THE ECONOMIC VALUE OF ONLINE NETWORKS: EVIDENCES FROM SOCIAL MEDIA AND ELECTRONIC COMMERCE

LIN ZHIJIE

(B.Eng.), South China University of Technology

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF INFORMATION SYSTEMS

NATIONAL UNIVERSITY OF SINGAPORE

2013

DECLARATION

I hereby declare that the 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.

Lin Zhijie

LIN ZHIJIE

5 September 2013

ACKNOWLEDGEMENTS

The completion of this dissertation would not have been possible without the support from many people in various ways. I wish to express my sincere gratitude to them.

First of all, I am most grateful to my committee chair and advisor, Dr. Goh Khim Yong, for his support and guidance given to me during the entire doctoral study. His research experiences and insights have greatly helped me develop my background in the field of economics of information systems.

Second, I would like to thank my other committee members, Dr. Heng Cheng Suang and Dr. Huang Ke-Wei, for their invaluable insights and willingness to share their time to discuss with me regarding the improvement of this research.

Third, I would also like to thank all the other professors in the NUS Department of Information Systems who have offered me valuable comments during my presentations in many seminars, which have significantly contributed to this research.

Fourth, I am also grateful to my fellow students and friends who have shared their research insights during our discussions, and have helped me go though the hard times during the past four years.

Last but not least, I would like to express my gratitude to my family for their constant support and encouragement to keep me moving forward in my doctoral study. I dedicate this dissertation to them.

TABLE OF CONTENTS

ABSTRACT

The rapid development of information technology has facilitated the emergence of social media-enabled social networks and electronic commerce-enabled product networks, and also the popularity of online recommendations via human beings and systems. Despite firms' popular use of online networks and online recommendations for greater business performance, there has been little research work investigating their economic impact. In this research, we conduct two empirical studies to examine how user interactions or recommendations in social media-enabled social networks affect consumer purchase behavior (study 1), and how structures of electronic commerce-enabled product recommendation networks influence product demand (study 2).

In study 1, the research question is: *How is consumer purchase behavior influenced by user interactions in a social network enabled by social media brand communities, and whether and how do the communication modes matter?* We integrate qualitative user-marketer interaction content data from a brand community social network on Facebook and consumer transactions data to assemble a unique dataset at the individual consumer level. We then quantify the impact of community contents from consumers (user-generated content, i.e., UGC) and marketers (marketer-generated content, i.e., MGC) on consumers' apparel purchase expenditures. A content analysis method was used to construct measures to capture the informative and persuasive nature of UGC and MGC while distinguishing between directed and undirected communication modes in the network. In our empirical analysis, we exploit differences across consumers' fan page joining decision and across timing differences in fan page joining dates for our model estimation and identification strategies. Importantly, we also control for potential self-selection biases and relevant factors such as pricing, promotion, social network attributes, consumer demographics and unobserved heterogeneity. Our findings show that engagement in social media brand community networks leads to a positive increase in purchase expenditures. Additional examinations of UGC and MGC impacts show evidence of network interaction contents affecting consumer purchase behavior through embedded

information and persuasion. We also uncover the different roles played by UGC and MGC, which vary by the type of directed or undirected communication modes by consumers and the marketer. Specifically, the elasticities of demand with respect to UGC information richness are 0.006 (directed communication) and 3.140 (undirected communication), whereas those for MGC information richness are insignificant. Moreover, the UGC valence elasticity of demand is 0.180 (undirected communication), while that for MGC valence is 0.004 (directed communication). Overall, UGC exhibits a stronger impact than MGC on consumer purchase behavior.

In study 2, the research questions are: *(1) Is the demand of a product influenced by both the incoming network and outgoing network? (2) How is the demand of a product influenced by product network attributes in terms of network diversity and network stability? (3) How do the diversity and stability effects differ between two types of recommendation networks (co-view and co-purchase)?* Using data from a Nikon store on Tmall.com, we use linear panel data models to examine the impact of network diversity and network stability on product demand. Importantly, we control for relevant factors at the individual product, product network, product category, and time unit levels, and account for implicit demand correlation (i.e., substitution and complementarity) and the simultaneity of demand and network structures. Our robustness checks also validate the consistency of our findings in the presence of potential collinearity, heteroskedasticity, price endogeneity, serial correlation, and across differences in variable operationalizations, time frames, and product categories. Our research identifies several notable findings. First, a 1% increase in the category diversity of the incoming (outgoing) co-purchase network of a product increases (decreases) the product's demand by 0.014% (0.011%). Second, a 1% increase in the stability of the outgoing co-purchase network of a product decreases the product's demand by 0.012%. These results show that the demand of a product is influenced by both the incoming and outgoing networks. Moreover, co-purchase network exhibits a stronger role than co-view network in affecting product demand.

The notable findings from this research provide significant contributions to the literature on the economic value of online networks and online recommendations, and offer important guidance to firms' online

network-based and recommendation-based business strategies, which are further discussed.

Keywords: Social network, Social media, Product network, Electronic commerce, Econometrics analysis

LIST OF TABLES

LIST OF FIGURES

1. GENERAL INTRODUCTION

The rapid development of information technology has facilitated the emergence of various mechanisms for firms' marketing purposes. First, two forms of online networks have been widely used by firms to achieve business performance. *Social networks* enabled by social media are one major form of online networks (Dou et al. 2013; Fang et al. 2013; Zeng and Wei 2013). Social media have become incredibly popular in recent years. For instance on Facebook, the number of monthly active users has already reached 1.15 billion by June 2013, an increase of 21% over the prior year (Facebook 2013). Due to the existence of the large potential customer base on social media, many firms or marketers have provided various online venues to engage consumers to achieve better business performance. For instance, more than 50 million brand communities (i.e., fan pages) have been set up on Facebook for marketing purposes by December 2012, an increase of 19% compared to that in April (Darwell 2013). Thus, this popular online venue has allowed a large number of consumers to engage in the community to form a social network for interactions, and produced extensive online user-generated content (UGC) or word of mouth (WOM). Hence, marketers, representatives of their firms, have also been attracted to generate content (hereafter termed as marketer-generated content (MGC)) to actively engage consumers with an aim to ultimately drive sales. Perhaps one of the most popular social networks on Facebook is the network of users in the Coca-Cola brand community.

In addition to the emergence of social networks on social media platforms, *product networks*, fostered by recommendation systems in electronic commerce (e-commerce), have become another important form of online networks (Oestreicher-Singer and Sundararajan 2012a; Oestreicher-Singer and Sundararajan 2012b). On most e-commerce sites, each product is featured on its own designated web page. On each product page, retailers can utilize some recommendation systems to explicitly recommend additional relevant products, which might be of consumers' interests, either to help consumers find the most suitable products or to cross-sell, thus creating a visible directed product network where products (i.e., network nodes) are explicitly connected by hyperlinks (i.e., network ties). Perhaps the best-known examples of product networks are the co-view and co-purchase recommendation networks on Amazon.com, where recommended products are listed under the titles "Customers who viewed this item also viewed” and "Customers who bought this item also bought”, respectively. Currently, similar product networks have existed on various e-commerce websites (e.g., Alibaba.com, Walmart.com, and Tmall.com).

Despite the fact that the surge in the emergence of online networks has attracted extensive attention from both academia and industry, there has been insufficient empirical research investigating the economic impact (e.g., sales impact) of such online networks of users and networks of products. One of the major reasons is the difficulty associated with quantifying or measuring the impact of online networks on key performance indicators such as sales. In other words, there still exist some critical gaps regarding the rigorous quantification of the value of online networks in the literature on the economic value of social media-enabled social networks and e-commerce-enabled product networks.

In addition to online networks, online recommendation has also become an important mechanism for firms’ marketing purpose. There are two major forms of online recommendations. The first one is the recommendation via human beings (online users). Online users can generate product-related recommendations[1] such as product reviews on product review sites (Liu 2006) and e-commerce sites (Chevalier and Mayzlin 2006), postings in blogs (Dhar and Chang 2009), forums (Tumarkin and Whitelaw 2001), and brand communities (Adjei et al. 2010). These recommendations usually convey recommending users’ attitudes and suggestions to consumers, and thus play an important role in consumers’ purchase decision process. For instance, an industry report from eMarketer shows that nearly three out of four consumers will search for online recommendations in the form of product reviews before their purchase decisions (eMarketer 2013). Academic research has also shown evidence of the impact of online users’ recommendations on consumer decision making (Chintagunta et al. 2010; Das and Chen 2007; Zhu and Zhang 2010). Accordingly, many firms are taking advantage of these online recommendations as a new marketing tool. For instance, firms or marketers

[1] These recommendations are also defined as WOM or UGC.

can sponsor promotional chats in online forums, such as USENET (Mayzlin 2006), and proactively induce their consumers to spread the word about their products online (Godes and Mayzlin 2004). Moreover, some marketers would even employ a community manipulation strategy (Dellarocas 2006) by anonymously behaving as "fellow consumers" to share positive product information with other consumers. Additionally, marketers now can also regularly post their online recommendations [2] to directly drive consumer purchases. Therefore, these various types of recommendations generated by human beings have become a common marketing strategy.

In contrast to the recommendation via human beings, the other form of recommendation is the recommendation via systems. Specifically, for each product online, firms will adopt some recommendation systems to recommend other related products (e.g., substitutes or complements) on the same web page to consumers based on some recommendation algorithms (e.g., collaborative filtering approach). These products include World Wide Web sites (Katona and Sarvary 2008), blogs (Mayzlin and Yoganarasimhan 2012), news reports (Dellarocas et al. 2010), and videos (Goldenberg et al. 2012). More typically, in the e-commerce context, retailers will adopt recommendation systems (e.g., co-view and co-purchase recommendation systems) to recommend other related products to consumers when they are viewing a certain product. It has commonly been assumed that these recommendations generated by systems would provide significant added value to consumers (Pathak et al. 2010). For instance, recommendations reduce consumers' product search cost especially when numerous products exist on an e-commerce site and consumers have difficulty in finding a suitable product out of so many (Häubl and Trifts 2000). Thus, the widely recognized added values provided by system-generated recommendations have also attracted retailers' implementation of recommendation systems on many current e-commerce sites.

Although online recommendations have been widely adopted by firms or marketers as a marketing strategy, and attracted significant interest from academia, there has been little empirical research investigating the economic impact of such online recommendations via users and systems. Thus, the

[2] These recommendations are also defined as MGC in this research.

literature also needs a rigorous quantification of the economic value of user recommendations and system recommendations.

Therefore, in order to address the above research gaps, the objective of this research is to conduct two empirical studies. The first study investigates how the interactions among consumers and between consumers and marketers within a social media-enabled user network affect consumers' purchase behavior. In this study, we also explore the effectiveness of recommendations generated by human beings (i.e., consumers and marketers) in affecting consumers' purchase behavior. The second study examines how the structures of product networks enabled by e-commerce platforms influence product demand. Meanwhile, we also explore the effectiveness of recommendations generated by systems in influencing product demands. Through these investigations, this research aims to complement and enrich the literature on the economic value of online networks and online recommendations. More importantly, this research seeks to demonstrate how to leverage online networks and online recommendations for greater business performance, and quantify the effectiveness (e.g., ROI) of these marketing strategies.

