyan, R., Routledge, B. R., and Smith, N. a. 2010. "From tweets to polls: Linking text sentiment to public opinion time series," in *ICWSM*, pp. 122–129 (available at http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewPDFInterstitial/1536/1842).

Oestreicher-Singer, G., and Zalmanson, L. 2012. "Content or community? A digital business strategy for content providers in the social age," *A Digital Business Strategy for Content Providers in the Social Age (July 01, 2012)*.

Oetting, E. R., and Donnermeyer, J. F. 1998. "Primary socialization theory: The etiology of drug use and deviance. I," *Substance use & misuse* (33:4), Informa UK Ltd UK, pp. 995–1026.

Opsahl, K. 2010. "Facebook's Eroding Privacy Policy: A Timeline | Electronic Frontier Foundation," *Electronic Frontier Foundation* (available at https://www.eff.org/deeplinks/2010/04/facebook-timeline/; retrieved July 4, 2015).

Papacharissi, Z. 2009. "The virtual geographies of social networks: a comparative analysis of Facebook, LinkedIn and ASmallWorld," *New Media & Society* (11:1-2), pp. 199–220 (doi: 10.1177/1461444808099577).

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., and Seligman, M. E. P. 2015. "Automatic personality assessment through social media language.," *Journal of personality and social psychology* (108:6), American Psychological Association, p. 934.

Park, J., and Barabási, A.-L. 2007. "Distribution of node characteristics in complex networks," *Proceedings of the National Academy of Sciences* (104:46), National Acad Sciences, pp. 17916–17920.

Park, N., Kee, K. F., and Valenzuela, S. 2009. "Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes.," *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society* (12:6), pp. 729–33 (doi: 10.1089/cpb.2009.0003).

Pearson, M., and West, P. 2003. "Drifting smoke rings," *Connections* (25:2), pp. 59–76.

Peppard, J., Lambert, R., and Edwards, C. 2000. "Whose job is it anyway?: organizational information competencies for value creation," *Information Systems Journal* (10:4), Blackwell Science Ltd, pp. 291–322 (doi: 10.1046/j.1365-2575.2000.00089.x).

Petronio, S. 2012. *Boundaries of privacy: Dialectics of disclosure*, Suny Press.

Phan, T. Q., and Airoldi, E. M. 2015. "A natural experiment of social network formation and dynamics," *Proceedings of the National Academy of Sciences* (112:21), National Acad Sciences, pp. 6595–6600.

Phang, C. W., Tan, C.-H., Sutanto, J., Magagna, F., and Lu, X. 2014. “Leveraging O2O commerce for product promotion: an empirical investigation in Mainland China,” *Engineering Management, IEEE Transactions on* (61:4), IEEE, pp. 623–632.

Polyak, B. T. 1990. “A new method of stochastic approximation type,” *Avtomatika i telemekhanika* (7), Russian Academy of Sciences, Branch of Power Industry, Machine Building, Mechanics and Control Processes, pp. 98–107.

Posner, R. A. 1981. “The economics of privacy,” *The American economic review* (71:2), JSTOR, pp. 405–409.

Raacke, J., and Bonds-Raacke, J. 2008. “MySpace and Facebook: applying the uses and gratifications theory to exploring friend-networking sites.,” *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society* (11:2), pp. 169–74 (doi: 10.1089/cpb.2007.0056).

Rampell, A. 2010. “Why Online2Offline Commerce Is A Trillion Dollar Opportunity,” *Techcrunch* (available at http://techcrunch.com/2010/08/07/why-online2offline-commerce-is-a-trillion-dollar-opportunity/).

Rastogi, V., Suciu, D., and Hong, S. 2007. “The boundary between privacy and utility in data publishing,” VLDB Endowment, pp. 531–542 (available at http://dl.acm.org/citation.cfm?id=1325851.1325913).

Ray, A. 2010. “What Is The Value Of A Facebook Fan? Zero! | Forrester Blogs,” (available at http://blogs.forrester.com/augie_ray/10-07-08-what_value_facebook_fan_zero; retrieved April 28, 2015).

Reingen, P. H. et al. 1984. “Brand Congruence in Interpersonal Relations: A Social Network Analysis,” *Journal of Consumer Research* (11:3), University of Chicago Press, pp. 771–83 (available at http://econpapers.repec.org/RePEc:ucp:jconrs:v:11:y:1984:i:3:p:771-83).

Rishika, R., Kumar, A., Janakiraman, R., and Bezawada, R. 2013. “The effect of customers’ social media participation on customer visit frequency and profitability: An empirical investigation,” *Information Systems Research* (24:January 2015), pp. 108–127 (doi: 10.1287/isre.1120.0460).

Robbins, H., and Monro, S. 1951. “A stochastic approximation method,” *The annals of mathematical statistics* (22:3), JSTOR, pp. 400–407.