2. STUDY 1: SOCIAL NETWORKS IN SOCIAL MEDIA

2.1 Introduction

Despite the prevalent use of social media-enabled social networks by consumers and marketers, empirical research investigating their economic values still lags in three critical aspects that motivate our study.

First, prior UGC studies that have documented the economic impact of various aspects of UGC, such as review volume (Chevalier and Mayzlin 2006; Duan et al. 2008; Liu 2006), review subjectivity and readability (Ghose and Ipeirotis 2011), have focused mainly on one-time purchase items or products such as movies (Chevalier and Mayzlin 2006; Duan et al. 2008; Liu 2006) and books (Chevalier and Mayzlin 2006; Clemons et al. 2006). Studies such as Luca (2011) that examine UGC in relation to repeat purchase items are rare, and none have examined both UGC and MGC in the context of a social media brand community-enabled social network. Thus, the literature lacks a rigorous quantification of the value of recurring engagement by consumers and marketers in such a community, especially with metrics such as UGC and MGC elasticities of demand for repeat purchase goods.

Second, prior research has shed little light on the contention between the two complicated roles of consumers and marketers. Even though some research (Chen and Xie 2008; Mayzlin 2006) has attempted to evaluate the role of UGC side by side that of MGC or other marketer actions, empirical evidence on the relative efficacy of UGC and MGC in inducing consumer purchases is rare, with the exceptions of Trusov et al. (2009) and Albuquerque et al. (2012). Due to the simultaneous engagement of consumers and marketers in the network, consumers' purchase decisions are often influenced by both UGC and MGC. The potential conflict stems from different consumer motivations, needs, and at times, their level of skepticism toward MGC (Escalas 2007; Obermiller and Spangenberg 1998). Coupled with the potential two-sidedness (i.e., general positivity and negativity) of interactions from UGC and online WOM (Godes and Mayzlin 2009), it is thus not clear yet in the literature as to what the relative marketing effectiveness of MGC (which typically is overtly positive) and UGC on consumer purchases is.

Third, prior UGC research mostly focused on the aggregate-level economic values of UGC, but overlooked the critical phenomena occurring at the dyadic individual consumer level. Despite the increasing reliance of firms on consumers' WOM as a marketing strategy (Godes and Mayzlin 2009; Nam et al. 2010), little effort has been devoted to understanding whether and how modes of interpersonal communication matter. Consumer-to-consumer communication tends to be *undirected* in the past (e.g., in online reviews), and so does marketer-to-consumer communication propagated in a broadcast manner. Such undirected communications typically address the entire audience base at large without targeting a specific party and without regard for past interactions contexts. However, in many social media-enabled social network contexts (e.g., Facebook fan pages), juxtaposed among the undirected communication are often *directed* consumer-to-consumer and marketer-to-consumer communication (Burke et al. 2011). For example, consumers and marketers can pinpoint each other's remarks and respond in a targeted way to each party's content. They can interact on fan pages on a one-to-one basis via posting or commenting in response to a post. Despite its prevalence, research distinguishing the effects of directed and undirected communication modes of consumers and marketers in affecting consumer behavior still lags.

The objective of our study is to assess the impacts of both UGC and MGC in a brand community social network on consumers' repeat purchase behavior. By measuring the informative and persuasive aspects of UGC and MGC, and observing them at the dyadic individual consumer level, we seek to quantify their direct and relative impacts under directed and undirected communication modes. Our research question is thus: *How is consumer purchase behavior influenced by user interactions in a social network enabled by social media brand communities, and whether and how do the communication modes matter?*

To answer our research question, we collected UGC and MGC data from an apparel retailer's brand community social network (i.e., fan page) on Facebook, and matched these with network members' purchase information from the retailer's customer reward program database. We used a commercial text mining tool to construct measures to capture the informative and persuasive nature of UGC and MGC while distinguishing between directed and undirected communication modes in the network. Our econometric specification models consumers' weekly purchase expenditure as a function of UGC and MGC factors, controlling for relevant factors at the pricing, promotion, individual consumer, social network and time unit levels. Our identification strategy for the impacts of UGC and MGC is first based on the Propensity Score Matching technique which enables us to control for self-selection at the fan page level (Moe and Schweidel 2012) via constructing a "control" group of matched customers who were in the reward program but did not join the network. With the matched customer data sample, we then used a difference-in-differences approach to estimate the economic impact (i.e., "treatment" effect) of joining the network. We finally estimated a Heckman selection model to quantify the differential effects of directed and undirected UGC and MGC, while controlling for potential self-selection based on unobserved factors, as well as observed ones such as content generation and network ties. Lastly, we performed robustness checks to validate the consistency of our findings in the presence of potential serial correlation, and across differences in time lags and model specifications.

We find evidence that network interaction contents affect consumer purchase behavior through the embedded information and persuasion. Importantly, we determine the positive impact of joining the brand community network to be about $25 per consumer. We uncover the different roles played by UGC and MGC in driving consumer purchases, varying by the type of directed or undirected communication modes by consumers and the marketer. Specifically, consumers influence the purchases of one another through both informative and persuasive communications, while marketers influence it only through persuasive communication. Further, undirected contents are more effective than directed ones for both informative and persuasive consumer-to-consumer communication, while directed contents are more effective than undirected ones for persuasive marketer-to-consumer communication. The elasticities of demand with respect to UGC's informative effect (directed), informative effect (undirected), and persuasive effect (undirected) are estimated to be 0.006, 3.140 and 0.180 respectively, while that for MGC's persuasive effect (directed) is 0.004. UGC thus exhibits a more influential role than MGC in driving consumer purchases.

Overall, our study makes the following contributions. First, our study unveils the intricate roles of consumers and marketers in brand community networks, and provides a rigorous quantification of the economic impact of a brand community network's UGC and MGC on consumers' repeat purchases of an apparel brand. Second, our research serves as the first attempt to measure the direct and relative effectiveness and economic values of consumers' online WOM and marketers' proactive marketing activities in social media-enabled social networks at the individual consumer level. Third, our findings document the criticality of communication modes of social network content by showing the differential and even contrasting impacts of social network content under directed and undirected communication modes.

2.2 Literature Review

The popular advent of social network enabled by social media has witnessed a dramatic increase in online engagement and digitalized WOM communication (Dellarocas 2003). Marketers have also capitalized on the trend and launched brand communities on social media platforms to engage consumers, facilitate and generate WOM “buzz”, so as to increase information sharing and ultimately, drive sales (Kozinets 2002). This has also triggered researchers to investigate the economic value of social networks on social media platforms. Early efforts focused on the various outcomes of consumers’ engagement in social media brand communities. For instance, researchers studied consumers’ identification (Algesheimer et al. 2005), participation (Bagozzi and Dholakia 2006) and communication (Adjei et al. 2010) in a brand community. They found that these engagements would positively affect consumers’ community participation behavior and commitment, firm trust, and brand purchase behavior.

Other research efforts focused on the online WOM “buzz” *per se*, which is the observed output of consumers’ engagement on social media. This WOM “buzz” is typically defined as UGC. Most extant studies focused on the quantitative aspects (e.g., review volume and rating) of UGC and investigated their impact on some aggregate-level[3] economic outcomes. For instance, researchers studied the impact of user-generated reviews on sales of mostly one-time purchase goods, such as movies (Chintagunta et al. 2010; Duan et al. 2008; Liu 2006), books (Chevalier and Mayzlin 2006), video games (Zhu and Zhang 2010), and more rarely, repeat purchase goods such as beers (Clemons et al. 2006) and beauty products (Moe and Trusov 2011). They generally concluded that the quantitative aspects of online reviews such as review volume and/or rating (valence) positively affect aggregate product sales. Apart from online reviews, some studies also examined other types of UGC. Godes and Mayzlin (2004) studied Usenet newsgroup conversations, Tumarkin and Whitelaw (2001) investigated Internet postings in financial discussion forums, Dhar and Chang (2009) studied blog postings, and Albuquerque et al. (2012) studied user-created magazines in an online platform. Likewise, they also reported that quantitative aspects of UGC (e.g., volume, dispersion) were related to aggregate-level economic outcomes.

However, isolated findings on the quantitative aspects of UGC have gradually waned in conclusiveness as the role of qualitative information (e.g.,

[3] Aggregate level outcomes refer here to metrics such as total sales volume per day and brand market shares, as opposed to individual customer’s behavioral outcomes such as purchase expenditure or quantity in a trip or week.

textual content) escalates to the forefront with its importance in the current social media context. For instance, Forman et al. (2008) found that the disclosure of reviewer identity information and a shared geographical location between reviewers and consumers increased product sales, highlighting the impact of qualitative factors. To examine the qualitative aspects of UGC and their economic impact, researchers often use some qualitative analysis methods (e.g., text mining) or tools to extract embedded information from the textual contents. For instance, Pavlou and Dimoka (2006) extracted “benevolence” and “credibility” information embedded in the feedback text comments of sellers on eBay’s online auction marketplace. They found that superior past seller performance revealed by the sellers’ feedback text comments created price premiums for reputable sellers by engendering buyers’ trust in the sellers. Gu et al. (2007) extracted the “quality” of postings in virtual communities and found a trade-off between the quality and quantity of postings. Ghose and Ipeirotis (2011) constructed measures for two text-based attributes (subjectivity and readability) of review contents and concluded that these two factors positively affected sales. Additionally, in the finance discipline, Antweiler and Frank (2004) found that the bullishness (sentiment) of messages posted in Internet stock forums helped predict market volatility. Similarly, Das and Chen (2007) identified investor sentiments from stock market message boards and found a relationship between sentiments and stock values. Ghose et al. (2012) leveraged on UGC captured using data-mining techniques from social media platforms to generate a new ranking system for travel search engines. Sonnier et al. (2011) and Tirunillai and Tellis (2012) further classified online communications into positive, negative and indifferent sentiment categories, and found asymmetric impacts on firm sales and stock trading outcomes. In essence, this stream of studies reported that qualitative aspects of UGC exert an impact on aggregate-level economic outcomes.

Despite these research efforts in studying UGC impact, the invariable focus on aggregate-level economic values has resulted in researchers overlooking UGC interpersonal communication at the dyadic individual consumer level. Specifically, UGC captured in past studies tends to be communication in an undirected manner from consumers to consumers. For instance, online reviews (e.g., Chevalier and Mayzlin 2006; Clemons et al. 2006; Duan et al. 2008; Liu 2006) were posted by consumers who have purchased some products, while other consumers who have not purchased or are interested in the products can read these reviews. However, no directed messages were exchanged since reviewers were essentially writing the reviews with the general public in mind. This also applies to many other types of UGC in past studies, such as postings in financial forums (Tumarkin and Whitelaw 2001) and e-commerce websites (Pavlou and Dimoka 2006). However, social media platforms have now enabled many features for observable, directed interpersonal communication.