Rosen, J. 2012. “The Right to Be Forgotten,” *Stanford Law Review Online* (64), p. 88 (available at http://www.stanfordlawreview.org/online/privacy-paradox/right-to-be-forgotten?em_x=22).

Rosenbaum, P. R. 2002. *Observational studies*, Springer.

Rosenberg, M. J. 1965. "When dissonance fails: On eliminating evaluation apprehension from attitude measurement.," *Journal of personality and social psychology* (95), pp. 28–42 (available at http://www.ncbi.nlm.nih.gov/pubmed/14328700).

Rui, H., and Whinston, A. 2011. "Designing a social-broadcasting-based business intelligence system," *ACM Transactions on Management Information Systems (TMIS)* (2:4), ACM, p. 22 (doi: 10.1145/2070710.2070713).

Ruppert, D. 1988. "Efficient estimations from a slowly convergent Robbins-Monro process.,"

Sacerdote, B. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics* (116:2), Oxford University Press, pp. 681–704 (doi: 10.1162/00335530151144131).

Sanghvi, R. 2009. "New Tools to Control Your Experience," *The Facebook Blog* (9).

Sankar, L., Rajagopalan, S. R., and Poor, H. V. 2013. "Utility-privacy tradeoffs in databases: An information-theoretic approach," *Information Forensics and Security, IEEE Transactions on* (8:6), IEEE, pp. 838–852.

Schlenker, B. 1986. "Self-identification: Toward an integration of the private and public self," *Public self and private self* (available at http://link.springer.com/chapter/10.1007/978-1-4613-9564-5_2).

Schlenker, B. R. 1980. *Impression Management: The Self-concept, Social Identity, and Interpersonal Relations*, Brooks/Cole Publishing Company (available at http://books.google.com.sg/books/about/Impression_Management.html?id=0wt_QgAACAAJ&pgis=1).

Shalizi, C. R., and Thomas, A. C. 2011. "Homophily and contagion are generically confounded in observational social network studies," *Sociological methods & research* (40:2), Sage Publications, pp. 211–239.

Shiller, R. 1995. "Conversation, information, and herd behavior," *The American Economic Review* (85:2) (available at http://www.citeulike.org/user/widgetenator/article/11770599).

Sinclaire, J. K., and Vogus, C. E. 2011. “Adoption of social networking sites: an exploratory adaptive structuration perspective for global organizations,” *Information Technology and Management* (12:4), pp. 293–314 (doi: 10.1007/s10799-011-0086-5).

Singh, P. V., and Phelps, C. 2013. “Networks , Social Influence , and the Choice Among Competing Innovations : Insights from Open Source Software Licenses Networks , Social Influence , and the Choice Among Competing Innovations : Insights from Open Source Software Licenses,” *Management Science* (24:3), pp. 539–560.

Singla, P., and Richardson, M. 2008. “Yes, there is a correlation:-from social networks to personal behavior on the web,” in *Proceedings of the 17th international conference on World Wide Web*, pp. 655–664.

Siponen, M. T. 2005. “Analysis of modern IS security development approaches: towards the next generation of social and adaptable ISS methods,” *Information and organization* (15:4), Elsevier, pp. 339–375.

Slama, M., and Celuch, K. 1995. “Self-presentation and consumer interaction styles,” *Journal of Business and Psychology* (10:1), pp. 19–30 (doi: 10.1007/BF02249267).

Slama, M., Clark, T., and Wolfe, R. 1999. “Consumption as Self-Presentation: A Socioanalytic Interpretation of Mrs. Cage,” *Journal of marketing: A quarterly publication of the american marketing association*, American Marketing Association, pp. 135–138 (available at http://dialnet.unirioja.es/servlet/articulo?codigo=469070).

Smith, H. J., Dinev, T., and Xu, H. 2011. “Information privacy research: an interdisciplinary review,” *MIS Quarterly* (35:4), Society for Information Management and The Management Information Systems Research Center, pp. 989–1016 (available at http://dl.acm.org/citation.cfm?id=2208940.2208950).

Snijders, T. A. B. 1996. “Stochastic actor-oriented models for network change,” *Journal of mathematical sociology* (21:1-2), Taylor & Francis, pp. 149–172.

Snijders, T. A. B. 2001. “The Statistical Evaluation of Social Network Dynamics,” *Sociological Methodology* (31:1), pp. 361–395 (doi: 10.1111/0081-1750.00099).

Snijders, T., Steglich, C., and Schweinberger, M. 2007. “Modeling the coevolution of networks and behavior,” *Longitudinal models in the behavioral and related sciences*, na, pp. 41–71.

“Social Networking Fact Sheet | Pew Research Center’s Internet & American Life Project.,” 2014. *Pew Research* (available at http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/; retrieved May 4, 2015).