There exist only a few studies that examined the relative effect of UGC versus that of MGC, and thus are related to our study. For instance, Mayzlin (2006) developed an analytical model to examine the credibility of online WOM, which can be a mixture of consumer recommendations and disguised firm promotions. She found that consumer WOM can still be persuasive despite the overt promotional intent by firms in such online settings. Chen and Xie (2008) developed analytical models to argue that a major function of consumer reviews is to serve as a new element in the marketing communications mix. While they theorized that a firm's decision to provide consumer reviews can increase its incentive to offer more complete product information, there is no relative comparison on the profit impact of consumer reviews and traditional marketing communications. Trusov et al. (2009) studied the effects of WOM marketing on customer acquisition and growth at an Internet social networking site and compared it with traditional marketing mechanisms. This study only focused on aggregate outcomes such as the number of one-time customer acquisitions and not recurring sales by individual customers. The authors obtained a long-term elasticity for online WOM of 0.53, which is about 20 to 30 times higher than that for traditional marketing. Albuquerque et al. (2012) used data from an online user-generated magazine platform to compare content creator activities (e.g., referrals and WOM efforts) with firm-based actions (e.g., public relations). However, they lacked individual customer-specific visitation and communication data, and did not focus on MGC *per se* nor study qualitative aspects of UGC. Our research differs from the above studies by quantifying the extent to which different aspects of network interaction content drive sales of a repeat purchase product, in terms of textual aspects (*information richness* and *valence*), and communication modes (*directed* and *undirected*) of types of contents (UGC and MGC) at the *dyadic individual consumer level*.

2.3 Hypotheses

Consumers typically face product uncertainties prior to purchases, so they often seek information from online contents (e.g., consumer reviews) (Chevalier and Mayzlin 2006). Contents from mass media or social media are evaluative and can serve to persuade consumers (Goh et al. 2011). Thus, we aim to examine two effects (informative effect and persuasive effect[4]) of UGC and MGC in social media brand community network contexts. We focus on two important textual aspects of UGC and MGC, namely content information richness (to capture the informative effect) and content valence (to capture the persuasive effect). Content information richness refers to the amount of information (e.g., product or brand attributes, usage experiences) embedded in the UGC and MGC. Content valence refers to the embedded positive or negative sentiment, evaluation or attitude toward the product or brand, which can be shown through the use of positive or negative words (e.g., good, bad, terrible).

2.3.1 Content Information Richness

Consumers often face incomplete product information (Kivetz and Simonson 2000), so they need to make purchase decisions under uncertainties (Narayanan et al. 2007; Nelson 1970). As consumers are typically averse to losses (Kahneman and Tversky 1979), they may seek more product-related information to reduce their uncertainties. When uncertainties are reduced, consumers bear more confidence in making purchase decisions (Schubert and

[4] The informative effect of UGC/MGC draws analogy to the notion of informative advertising in the marketing literature, whereby consumers are provided with factual data on the nature and function of the product or service. Correspondingly, the persuasive effect of UGC/MGC parallels the persuasive advertising concept which assumes that consumers already understand the basic function or nature of the product, but have to be convinced of the desirability and/or benefits of the product that sets it apart from rival alternatives in a market.

Ginsburg 2000). Hence, *ceteris paribus*, when consumers possess more product-related information, they will be more likely to purchase a product that fits their needs or requirements.

A brand community is specialized, because at its center is a branded product (Muniz and O'Guinn 2001). UGC and MGC generated within the community involve product-related information. For instance, UGC may embed consumers' product usage experiences, which involve information of the product (e.g., product features) and other related information (e.g., shopping experiences). MGC may also embed product and other related information (e.g., warranty conditions, after-sales services). As such, we expect information richness of both UGC and MGC to have a positive impact on consumer purchase behavior.

The comparative impact of UGC and MGC (in terms of the informative effect) is ambivalent. On the one hand, the information asymmetry problem (i.e., firms have complete product information whereas consumers possess incomplete product information) (Akerlof 1970; Mishra et al. 1998) always plagues a consumer-firm relationship. Hence, consumers are tempted to seek information they need from marketers (or representatives of firms), rather than from other consumers who may lack the desired information. As such, MGC information might be more effective than UGC information in addressing consumers' needs and reducing uncertainties. Moreover, search and processing costs are incurred when consumers seek and process information (Ratchford 1982). Since MGC has a higher likelihood to embed information that fits consumers' needs, it will be less costly for consumers' information seeking and processing. As a result, consumers might put more weight on MGC than UGC. Thus, we expect MGC information richness to be more influential than UGC information richness.

On the other hand, there is another school of competing thoughts. Specifically, information generated by marketers typically describes product information based on technical specifications and is thus product oriented, whereas consumer-generated information tends to describe a product based on usage conditions from a consumer's perspective and is, in contrast, more likely to be consumer-oriented (Bickart and Schindler 2001). In other words, UGC information might be more relevant to consumers than MGC information, and thus has the advantage of helping consumers find products matching their preferences (Chen and Xie 2008). This begets the competing hypothesis that UGC information richness will be more influential than MGC information richness in influencing consumer purchases. Summing both perspectives, we arrive at a set of competing hypotheses:

Hypothesis 1A (H1A, competing): UGC information richness has a smaller impact than MGC information richness on consumers' purchase behavior.

Hypothesis 1B (H1B, competing): UGC information richness has a larger impact than MGC information richness on consumers' purchase behavior.

2.3.2 Content Valence

Consumers often love to share and relate their product experiences with members in a network, expressing their opinions and sentiments (Algesheimer et al. 2005). If consumers are satisfied with a brand or product, they may exhibit favorable attitudes and sentiments toward it. If they dislike the brand or product, or are marred by the experience, they may exhibit negative attitudes and sentiments. Hence, valence embedded in UGC can be interpreted as their general evaluations of a brand or product (Clemons et al. 2006; Liu 2006). Positive (negative) valence of UGC should drive (impede) consumer purchases (Pavlou and Dimoka 2006).

The impact of MGC valence can be discerned from the literature on persuasive advertising (e.g., Russo and Chaxel 2010; Von der Fehr and Stevik 1998). Persuasive advertising involves messages that highlight the positivity of products to enhance evaluations and to instill a sense of good feeling in consumers to tempt them into purchase (Wu et al. 2009). Similarly, marketers embed their positive statements in MGC to create a favorable product reputation and image to influence sales. Hence, we posit that the impact of MGC valence, similar to that of persuasive advertising, positively influences consumers' purchase behavior.

However, MGC may exhibit a weaker persuasive effect than that of UGC. Specifically, over the years, consumers have developed a general tendency to disbelieve or be skeptical toward marketing messages (Escalas

2007). They feel that marketers would resort to gimmicks and tricks (e.g., exaggerating the product benefits while downplaying the weaknesses) in order to persuade consumers to purchase. In contrast, other consumers have little reasons for doing so. Moreover, consumers tend to trust UGC in evaluating products because they are more similar to one another in terms of community identities, needs and preferences for specific brands or products and their information (Arazy et al. 2010; Brown and Reingen 1987; Gilly et al. 1998). Thus, consumers might succumb more to UGC persuasion rather than MGC persuasion. Trusov et al. (2009) documented that the impact of user referrals (persuasion) on member growth at an Internet social networking site is higher than that of traditional marketing communications (e.g., media appearances and promotional events). This corroborates our conjecture that UGC might be stronger than MGC in terms of persuasive effect. In essence, we postulate that social media UGC valence has a larger impact than MGC valence in driving purchases.

Hypothesis 2 (H2): UGC valence has a larger impact than MGC valence on consumers' purchase behavior.

2.3.3 Directed Communication versus Undirected Communication

Consumers are inundated with irrelevant information in online environments nowadays (Tam and Ho 2005). Hence, a directed message, which is communicated to a targeted consumer, is expected to be more effective than an undirected one circulated to the mass population, because directed communication easily captures one's attention and elicits a response (Amaldoss and He 2009). Moreover, compared to undirected communication, consumer-to-consumer directed communication is more likely to evoke norms of reciprocity. Such directed communication in brand communities may be more intimate in the message contents such that WOM product recommendation or feedback can be exchanged in a more personalized manner fitting each other's preferences or needs (Burke et al. 2011). We thus postulate that communicating in a directed manner with UGC would be more effective in driving consumer purchases than doing so in an undirected manner for consumer-to-consumer interactions in social media brand communities.

Hypothesis 3 (H3): For brand community UGC, the impact of directed communication is more effective than that of undirected communication in influencing consumers' purchase behavior.

The comparative advantage of directed messaging over undirected messaging for MGC communication is equivocal. On the one hand, when marketers directly communicate to a specific consumer, it is easier to capture one's attention relative to undirected communication addressing the entire customer base without regard for past interaction contexts or specific targeted consumers. Directed marketing messages designed for and communicated to a specific consumer are often tailored to one's needs, heightening the relevance and fit. This ensures that replies can be customized to generate responses or interactions to culminate in eventual purchases (Manchanda et al. 2008). Indeed, directed communications are often exemplary of great customer service.

On the other hand, if marketers frequently engage in unsolicited directed communication with consumers, consumers' skepticism and annoyance (Obermiller and Spangenberg 1998) might be aggravated. This might result in the termination of such communication links (Goh et al. 2011), or disapproving behaviors, such as product boycotts or even the dissemination of negative WOM (Smith and Cooper-Martin 1997). Conversely, undirected marketing communications by a marketer may have a higher level of reach in message receipt by consumers in the brand community of platforms such as Facebook. Undirected communications often get propagated as "posts" or news streams that appear prominently, for instance, on a fan's or consumer's own Facebook "News Feed" page. In contrast, a marketer's directed messages to specific consumers have a lower level of reach or exposure. As such, undirected marketing communication might be more effective than directed communication. Thus, these two camps of arguments give rise to our competing set of hypotheses.

Hypothesis 4A (H4A, competing): For brand community MGC, the impact of directed communication is more effective than that of undirected communication in influencing consumers' purchase behavior.

Hypothesis 4B (H4B, competing): For brand community MGC, the impact of directed communication is less effective than that of undirected communication in influencing consumers' purchase behavior.

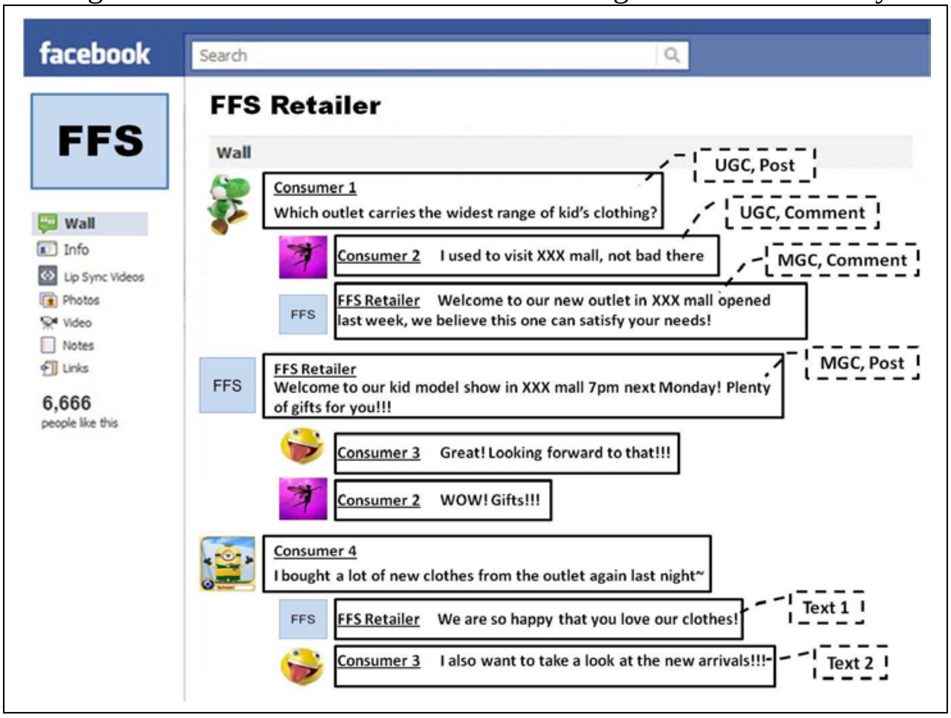
2.4 Empirical Method and Analysis

2.4.1 Research Context

Our research context is a brand community social network on Facebook set up in July 2009 by FFS[5], a casual wear apparel retailer in a small Asian market. The retailer also provided us with customer information from their reward program database. Figure 1-1 presents an edited screenshot of the brand community. FFS retailer set up this community to serve as a platform to engage and interact with their consumers, and also to facilitate network interactions among consumers. Consumers can "like" this fan page to engage as community members, and then interact with other consumers and the marketer (i.e., FFS retailer). Users interact by generating content, such as posts and comments. Contents generated by consumers (or the marketer) are referred to as UGC (or MGC). According to FFS retailer, Facebook is the only social media platform it uses to engage consumers. This thus provides us a thorough, unambiguous setting to examine the impact of UGC and MGC on consumer behavior.

[5] Due to a non-disclosure agreement, we are not able to reveal the identity of the retailer.

Figure 1-1 - FFS Retailer Facebook Fan Page Brand Community



Note: The most recent post appears on top, but the most recent comment appears at the bottom of a list of comments related to a particular post.

In this community, we observe two types of content, i.e., posts and comments, for both UGC and MGC. Posts are initial text postings which may be addressed to someone (directed) or the entire network (undirected) whereas comments are follow-ups to posts. Although comments are responses to posts, they too can be directed or undirected. Hence, the coders manually read through all posts and comments to ensure the correct coding of communication modes. Posts and comments which were directly addressed to a user are coded as directed communications whereas posts and comments which were not directly addressed to a user were deemed as undirected communications. For instance, Texts 1 and 2 to Consumer 4 are directed communications from the marketer and Consumer 3 respectively, whereas all other messages generated by others are considered as undirected communications to Consumer 4 (e.g., the phrase “WOW! Gifts!!!” from Consumer 2).

2.4.2 Qualitative Analysis

We employ text mining techniques provided by SPSS Clementine to analyze the textual or qualitative UGC and MGC data for quantitative analysis.

Given a piece of textual content, the text mining tool first decomposes the content into words and phrases based on its large library, and then performs extraction of concepts. Each extracted concept is assigned a corresponding type indicating the sentiment nature (positive, negative or indifferent).

Table 1-1 presents an illustration of the qualitative analysis results generated by SPSS Clementine. We input three pieces of text examples, one from the marketer (ID=1) and two from consumers (ID=2, ID=3). The text from the marketer is "*We have plenty of new arrivals for you!*". The other two texts from consumers are: "*That's great! I always love your skirts, make me look so good!*", "*I don't like your color, always too red!*". The text mining tool will analyze all the textual contents and display the results as indicated in Table 1-1. "ID" shows the index of each piece of text. "Matched Text" shows the original text. "Concept1" and "Concept2" show the extracted concepts (indicated by brackets <* *> in the original text) from a particular piece of text. "Type1" and "Type2" indicate the corresponding sentiment nature for each extracted concept. Positive (negative) sentiment can be identified by type value with "Positive" ("Negative"), otherwise is indifferent. For instance, five concepts are identified from the second text (ID=2), with three of them positive, none of them negative, and two of them indifferent.

Table 1-1 - Text Analysis Results

SN	Concept1	Type1	Concept2	Type2	ID	Matched Text
1	arrivals	Indifferent	new	Indifferent	1	we have plenty of <*new*> <*arrivals*> for you
2	excellent	Positive			2	that's <*great*>
3	skirts	Indifferent	like	Positive	2	i always <*love*> your <*skirts*> , make me look so good
4	look	Indifferent	good	Positive	2	i always love your skirts , make me <*look*> so <*good*>
5	color	Indifferent	dislike	Negative	3	i <*don't like*> your <*color*> , always too red

As the number of concepts can indicate the richness of information and the type of a concept can reflect the embedded sentiment, our measures of UGC and MGC factors are directly derived from these text mining results.

First, information richness is measured as the number of concepts extracted. Previous information extraction studies also extracted information by identifying context-related or context-free concepts (e.g., Rau et al. 1989). Similar approaches have been employed in studies in various disciplines. For instance, researchers had operationalized information richness as the amount of concepts (e.g., price, quality) communicated by advertisements (e.g., Healey and Kassarjian 1983; Resnik and Stern 1977).

Second, valence is measured as the net positivity (i.e., number of positive concepts minus number of negative concepts), which is derived from a sentiment classification algorithm, i.e., *Naïve Classifier* (Das and Chen 2007). Each word in a text is checked against the lexicon and given a value (-1, 0, +1) based on sentiment type (negative, indifferent, positive). The net word count of all lexicon-matched words is taken, and the text is deemed positive (negative) if the value is greater (less) than zero; else, it is indifferent.

2.4.3 Empirical Model

2.4.3.1 Communication Intensity

It has been widely acknowledged that online network interactions can allow online users to establish awareness of one another (McKenna et al. 2002), and the awareness may increase with the amount of interactions and eventually lead to online relationship development (Parks and Floyd 1996). Different levels of awareness may result in different levels of communication impact (Brown and Reingen 1987). For instance, one may expect the information from a friend, whom he or she has a higher awareness of, to be more influential compared to the same information from a stranger. In addition, consumers may have a relationship with firms or their representatives such as a marketer, and this relationship may also affect consumers' purchase decisions (Crosby and Stephens 1987). Importantly, trust in online merchants is also typically built up over time with increasing interactions and patronage (Pavlou and Dimoka 2006).

In order to account for this, we use *communication intensity* to weigh the impact of each *directed* consumer-to-consumer (UGC) and marketer-to-consumer (MGC) communication. Thus, the information richness and valence of each directed communication is weighted by the communication intensity between each pair of communicating users. To account for this intensity between each pair of users, we measure the number of prior directed communications between them, accumulated over time.

2.4.3.2 UGC Factors

For directed communication, $U\_D\_IR_{it}$ in Equation (1) and $U\_D\_VA_{it}$ in Equation (2) denote the average information richness and average valence of UGC that consumer i has observed through directed communications in time period t. $UDIR_{ijtm}$ and $UDVA_{ijtm}$ are the information richness and valence of the m^{th} UGC that consumer i has observed from consumer j through directed communication in period t. $UIntensity_{ijtm}$ is the communication intensity between consumers i and j, which is measured as the number of previous directed communications between consumers i and j prior to their m^{th} directed communication in period t. M_{ijt} denotes the total number of UGC that consumer j has generated to consumer i through directed messaging in period t. Thus, dividing the inner summation term of weighted $UDIR_{ijtm}$ and $UDVA_{ijtm}$ in Equations (1) and (2) by M_{ijt} obtains the average information richness and average valence of directed UGC from each consumer j. Finally, J_{it} is the total number of consumers who have generated directed messages to consumer i in period t. Therefore, dividing the outer summation term in Equations (1) and (2) by J_{it} derives the mean information richness and valence of directed UGC for consumer i across J_{it} users whom consumer i interacted with in a directed manner.

$$U\_D\_IR_{it} = \sum_{j=1}^{J_{it}} \left(\frac{\sum_{m=1}^{M_{ijt}} \left(UDIR_{ijtm} \times UIntensity_{ijtm} \right)}{M_{ijt}} \right) / J_{it} \tag{1}$$

$$U\_D\_VA_{it} = \sum_{j=1}^{J_{it}} \left(\frac{\sum_{m=1}^{M_{ijt}} \left(UDVA_{ijtm} \times UIntensity_{ijtm} \right)}{M_{ijt}} \right) / J_{it} \tag{2}$$

For undirected communication, $U\_U\_IR_{it}$ in Equation (3) and $U\_U\_VA_{it}$ in Equation (4) denote the average information richness and valence of UGC that consumer i has observed through undirected communication in period t. $U\_U\_IR_{it}$ and $U\_U\_VA_{it}$ are simply the average information richness and average valence of all N_{it} pieces of UGC that consumer i has observed through undirected communication in period t, where $UUIR_{itn}$ and $UUVA_{itn}$

denote the information richness and valence of the n^{th} UGC that consumer i has observed through undirected communication in period t.

$$U\_U\_IR_{it} = \sum_{n=1}^{N_{it}} UUIR_{itn} / N_{it} \tag{3}$$

$$U\_U\_VA_{it} = \sum_{n=1}^{N_{it}} UUVA_{itn} / N_{it} \tag{4}$$

2.4.3.3 MGC Factors

For directed communication, $M\_D\_IR_{it}$ in Equation (5) and $M\_D\_VA_{it}$ in Equation (6) denote the average information richness and average valence of directed MGC that the marketer has communicated to consumer i in period t. $MDIR_{itr}$ and $MDVA_{itr}$ are the information richness and valence of the r^{th} directed MGC that the marketer has communicated to consumer i in period t. $MIntensity_{itr}$ is the communication intensity between consumer i and the marketer, measured as the number of prior directed communications between consumer i and the marketer prior to their r^{th} directed communication in period t. R_{it} denotes the total number of directed MGC that the marketer has communicated to consumer i in period t.

$$M\_D\_IR_{it} = \sum_{r=1}^{R_{it}} \left(MDIR_{itr} \times MIntensity_{itr} \right) / R_{it} \tag{5}$$

$$M\_D\_VA_{it} = \sum_{r=1}^{R_{it}} \left(MDVA_{itr} \times MIntensity_{itr} \right) / R_{it} \tag{6}$$

For undirected communication, $M\_U\_IR_{it}$ in Equation (7) and $M\_U\_VA_{it}$ in Equation (8) denote the average information richness and average valence of MGC that consumer i has observed through undirected communication in period t. $M\_U\_IR_{it}$ and $M\_U\_VA_{it}$ are simply the average information richness and average valence of all S_{it} pieces of MGC that consumer i has observed through undirected communication in period t, where $MUIR_{its}$ and $MUVA_{its}$ denote the information richness and valence of the s^{th} MGC that consumer i has observed through undirected communication in period t.

$$M\_U\_IR_{it} = \sum_{s=1}^{S_{it}} MUIR_{its} / S_{it} \tag{7}$$

$$M\_U\_VA_{it} = \sum_{s=1}^{S_{it}} MUVA_{its} / S_{it} \quad (8)$$

2.4.3.4 Control Variables

To obtain robust estimates of the effect of focal UGC and MGC constructs, we control for potentially confounding factors at the pricing, promotion, individual consumer, peer social network and time levels.