Solove, D. J. 2006. “A taxonomy of privacy,” *University of Pennsylvania Law Review* (154), JSTOR, pp. 477–564.

Statista. 2015. “Social network advertising revenue from 2014 to 2017,” p. Social network advertising revenue from 2014 to 20 (available at http://www.statista.com/statistics/271406/advertising-revenue-of-social-networks-worldwide/).

Steele, C. M. 1988. “The psychology of self-affirmation: Sustaining the integrity of the self,” *Advances in experimental social psychology* (21), Elsevier, pp. 261–302.

Steglich, C., Snijders, T. A. B., and Pearson, M. 2010. “Dynamic networks and behavior: Separating selection from influence,” *Sociological methodology* (40:1), Wiley Online Library, pp. 329–393.

Steinfield, C., DiMicco, J. M., Ellison, N. B., and Lampe, C. 2009. “Bowling online,” in *Proceedings of the fourth international conference on Communities and technologies - C&T ’09*, New York, New York, USA: ACM Press, June 25, p. 245 (doi: 10.1145/1556460.1556496).

Steinfield, C., Ellison, N. B., and Lampe, C. 2008. “Social capital, self-esteem, and use of online social network sites: A longitudinal analysis,” *Journal of Applied Developmental Psychology* (29:6), pp. 434–445 (doi: 10.1016/j.appdev.2008.07.002).

Stokman, F. N., and Doreian, P. 1997. “Evolution of social networks: processes and principles,” *Evolution of social networks*, Gordon and Breach, pp. 233–250.

Stone, E. F., and Stone, D. L. 1990. “Privacy in organizations: Theoretical issues, research findings, and protection mechanisms,” *Research in personnel and human resources management* (8:3), Sage Publications, pp. 349–411.

Stutzman, F. 2006. “An evaluation of identity-sharing behavior in social network communities,” *Journal of the International Digital Media and Arts Association* (3:1), Citeseer, pp. 10–18.

Stutzman, F., and Kramer-Duffield, J. 2010. “Friends only: examining a privacy-enhancing behavior in facebook,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1553–1562.

Syncapse. 2010. “The Value of a Facebook Fan : An Empirical Review.,”

Taylor, S. E., and Brown, J. D. 1988. “Illusion and well-being: a social psychological perspective on mental health.,” *Psychological bulletin* (103:2), American Psychological Association, p. 193.

Terlep, S., Vranica, S., and Raice, S. 2012. “GM Says Facebook Ads Don’t Pay Off - WSJ,” (available at http://www.wsj.com/articles/SB10001424052702304192704577406394017764460; retrieved April 28, 2015).

Terry, D. J., and O’Leary, J. E. 1995. “The theory of planned behaviour: The effects of perceived behavioural control and self-efficacy,” *British Journal of Social Psychology* (34:2), pp. 199–220 (doi: 10.1111/j.2044-8309.1995.tb01058.x).

Tesser, A. 1988. “Toward a self-evaluation maintenance model of social behavior,” *Advances in experimental social psychology* (21), Elsevier, pp. 181–227.

Toubia, O., and Stephen, A. T. 2013. “Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter?,” *Marketing Science* (32:3), INFORMS, pp. 368–392 (doi: 10.1287/mksc.2013.0773).

Tsai, W., and Ghoshal, S. 1998. “Social capital and value creation: The role of intrafirm networks,” *Academy of management Journal* (41:4), Academy of Management, pp. 464–476.

Tucker, C. 2012. “Social Advertising,” *SSRN Electronic Journal* (doi: 10.2139/ssrn.1975897).

Tucker, C. 2014. “Social Networks, Personalized Advertising and Privacy Controls,” *Journal of Marketing Research* (51:5), American Marketing Association, pp. 546–562 (doi: 10.1509/jmr.10.0355).

Tufekci, Z. 2007. “Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites,” *Bulletin of Science, Technology & Society* (28:1), pp. 20–36 (doi: 10.1177/0270467607311484).

Valkenburg, P. M., Peter, J., and Schouten, A. P. 2006. “Friend networking sites and their relationship to adolescents’ well-being and social self-esteem.,” *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society* (9:5), Mary Ann Liebert, Inc.   2 Madison Avenue Larchmont, NY 10538 USA, pp. 584–90 (doi: 10.1089/cpb.2006.9.584).

De Vries, L., Gensler, S., and Leeflang, P. S. H. 2012. “Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing,” *Journal of Interactive Marketing* (26:2), pp. 83–91 (doi: 10.1016/j.intmar.2012.01.003).

Walczuch, R. M., and Steeghs, L. 2001. “Implications of the new EU Directive on data protection for multinational corporations,” *Information Technology & People* (14:2), MCB UP Ltd, pp. 142–162.