Besides the focal UGC and MGC variables, we also control for other important aspects of UGC and MGC, namely the volumes of directed UGC ($U\_D\_VO_{it}$), undirected UGC ($U\_U\_VO_{it}$), directed MGC ($M\_D\_VO_{it}$) and undirected MGC ($M\_U\_VO_{it}$) that consumer *i* observed in the brand community network in period *t*. To account for potential selection bias at the content generation level, we include variables that measure a user's own posting valence (OWN_VA_{it}) and own posting volume (OWN_VO_{it}), i.e., the average valence and total volume of content generated by consumer *i* in the brand community network in period *t*.

Importantly, we also include control variables that measure the extent of peer effects, influence and general activity in the FFS brand community, as well as a user's Facebook social network at large. To quantify the influence of a fan, we compute his or her degree centrality[6] ($CENT_{it}$) on the FFS fan page, based on the communication ties consumer *i* maintained with other consumers on the fan page in period *t*. Other control measures that account for the extent of network ties, activity and influence from a consumer's Facebook social network at large include the count of Facebook page views[7] (FB_V_i, i.e., total number of Facebook page views since consumer *i*'s registration of an account on Facebook), the number of Facebook friends (FB_F_i), and the number of consumer *i*'s Facebook friends who were also fans on the FFS fan page (FFS_F_i).

---

[6] We mapped the network structure of users based on directed content communications on the FFS fan page, i.e., two users or consumers are deemed to be connected to each other if they have ever engaged in directed communications.

[7] This measure varies across individual consumers but is time-invariant, as is the case for FB_F_i and FFS_F_i.

To control for the effects of marketing-mix activities, we include a variable $PRICE_t$ that measures the average price (inclusive of discounts) of all products sold in period t. We account for promotional intensity[8] ($PROM_t$), i.e., the average level of promotion across all days in period t. Promotion on each day is measured as a dummy indicator of a promotional event based on information from the retailer's marketing calendar.

At the consumer level, we account for past expenditure ($PEXP_{it}$), i.e., consumer i's average expenditure per transaction prior to period t. Other demographic variables captured include a consumer's age[9] (AGE_i), monthly income (INC_i, i.e., the level of consumer i's monthly income (1: lowest, 5: highest)), and gender ($MALE_i$, i.e., a dummy indicator for male gender (1: male, 0: female)). Lastly, we include a set of weekly time dummies (θ_t).

2.4.3.5 Econometrics Model Specifications

In Equation (9), we model the influence of UGC and MGC factors on consumers' purchase expenditure. The dependent variable in this study is consumer i's total purchase expenditure in period t ($EXPEND_{it}$).

$$\begin{aligned}
EXPEND_{it} = {} & \beta_1 U_D_IR_{i,t-1} + \beta_2 U_U_IR_{i,t-1} + \beta_3 U_D_VA_{i,t-1} + \beta_4 U_U_VA_{i,t-1} \\
& + \beta_5 M_D_IR_{i,t-1} + \beta_6 M_U_IR_{i,t-1} + \beta_7 M_D_VA_{i,t-1} + \beta_8 M_U_VA_{i,t-1} \\
& + \beta_9 U_D_VO_{i,t-1} + \beta_{10} U_U_VO_{i,t-1} + \beta_{11} M_D_VO_{i,t-1} + \beta_{12} M_U_VO_{i,t-1} \\
& + \beta_{13} OWN_VA_{i,t-1} + \beta_{14} OWN_VO_{i,t-1} + \beta_{15} CENT_{i,t-1} \\
& + \beta_{16} FB_V_i + \beta_{17} FB_F_i + \beta_{18} FFS_F_i \\
& + \beta_{19} PRICE_t + \beta_{20} PROM_t + \beta_{21} PEXP_{it} \\
& + \beta_{22} AGE_i + \beta_{23} INC_i + \beta_{24} MALE_i + \theta_t + \alpha_i + \varepsilon_{it}
\end{aligned} \tag{9}$$

We consider UGC and MGC factors in the previous time period (t-1) to avoid simultaneity issues and to allow for a lagged effect from consumers' UGC and MGC exposure to their actual purchases[10]. βs are the model coefficients of interest, α_i captures unobserved consumer-specific effects, and ε_{it} is the residual error term.

---

[8] These marketing promotions were targeted at all consumers (i.e., both fans and non-fans of the community).

[9] We used consumers' age as of July 2010 (the mid-point of our dataset).

[10] We compared a set of lag time-period models and determined our choice of one-period lag (t-1) as the best lag level in terms of model fit statistics. The comparison is shown in Table 1-12.

To account for self-selection decisions of consumers joining the FFS brand community, we further specify and estimate a Heckman selection model, i.e., the combination of expenditure model in Equation (9) and selection model in Equations (10) to (12). To model the first-stage fan page selection decision ($BrandCom_i$), we include several exogenous variables as covariates in the first-stage Probit model shown in Equations (10) to (12): (1) AGE_i, (2) INC_i, (3) $MALE_i$, two binary indicators of whether a consumer disclosed his or her (4) home phone number ($PHONE\_DIS_i$) and (5) home address ($ADDRESS\_DIS_i$), and two indicators of whether a consumer opted in to receive promotional information through (6) mobile phone ($PHONE\_OPT_i$) and (7) postal mail ($MAIL\_OPT_i$) when one signed up as a reward program member.

Selection Equation:
$$\begin{aligned} BrandCom_i^* = {} & \delta_1 AGE_i + \delta_2 INC_i + \delta_3 MALE_i \\ & + \delta_4 PHONE\_DIS_i + \delta_5 ADDRESS\_DIS_i \\ & + \delta_6 PHONE\_OPT_i + \delta_7 MAIL\_OPT_i + \mu_i \end{aligned} \tag{10}$$

$$BrandCom_i = 1 \text{ if } BrandCom_i^* > 0, \text{ and } BrandCom_i = 0 \text{ otherwise} \tag{11}$$

$$\text{Prob}(BrandCom_i = 1 \mid z_i) = \Phi(z_i\delta), \;\; \text{Prob}(BrandCom_i = 0 \mid z_i) = 1 - \Phi(z_i\delta) \tag{12}$$

where z_i is a vector of Heckman first-stage model covariates as described in the prior paragraph.

We expect that a consumer's fan page selection decision, $BrandCom_i$, to be related to age, income level and gender (Muniz and O'Guinn 2001) since FFS is an apparel retailer with trendy, stylish men, women and baby/kids wear offerings. We also expect a user's decision to join the FFS fan page (and thus Facebook) to be related to concerns over data or information privacy (which can be proxied by phone number and address disclosures) and interests in receiving marketing communications from FFS over different channels (Tsai et al. 2011).

2.4.4 Data Description

The data in our study were drawn from three sources. First, we wrote Java codes based on the Facebook API to retrieve all user interaction contents from FFS retailer's fan page community on Facebook. Second, Facebook user details and usage logs were obtained from a source related to the Facebook

Data Science Team. Third, FFS retailer provided us with (1) the customer reward program database with information for 14,388 customers, (2) the purchase transactions data of customers in this database, and (3) the marketing calendar that detailed the marketing events in a period. These datasets allowed us to construct our major variables of interest and the various control variables. We finally matched Facebook interaction contents data with transactions data by consumer names, and organized our model estimation data at the consumer-week level.

Our data spans 104 weeks from when the brand community was first launched in July 2009 till June 2011. By June 2011, the FFS fan page acquired about 6,600 fans in total[11]. On average at the weekly basis, there were about 2.07 MGC posts (std. dev. = 2.08, max = 10) and about 2.59 MGC comments (std. dev. = 3.67, max = 25). Similarly, in terms of UGC participation, the mean UGC postings averaged about 1.62 per week (std. dev. = 2.72, max = 17) while the mean UGC comments averaged around 5.72 per week (std. dev. = 10.11, max = 62). On aggregate, UGC plus MGC participations averaged 12 incidences (std. dev. = 15.57, max = 78) on a weekly basis. In general, we note that there is a high level of heterogeneity or variation in the UGC and MGC contributions on a week to week basis, which provides a vital source of identification for the UGC and MGC effects that can influence purchase behaviors. In assembling the final sample at the consumer-week level, there is no left censoring since we know the date of each fan's joining of the fan page and the date of first purchase.

Our final data sample for model estimations has 398 unique consumers who are both members of the FFS reward program and fans of FFS Facebook fan page. Across all purchase transactions, these 398 customers spent on average \$37.05 (std. dev. = \$29.15). We further find that the average purchase expenditure *before* joining the fan page was \$28.57 (std. dev. = \$29.19), while that *after* joining the fan page was \$40.52 (std. dev. = \$28.41) - a positive difference of about \$12. Comparatively, the average purchase expenditure for all 14,388 customers in the reward program was \$32.93 across all transactions.

[11] To put the number of fans in perspective, we note that the FFS fan page is within the top 100 country-specific Facebook fan pages in terms of acquired fans, as listed on http://www.socialbakers.com/facebook-pages.

Table 1-2 shows the descriptive statistics of model variables for the unbalanced panel of 398 consumers across 20,406 observations. We note that there is a high level of variability in the UGC and MGC information richness and valence variables, with many cases of over-dispersion (i.e., mean > std. dev.). Comparing UGC with MGC, the means and standard deviations of MGC information richness and valence variables are higher than those of equivalent UGC variables[12]. A correlation matrix is shown in Table 1-3.

Table 1-2 - Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
EXPEND (Purchase expenditure)	4.711	22.546	0.000	538.420
U_D_IR (UGC, directed, information richness)	0.006	0.177	0.000	12.000
U_U_IR (UGC, undirected, information richness)	3.143	2.021	0.000	14.000
U_D_VA (UGC, directed, valence)	-0.00005	0.019	-1.000	1.000
U_U_VA (UGC, undirected, valence)	0.181	0.539	-3.000	2.000
M_D_IR (MGC, directed, information richness)	0.037	0.896	0.000	48.000
M_U_IR (MGC, undirected, information richness)	7.010	3.359	0.000	16.000
M_D_VA (MGC, directed, valence)	0.004	0.166	-4.000	9.000
M_U_VA (MGC, undirected, valence)	0.705	0.987	-2.000	4.000
U_D_VO (UGC, directed, volume)	0.026	0.815	0.000	45.000
U_U_VO (UGC, undirected, volume)	51.378	172.546	0.000	1184.000
M_D_VO (MGC, directed, volume)	0.004	0.104	0.000	7.000
M_U_VO (MGC, undirected, volume)	12.331	21.091	0.000	112.000
OWN_VA (Own posting valence)	0.0001	0.013	-0.500	1.000
OWN_VO (Own posting volume)	0.003	0.071	0.000	4.000
CENT (Degree centrality)	0.0001	0.010	0.000	1.000
FB_V (Number of Facebook page views)	120.087	148.361	0.000	1261.000
FB_F (Number of Facebook friends)	354.254	388.599	0.000	4791.000
FFS_F (Number of Facebook friends on FFS)	4.813	6.909	0.000	68.000
PRICE (Product price)	55.463	18.517	31.036	144.060
PROM (Promotion intensity)	0.753	0.351	0.000	1.000
PEXP (Past expenditure)	40.685	28.190	0.000	266.290
AGE (Age)	32.508	6.216	16.333	54.167
INC (Income level)	2.357	0.836	1.000	5.000
MALE (Gender)	0.110	0.312	0.000	1.000

Note: Observations = 20,406. Mean *EXPEND* across non-zero expenditure weeks = 56.685.