Wang, Y.-C., Burke, M., and Kraut, R. E. 2013. "Gender, topic, and audience response," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (Cmc), p. 31 (doi: 10.1145/2470654.2470659).

Warren, S. D., and Brandeis, L. D. 1890. "The right to privacy," *Harvard law review* (4:5), JSTOR, pp. 193–220.

Wasow, O., Baron, A., Gerra, M., Lauderdale, K., and Zhang, H. 2010. "Can Tweets Kill a Movie? An Empirical Evaluation of the Bruno Effect," in *CHI 2010 Workshop on Microblogging* (available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.193.7667).

Wasserman, S. 1977. *Stochastic models for directed graphs*, Harvard University.

Wasserman, S. 1980a. "A stochastic model for directed graphs with transition rates determined by reciprocity," *Sociological methodology* (11), pp. 392–412.

Wasserman, S. 1980b. "Analyzing social networks as stochastic processes," *Journal of the American statistical association* (75:370), Taylor & Francis, pp. 280–294.

Wasserman, S., and Iacobucci, D. 1988. "Sequential social network data," *Psychometrika* (53:2), Springer, pp. 261–282.

Waters, R. D., Burnett, E., Lamm, A., and Lucas, J. 2009. "Engaging stakeholders through social networking: How nonprofit organizations are using Facebook," *Public Relations Review* (35:2), pp. 102–106 (doi: 10.1016/j.pubrev.2009.01.006).

Watson, D., Clark, L. a, and Tellegen, a. 1988. "Development and validation of brief measures of positive and negative affect: the PANAS scales.," *Journal of personality and social psychology* (54:6), pp. 1063–70 (available at http://www.ncbi.nlm.nih.gov/pubmed/3397865).

Wellman, B., Haase, A. Q., Witte, J., and Hampton, K. 2001. "Does the Internet Increase, Decrease, or Supplement Social Capital?: Social Networks, Participation, and Community Commitment," *American Behavioral Scientist* (45:3), pp. 436–455 (doi: 10.1177/00027640121957286).

Wernerfelt, B. 1984. "A resource-based view of the firm," *Strategic Management Journal* (5:2), John Wiley & Sons, Ltd., pp. 171–180 (doi: 10.1002/smj.4250050207).

Westin, A. F. 1968. "Privacy and freedom," *Washington and Lee Law Review* (25:1), p. 166.

Wilcox, K., and Stephen, A. T. 2013. “Are Close Friends the Enemy? Online Social Networks, Self-Esteem, and Self-Control,” *Journal of Consumer Research* (40:1), pp. 90–103 (doi: 10.1086/668794).

Wilson, M. 2015. “Stop doing quizzes on Facebook if you place any value on your privacy,” *betanews.com* (available at http://betanews.com/2015/12/19/stop-doing-quizzes-on-facebook-if-you-place-any-value-on-your-privacy/).

Wilson, R. E., Gosling, S. D., and Graham, L. T. 2012. “A Review of Facebook Research in the Social Sciences,” *Perspectives on Psychological Science* (7:3), pp. 203–220 (doi: 10.1177/1745691612442904).

Wong, F. M. F., Sen, S., and Chiang, M. 2012. “Why watching movie tweets won’t tell the whole story?,” in *Proceedings of the 2012 ACM workshop on Workshop on online social networks - WOSN ’12*, New York, New York, USA: ACM Press, August 17, p. 61 (doi: 10.1145/2342549.2342564).

Xu, H. 2007. “The effects of self-construal and perceived control on privacy concerns,” *ICIS 2007 Proceedings*, p. 125.

Xu, H., Teo, H.-H., Tan, B. C. Y., and Agarwal, R. 2009. “The role of push-pull technology in privacy calculus: the case of location-based services,” *Journal of Management Information Systems* (26:3), ME Sharpe, pp. 135–174.

Yin, G. 1991. “On extensions of Polyak’s averaging approach to stochastic approximation,” *Stochastics: An International Journal of Probability and Stochastic Processes* (36:3-4), Taylor & Francis, pp. 245–264.

Yoganarasimhan, H. 2012. “Impact of social network structure on content propagation: A study using YouTube data,” *Quantitative Marketing and Economics* (10:1), pp. 111–150 (doi: 10.1007/s11129-011-9105-4).

Zhang, K., Evgeniou, T., Padmanabhan, V., and Richard, E. 2012. “Content Contributor Management and Network Effects in a UGC Environment,” *Marketing Science* (31:3), INFORMS, pp. 433–447 (doi: 10.1287/mksc.1110.0639).

Zhang, Z., Li, X., and Chen, Y. 2012. “Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews,” ACM Transactions on Management Information Systems (TMIS) (3:1), ACM, p. 5 (doi: 10.1145/2151163.2151168).