[12] Although it may appear counter-intuitive that there are high variability and negative values in MGC valence, it can be explained by instances where some consumers requested for home delivery services, but the marketer had to apologize for the unavailability of such services. Some consumers also complained about poor in-store services, and the marketer apologized while offering discount coupons as compensation. Such compensatory marketer actions may over-react at times in order to maintain customer satisfaction levels, thus explaining the higher means and variability of MGC factors.

Table 1-3 - Correlation Matrix

Variable		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	$EXPEND_{it}$	-														
2	$U\_D\_IR_{i,t\text{-}1}$	0.002	-													
3	$U\_U\_IR_{i,t\text{-}1}$	0.033	0.012	-												
4	$U\_D\_VA_{i,t\text{-}1}$	-0.002	-0.164	-0.007	-											
5	$U\_U\_VA_{i,t\text{-}1}$	0.001	0.001	0.116	0.003	-										
6	$M\_D\_IR_{i,t\text{-}1}$	0.009	0.571	-0.007	-0.153	-0.001	-									
7	$M\_U\_IR_{i,t\text{-}1}$	0.020	0.009	0.192	0.000	-0.008	0.021	-								
8	$M\_D\_VA_{i,t\text{-}1}$	0.017	0.461	-0.002	-0.287	-0.000	0.683	0.010	-							
9	$M\_U\_VA_{i,t\text{-}1}$	0.038	0.012	0.181	-0.012	-0.076	0.014	0.438	0.020	-						
10	$PRICE_t$	-0.018	0.003	-0.045	0.003	0.006	0.003	0.175	-0.005	0.069	-					
11	$PROM_t$	0.033	-0.006	-0.078	0.008	0.223	-0.031	0.065	-0.013	0.119	0.140	-				
12	$PEXP_{it}$	0.022	0.026	0.006	-0.002	-0.021	0.019	0.042	0.021	0.024	0.086	0.011	-			
13	AGE_i	0.004	0.002	0.008	-0.009	0.008	-0.003	0.004	-0.004	0.006	0.000	0.007	0.002	-		
14	INC_i	0.004	-0.002	0.004	0.001	0.005	-0.005	-0.008	-0.004	-0.001	-0.018	-0.001	0.039	0.334	-	
15	$MALE_i$	0.001	0.007	-0.009	-0.008	-0.004	0.010	-0.003	0.016	-0.002	0.007	0.004	0.123	-0.084	0.021	-

Note: Only major variables are reported. The correlations of these variables with other variables were generally small.

2.5 Estimation and Results

2.5.1 Identification Strategies

Our first identification strategy for the impacts of UGC and MGC is based on the Propensity Score Matching (PSM) method (Heckman et al. 1998; Rosenbaum and Rubin 1983). The PSM method enables us to control for self-selection at the fan page level (Moe and Schweidel 2012) via constructing a "control" group of matched 398 customers[13] who were in the reward program but did not join the FFS brand community. Specifically, we use the PSM method to identify a matched sample based on the one-to-one nearest-neighbour matching (without caliper) algorithm to generate a "control" group with a comparable sample size (i.e., 398 non-fan consumers), which is recognized as the optimal matching method in the literature (Austin 2010). The propensity score is computed using observed individual consumer characteristics, i.e., (1) age (*AGE*), (2) income level (*INC*), (3) gender (*MALE*), (4) home phone number disclosure (*PHONE_DIS*), (5) home address disclosure (*ADDRESS_DIS*), (6) mobile phone opt-in (*PHONE_OPT*) and (7) mail opt-in (*MAIL_OPT*) for promotional information. We believe these set of consumer covariates are comprehensive and informative, such that they influence the "treatment" assignment (i.e., joining FFS Facebook fan page as a fan) and yet are not affected by the "treatment", thus satisfying the unconfoundedness or selection-on-observables identification assumption of PSM.

Ideally, one would expect the matched sample to be as similar as possible to the "treated" sample of 398 fans, and in general for the distribution of all observed characteristics of the two groups to be identical after the PSM procedure. Table 1-4 presents the t-test results of the mean differences between the treated and matched groups in terms of the above seven characteristics. After matching, the two groups have no significant differences across all characteristics. The major difference now between these two groups is that consumers in the "treatment" group were fans on FFS retailer's Facebook fan page and thus could get exposed to UGC and MGC, whereas those in the "control" group were not fans and thus had no exposure to UGC

[13] The average purchase expenditure for these 398 propensity-score matched customers was $34.63.

or MGC. Given that consumers across the "control" and "treatment" groups were essentially identical to one another across the set of exogenous variables (age, income, gender, home phone and address disclosures, mobile phone and mail opt-ins for marketing information) used as the criteria for matching, self-selection at the fan page level based on these observed attributes is thus controlled for.

Table 1-4 - Propensity Score Matching T-test Results

Variable	Sample	Mean		t-test	
		Treated	Control	t	*p*>|t|
AGE	Unmatched	32.528	34.306	-4.93	0.000
	Matched	32.528	33.139	-1.31	0.190
INC	Unmatched	2.339	2.716	-6.79	0.000
	Matched	2.339	2.387	-0.74	0.458
MALE	Unmatched	0.121	0.137	-0.94	0.348
	Matched	0.121	0.113	0.33	0.741
PHONE_DIS	Unmatched	0.887	0.105	50.25	0.000
	Matched	0.887	0.887	-0.00	1.000
ADDRESS_DIS	Unmatched	0.751	0.789	-1.83	0.067
	Matched	0.751	0.706	1.43	0.152
PHONE_OPT	Unmatched	0.392	0.445	-2.09	0.037
	Matched	0.392	0.420	-0.79	0.428
MAIL_OPT	Unmatched	0.060	0.090	-2.05	0.040
	Matched	0.060	0.055	0.30	0.762

PSM however does not allow for selection on unobservables (which our next two identification strategies allow), and thus can only match based on observed attributes, but not unobserved, potentially confounding factors[14]. Another limitation is that PSM can only estimate "treatment" effects where there is support for the "treated" individuals among the "non-treated" population. Lastly, as is the case with other partial equilibrium evaluation methods, PSM cannot establish the impact of the "treatment" beyond the eligible group of consumers.

[14] Sensitivity tests to check on potential deviations from unconfoundedness reveal that the Γ cutoff value is 1.3 (1.35) before an upper bound of significance value reaches above 0.05 (0.10). This implies that to attribute a higher level of purchase expenditure due to an unobserved covariate, rather than to joining FFS's fan page, that unobserved covariate would need to produce a 30%-35% increase in the odds of joining FFS's fan page. This thus quantifies the extent of insensitivity of our PSM results to biases from potential unobserved factors.

With the matched customer data sample, our second identification strategy exploits differences across consumers' fan page joining decision and across timing differences in fan page joining dates to use a difference-in-differences (DID) model estimation approach. This thus enables us to estimate the economic impact (i.e., "treatment" effect) of joining the FFS brand community. While the DID approach allows for selection on (time-invariant) unobservables, there are limitations to this method. First, the DID approach is valid only when the treatment is as good as random when conditioned on individual, group and time fixed effects. Second, the validity of DID estimates may be threatened by the potential endogeneity of the treatments or interventions themselves (e.g., in our context, if loyal consumers have a time-varying propensity to join the retailer's fan page). Lastly, DID model estimations may be susceptible to serial correlation problems (Bertrand et al. 2004).

Furthermore, with the same matched data sample, our third identification strategy uses a Heckman selection model to quantify the effects of directed and undirected UGC and MGC, while controlling for potential self-selection at other levels such as content generation and network ties, or that associated with unobserved factors. The Heckman selection model takes on specific normal distribution assumptions for the unobservable characteristics that jointly influence the fan page selection decision and the purchase outcome. The estimated model parameters may thus be sensitive to these distributional assumptions of the residuals that provide a technical basis of the Heckman model's identification (which need not rely strictly on the variation in the explanatory variables). Another limitation is that model estimation results are unreliable if there are no exclusion restrictions (i.e., at least one exogenous independent variable from the first-stage selection model is excluded from the set of independent variables for the second-stage model).

2.5.2 Preliminary Analysis and Results

Prior to estimating our main model specification shown in Equation (9), we first conduct a preliminary analysis using a baseline alternative model with a series of main effects and interactions between the four variables of the source of content (UGC/MGC), directed/undirected communication mode, content information richness and valence. This preliminary analysis seeks to examine the impact of information richness (*IR*) and valence (*VA*) of network interaction contents on consumer purchase behavior, and then further investigates how *IR* and *VA* depend on content source (*SOURCE*, i.e., UGC volume/MGC volume ratio) and communication mode (*MODE*, i.e., directed content volume/undirected content volume ratio). The baseline alternative model specification is given in Equation (13):

$$\begin{aligned}EXPEND_{it} = {} & \beta_1 IR_{i,t-1} * SOURCE_{i,t-1} * MODE_{i,t-1} + \beta_2 VA_{i,t-1} * SOURCE_{i,t-1} * MODE_{i,t-1} \\ & + \beta_3 IR_{i,t-1} * SOURCE_{i,t-1} + \beta_4 IR_{i,t-1} * MODE_{i,t-1} \\ & + \beta_5 VA_{i,t-1} * SOURCE_{i,t-1} + \beta_6 VA_{i,t-1} * MODE_{i,t-1} \\ & + \beta_7 SOURCE_{i,t-1} * MODE_{i,t-1} \\ & + \beta_8 IR_{i,t-1} + \beta_9 VA_{i,t-1} + \beta_{10} SOURCE_{i,t-1} + \beta_{11} MODE_{i,t-1} \\ & + ControlVariables + \alpha_i + \varepsilon_{it}\end{aligned} \tag{13}$$

Table 1-5 shows the descriptive statistics of the model covariates for the alternative model in Equation (13), and draws comparison to the variables for the main model specification in Equation (9). We note that the variables in the alternative model generally have lesser dispersions in the variable values (as evidenced by the ratio of std. dev. over mean values) than those in the main model. This is because variables in the alternative model are aggregated from all consumers, and the only variation comes from the (exclusion of the) focal consumer *i*’s own postings and comments which are much smaller in volume compared to the aggregated postings and comments of other consumers.

Table 1-5 - Descriptive Statistics of Main and Alternative Model Variables

Model	Variable	Mean	Std. Dev.	Ratio: Std.Dev./Mean	Min	Max
Alternative model	*IR*	3.958	2.394	0.605	0.171	12.000
	VA	0.273	0.517	1.894	-2.000	2.400
	SOURCE	1.600	1.233	0.771	0.000	6.000
	MODE	0.00009	0.002	22.222	0.000	0.122
Main model	*U_D_IR*	0.006	0.177	29.500	0.000	12.000
	U_U_IR	3.143	2.021	0.643	0.000	14.000
	U_D_VA	-0.00005	0.019	-380.000	-1.000	1.000
	U_U_VA	0.181	0.539	2.978	-3.000	2.000
	M_D_IR	0.037	0.896	24.216	0.000	48.000
	M_U_IR	7.010	3.359	0.479	0.000	16.000
	M_D_VA	0.004	0.166	41.500	-4.000	9.000
	M_U_VA	0.705	0.987	1.400	-2.000	4.000

We first estimate a model with only the four main effect variables (plus other control variables), using fixed effects (FE), random effects (RE) and Heckman selection model specifications. Table 1-6 presents the results. The main effects model estimation results reveal significant positive main effects of *IR* and *VA* that are consistent with prior studies on online WOM. Next, we estimate a model with both the main effects and interaction effects variables, and find a significant main effect of *VA* and also importantly, a significant interaction effect of *SOURCE*MODE*. This significant interaction coefficient thus indicates the importance of content source and communication mode, providing support to investigating content source and communication mode in brand community networks according to the main model specification given in Equation (9).

Table 1-6 - Alternative Model: Main and Interaction Effects

Variable	(1) FE: Main	(2) RE: Main	(3) Heckman: PSM, FE Main	(4) Heckman: Population Main	(5) FE: Main + Int	(6) RE: Main + Int	(7) Heckman: PSM, FE Main + Int	(8) Heckman: Population Main + Int
*IR*SOURCE*MODE*					-179.448	-176.764	-179.448	-169.505
					(161.499)	(160.915)	(159.753)	(163.819)
*VA*SOURCE*MODE*					-280.332	-481.736	-280.332	-717.111
					(4,944.674)	(4,927.523)	(4,891.233)	(5,015.946)
*IR*SOURCE*					0.109	0.111	0.109	0.109
					(0.076)	(0.076)	(0.075)	(0.077)
*IR*MODE*					227.878	217.039	227.878	200.887
					(356.037)	(354.976)	(352.189)	(361.489)
*VA*SOURCE*					0.062	0.047	0.062	0.028
					(0.501)	(0.501)	(0.495)	(0.513)
*VA*MODE*					509.820	819.780	509.820	1,150.584
					(8,093.279)	(8,066.824)	(8,005.809)	(8,213.037)
*SOURCE*MODE*					257.017***	257.109***	257.017***	262.021***
					(80.214)	(79.812)	(79.347)	(81.147)
IR	0.148*	0.148*	0.148*	0.138*	0.012	0.010	0.012	0.004
(Information richness)	(0.082)	(0.081)	(0.081)	(0.083)	(0.120)	(0.120)	(0.119)	(0.123)
VA	1.753***	1.734***	1.753***	1.683***	1.743***	1.732***	1.743***	1.691***
(Valence)	(0.336)	(0.336)	(0.332)	(0.343)	(0.437)	(0.437)	(0.433)	(0.448)
SOURCE	0.017	0.021	0.017	0.022	-0.303	-0.302	-0.303	-0.292
(Content source)	(0.149)	(0.149)	(0.147)	(0.152)	(0.242)	(0.242)	(0.239)	(0.247)
MODE	113.126	106.205	113.126	92.612	-376.278	-372.923	-376.278	-384.226
(Communication mode)	(71.119)	(70.803)	(70.363)	(71.954)	(374.785)	(373.273)	(370.735)	(379.637)
Constant	3.611***	3.639***	-0.575	4.704***	4.108***	4.147***	0.009	5.159***
	(0.497)	(0.555)	(5.503)	(0.929)	(0.579)	(0.632)	(5.511)	(0.971)
Control variables	-included-	-included-	-included-	-included-	-included-	-included-	-included-	-included-
Number of consumers	398	398	796	14,388	398	398	796	14,388
Number of observations	20,406	20,406	52,250	840,708	20,406	20,406	52,250	840,708
R^2	0.0022	0.0022	-	-	0.0030	0.0030	-	-
Wald χ^2	-	-	1,517.990	40.890	-	-	1,534.900	56.650

Note: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

2.5.3 Main Analysis and Results

In our main analysis, we first estimate a FE model and a RE model of consumers' purchase expenditure (*EXPEND*) on all control variables which have been widely recognized as important factors affecting consumer purchase behavior. As reported in Table 1-7, Columns (1) and (2), a few control variables such as prior purchase expenditure and UGC volumes have explanatory power[15].

Table 1-7 - Model Estimation Results

Variable	(1) FE: Control	(2) RE: Control	(3) DID: PSM, TE	(4) FE: Full	(5) RE: Full	(6) Heckman: PSM, FE	(7) Heckman: Population
U_D_IR				3.225*	3.195*	3.182*	3.523*
(UGC, directed, information)				(1.863)	(1.849)	(1.838)	(1.873)
U_U_IR				21.849***	22.042***	21.317***	22.973***
(UGC, undirected, information)				(7.994)	(7.977)	(7.891)	(8.105)
U_D_VA				6.641	6.195	6.603	5.290
(UGC, directed, valence)				(9.009)	(8.996)	(8.883)	(9.138)
U_U_VA				76.733**	77.793**	74.311**	81.355**
(UGC, undirected, valence)				(33.224)	(33.151)	(32.819)	(33.708)
M_D_IR				-0.437	-0.422	-0.448	-0.400
(MGC, directed, information)				(0.389)	(0.387)	(0.386)	(0.393)
M_U_IR				-14.209	-13.352	-15.882	-11.962
(MGC, undirected, information)				(22.570)	(22.526)	(22.493)	(23.118)
M_D_VA				3.383**	3.234**	3.372**	2.800*
(MGC, directed, valence)				(1.607)	(1.600)	(1.570)	(1.606)
M_U_VA				71.473	66.878	76.714	59.933
(MGC, undirected, valence)				(86.292)	(86.095)	(84.069)	(86.379)
*BrandCom*BecomeFan*			24.597***				
(DID treatment effect)			(2.040)				
U_D_VO	0.751*	0.798*		0.910**	0.959**	0.917**	1.101**
(UGC, directed, volume)	(0.410)	(0.408)		(0.462)	(0.460)	(0.456)	(0.468)
U_U_VO	0.199	0.222		0.089	0.114	0.099	0.157
(UGC, undirected, volume)	(0.233)	(0.232)		(0.249)	(0.248)	(0.245)	(0.250)

Note: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

[15] The model fit statistics R^2 of the fixed and random effects models shown in Table 1-7 are generally small. This is because our estimation data have many zero-expenditure weeks of each customer. Dropping these zero-expenditure weeks increases the R^2 of the estimated models to about 0.107 to 0.139, but this would omit relevant UGC and MGC information that may bias the results. Our research does not involve forecasting, thus R^2 model fit may matter less.

Table 1-7 - Estimation Results (Continued)

Variable	(1) FE: Control	(2) RE: Control	(3) DID: PSM, TE	(4) FE: Full	(5) RE: Full	(6) Heckman: PSM, FE	(7) Heckman: Population
M_D_VO	-6.810**	-3.655		-8.772	-7.504	-6.771	-8.303
(MGC, directed, volume)	(3.331)	(2.382)		(5.410)	(17.361)	(17.176)	(17.644)
M_U_VO	-2.059	1.294		0.559	1.967	2.371	2.057
(MGC, undirected, volume)	(2.484)	(0.841)		(2.954)	(16.576)	(16.401)	(16.855)
OWN_VA	-4.845	-4.347		9.672	10.372	9.138	11.900
(Own posting valence)	(12.260)	(12.257)		(13.646)	(13.639)	(13.470)	(13.891)
OWN_VO	9.443***	9.528***		4.826	4.927	4.908	5.079
(Own posting volume)	(3.054)	(3.048)		(3.363)	(3.357)	(3.313)	(3.406)
CENT	-3.252	-2.716		-10.008	-9.348	-10.009	-10.401
(Degree centrality)	(15.716)	(15.552)		(16.183)	(16.021)	(15.984)	(16.231)
FB_V		-0.001			-0.002	0.076	-0.001
(# of Facebook page views)		(0.002)			(0.002)	(0.126)	(0.001)
FB_F		-0.001			-0.001	0.227	-0.001**
(# of Facebook friends)		(0.001)			(0.001)	(0.275)	(0.000)
FFS_F		0.038			0.037	6.402	0.054**
(# of Facebook friends on FFS)		(0.047)			(0.047)	(7.895)	(0.024)
PRICE	0.188	0.153	0.107	1.171	1.240	1.337	1.214
(Product price)	(0.237)	(0.136)	(0.065)	(1.764)	(1.555)	(1.556)	(1.600)
PROM	5.983	-68.137	-32.551*	38.603	-670.859	-703.456	-649.995
(Promotion intensity)	(22.780)	(50.279)	(18.589)	(261.978)	(877.171)	(863.461)	(887.737)
PEXP	-0.029***	-0.002	0.034***	-0.029***	-0.002	-0.029***	0.028***
(Past expenditure)	(0.011)	(0.008)	(0.006)	(0.011)	(0.008)	(0.011)	(0.006)
AGE		0.017	0.045		0.017	-5.370	0.033
(Age)		(0.051)	(0.068)		(0.052)	(6.312)	(0.028)
INC		-0.117	0.305		-0.113	310.863	0.022
(Income level)		(0.386)	(0.500)		(0.387)	(375.315)	(0.201)
MALE		0.023	-0.981		0.022	-553.704	0.060
(Gender)		(0.963)	(1.238)		(0.966)	(673.973)	(0.513)
Constant	-12.676	35.117	12.051	-200.030	481.263	0.000	455.053
	(23.447)	(22.325)	(11.118)	(125.129)	(638.524)	(0.000)	(643.396)
Time dummies	-included-	-included-	-included-	-included-	-included-	-included-	-included-
Number of consumers	398	398	796	398	398	796	14,388
Number of observations	20,406	20,406	61,160	20,406	20,406	52,250	840,708
Hausman test / Selection ρ	$\chi^2 = 8.36, p = 0.99$		-	$\chi^2 = 0.69, p = 0.99$		$\rho = 0.000$	$\rho = -0.066$
R^2	0.0240	0.0273	-	0.0246	0.0279	-	-
Wald χ^2	-	600.00	871.46	-	613.78	3198.25	612.30

Note: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Next, before we examine the impact of the various UGC and MGC factors of interest, we estimate a DID model to compare consumer purchase expenditure between fans and non-fans, as well as before and after becoming a fan of FFS brand community. Specifically, we created an estimation data sample of 796 consumers, combining the 398 PSM-matched consumers with the original 398 consumers who were fans of the FFS fan page. We use a binary variable, *BrandCom*, to indicate whether each of the 796 consumers was a fan in the brand community (1: fan, 0: non-fan). We then use an additional binary variable, *BecomeFan*, to indicate the timing of becoming a fan (1: after, 0: before) for the 398 fans, and interact it with *BrandCom* (i.e., *BrandCom*BecomeFan*). As *BrandCom* and *BecomeFan* might be endogenous, we first use several exogenous variables (*AGE*, *INC*, *MALE*, *PHONE_DIS*, *ADDRESS_DIS*, *PHONE_OPT* and *MAIL_OPT*) in a Probit model to model the outcome of an unobserved latent variable determining the selection decisions. We thus estimate a treatment effects (TE) model focusing on the coefficient for *BrandCom*BecomeFan*, while controlling for the various control variables. As shown in Table 1-7, Column (3), the DID parameter estimate is 24.597 ($\pm$2.040), which is significantly positive. This implies a significant positive impact of about $24.60 in purchase expenditure after joining the brand community of FFS retailer. The exposure to UGC and MGC thus has a significant impact on purchase behavior, which gives credence to further explore the impact of different UGC and MGC factors in depth.

We further estimate a full FE model, including all the UGC and MGC factors of focal interest. Table 1-7, Column (4), reports the results. For UGC factors, both information richness and valence are found to have a significant impact on *EXPEND*. Specifically, the coefficients of *U_D_IR* (3.225$\pm$1.863), *U_U_IR* (21.849$\pm$7.994) and *U_U_VA* (76.733$\pm$33.224) are positive and statistically significant. For MGC factors, only valence, i.e., *M_D_VA* (3.383 $\pm$1.607) is found to have a positive and significant impact on *EXPEND*. Next, we further estimate a full RE model. In Table 1-7, Column (5), the RE model shows similar results to those in Column (4). The Hausman test suggests that the RE estimates are not inconsistent (χ^2 = 0.69, p = 0.99). Nevertheless, we prefer the FE model over the RE model since the former allows the consumer-specific unobserved heterogeneity to be correlated to the observed variables (i.e., a more tenable assumption), and its estimation involves a conditional analysis restricted to a specific sample (thus matching our data from the FFS reward program).

Both the prior FE and RE model estimation results have not accounted for potential self-selection at the fan page level. To control for self-selection as a potential confounding factor in determining the effects of consumers' exposure to UGC and MGC on their purchase behavior, we use as model estimation sample, the PSM-matched 398 non-fan consumers as a control group in addition to the original 398 fans. We use *BrandCom* to indicate whether each of the 796 consumers was a fan in FFS retailer's fan page brand community. We then employ the Heckman two-stage selection model (Heckman 1976; Heckman 1979), including a full set of exogenous consumer-specific covariates in the first stage to model the selection decision. Results of the first-stage Probit model estimation are shown in Table 1-8. In the second stage, besides the focal UGC and MGC factors and control variables, we also include consumer fixed effects to account for consumer heterogeneity in the purchase expenditures. As indicated in Table 1-7, Column (6), the estimates are consistent with those in the FE model. Specifically, the parameter estimates for the focal UGC and MGC factors of *U_D_IR* (3.182 ± 1.838), *U_U_IR* (21.317 ± 7.891), *U_U_VA* (74.311 ± 32.819), and *M_D_VA* (3.372 ± 1.570) are all statistically significant. Thus, the information richness of both directed and undirected UGC have a positive influence on consumer purchase expenditure, but not for the case of MGC. In terms of content valence, the valence of directed MGC has a positive effect on expenditure while that for directed UGC does not. However, while the valence of undirected UGC has a large positive effect on expenditure, there is no effect of undirected MGC.

Table 1-8 - First-Stage Estimation Results of Heckman Selection Model

Variable	(1) Heckman-Stage 1 PSM, FE	(2) Heckman-Stage 1 Population
AGE	-0.011***	-0.019***
(Age)	(0.001)	(0.001)
INC	-0.009	-0.093***
(Income level)	(0.007)	(0.004)
MALE	-0.071***	-0.171***
(Gender)	(0.018)	(0.012)
PHONE_DIS	-0.137***	1.919***
(Home phone disclosure)	(0.027)	(0.011)
ADDRESS_DIS	0.235***	-0.299***
(Home address disclosure)	(0.015)	(0.010)
PHONE_OPT	-0.044***	-0.190***
(Phone opt-in)	(0.012)	(0.008)
MAIL_OPT	0.119***	-0.367***
(Mail opt-in)	(0.024)	(0.015)
Constant	0.060*	-1.754***
	(0.035)	(0.021)
Number of consumers	796	14,388
Number of observations	52,250	840,708
Log likelihood	-34701.307	-63,238.339
Pseudo R^2	0.0073	0.3415

Note: * p<0.1, ** p<0.05, *** p<0.01.

Importantly, the mean of all the four statistically significant UGC and MGC parameter estimates average to about 25.546, which is very close to the DID *BrandCom*BecomeFan* parameter estimate of 24.597 from Table 1-7, Column (3).

To further establish the robustness of results from the Heckman selection model on the PSM-matched data sample, we also estimate the Heckman model using consumers from the rest of the entire customer reward program database (i.e., the 13,990 non-fan consumers) as the control group[16]. We report in Table 1-7, Column (7) results that are consistent with those in Column (6).

Noteworthy, the Heckman model accounts for selection on unobservables and also potential selection at the content generation and consumption level, since it includes control variables of a fan's own posting valence (*OWN_VA*) and volume (*OWN_VO*) in the brand community, as well as a fan's number of Facebook page views (*FB_V*). Finally, our model attempts to account for selection at the network-tie or peer influence level by

[16] We cannot include all consumer fixed effects due to PC memory limitations with the large number of consumers.

including control variables associated with the social network circles of a consumer, i.e., a fan's network degree centrality based on interactions solely on the FFS fan page (*CENT*), number of Facebook friends (*FB_F*), and number of Facebook friends who were also in the FFS brand community (*FFS_F*). Therefore, the above controls give further credence to the impacts of UGC and MGC information richness and valence on purchase behavior, after having accounted for observed and unobserved potentially confounding factors.

In summary, we consider the Heckman two-stage selection model based on the PSM-matched control group (Table 1-7, Column (6)) as our best model, since it accounts for selection bias and consumer-specific heterogeneity. To compare the relative impact of UGC and MGC in terms of information richness and valence, and also the relative impact of directed and undirected communication modes, we report the marginal effects and elasticities for the significant UGC and MGC factors in Table 1-9 based on the main model. We summarize our hypothesis testing results in Table 1-10. For information richness, only UGC factors, *U_D_IR* (marginal effect = 3.182, $p < 0.1$) and *U_U_IR* (marginal effect = 21.317, $p < 0.01$), are significant, thus supporting H1B and rejecting its competing hypothesis H1A. For valence, the significant marginal effect of the UGC factor, *U_U_VA* (marginal effect = 74.311, $p < 0.05$), is more than 22 times that of the only significant MGC factor, *M_D_VA* (marginal effect = 3.372, $p < 0.05$), thus supporting H2. Finally, as for directed and undirected communication modes, UGC information richness and valence are generally significant and with larger marginal effects in the undirected mode, thus rejecting H3. On the contrary, for MGC, valence is significant in the directed communication mode only, thus supporting H4A and rejecting its competing hypothesis H4B.

Table 1-9 - Marginal Effects and Elasticities

UGC factors	*U_D_IR*	*U_U_IR*	*U_D_VA*	*U_U_VA*
Marginal effect	3.182*	21.317***	-	74.311**
Elasticity	0.006*	3.140***		0.180**
MGC factors	*M_D_IR*	*M_U_IR*	*M_D_VA*	*M_U_VA*
Marginal effect	-	-	3.372**	-
Elasticity			0.004**	

Note: * p<0.1, ** p<0.05, *** p<0.01.

Table 1-10 - Hypothesis Testing Results

Hypothesis		Support
H1A, competing	UGC information richness < MGC information richness	No
H1B, competing	UGC information richness > MGC information richness	Yes
H2	UGC valence > MGC valence	Yes
H3	UGC: directed communication > undirected communication	No
H4A, competing	MGC: directed communication > undirected communication	Yes
H4B, competing	MGC: directed communication < undirected communication	No

2.5.4 Robustness Checks

We further corroborate our main findings by checking its robustness in multiple ways. For ease of reference, Table 1-11, Column (1), presents the main results from Table 1-7, Column (6). For brevity, from this point onward, we only report the major variables of interest for hypotheses testing.

First, we examine the effects of UGC and MGC factors without accounting for communication intensity. We remove all intensity elements from Equations (1), (2), (5) and (6) in Section 2.4.3, and compute only the average UGC and MGC factors for directed communication. Table 1-11, Column (2) shows the model estimates, which qualitatively remain consistent with our main results. However, the comparatively larger and thus potentially misleading coefficient size of the *M_D_VA* parameter (i.e., 7.234), relative to those from all the other models, highlights the importance of accounting for communication intensity.

Second, we check the robustness of our findings across different model specifications. We first estimate a population-averaged (PA) model that allows for an exchangeable correlation structure of a generalized linear model, and then a random effects model estimated via maximum likelihood (RE-ML). The corresponding results for the PA and RE-ML models are shown in Table 1-11, Columns (3) and (4). The model parameter estimates remain consistent with those of the main one in Column (1).

Next, to account for the existence of potential serial correlation, we estimate a FE model with a first-order autoregressive disturbance structure (FE-AR1). As indicated in Table 1-11, Column (5), the model estimates under an AR1 structure are consistent with those of the main one. This implies that findings from our main model in Column (1) are robust to serial correlation.

Lastly, Table 1-12 presents the Heckman two-stage selection model and FE model estimation results based on different time lag levels (*t*-1, *t*-2, *t*-3). The comparison suggests that a one-period (*t*-1) lag level is sufficient and possesses the best model fit.

Table 1-11 - Robustness Checks

Variable	(1) Main	(2) Intensity	(3) PA	(4) RE-ML	(5) FE-AR1
U_D_IR	3.182*	3.382*	3.193*	3.190*	3.531*
(UGC, directed, information)	(1.838)	(1.829)	(1.846)	(1.843)	(2.094)
U_U_IR	21.317***	21.658***	22.035***	22.012***	22.748***
(UGC, undirected, information)	(7.891)	(7.904)	(7.963)	(7.950)	(8.104)
U_D_VA	6.603	4.938	6.203	6.229	7.804
(UGC, directed, valence)	(8.883)	(8.759)	(8.981)	(8.966)	(9.006)
U_U_VA	74.311**	75.959**	77.761**	77.659**	85.498**
(UGC, undirected, valence)	(32.819)	(32.879)	(33.094)	(33.039)	(33.801)
M_D_IR	-0.448	-0.861	-0.422	-0.423	-0.467
(MGC, directed, information)	(0.386)	(0.865)	(0.386)	(0.386)	(0.388)
M_U_IR	-15.882	-12.976	-13.369	-13.426	-6.860
(MGC, undirected, information)	(22.493)	(22.407)	(22.487)	(22.449)	(22.920)
M_D_VA	3.372**	7.234*	3.237**	3.246**	2.764*
(MGC, directed, valence)	(1.570)	(3.979)	(1.598)	(1.595)	(1.645)
M_U_VA	76.714	81.518	66.958	67.220	30.434
(MGC, undirected, valence)	(84.069)	(89.616)	(85.947)	(85.802)	(89.035)
Constant	0.000	0.000	481.495	482.251	418.512
	(0.000)	(0.000)	(637.421)	(636.340)	(608.180)
Control variables	-included-	-included-	-included-	-included-	-included-
Number of consumers	796	796	398	398	398
Number of observations	52,250	52,250	20,406	20,406	20,009
Wald χ^2	3198.25	3196.86	615.95	608.85	-

Note: Standard errors in parentheses. * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

Table 1-12 - Models with Different Time Lags

Variable	(1) Heckman *t*-1	(2) Heckman *t*-2	(3) Heckman *t*-3	(4) FE *t*-1	(5) FE *t*-2	(6) FE *t*-3
U_D_IR	3.182*	4.651**	2.035	3.225*	4.593**	2.079
(UGC, directed, information)	(1.838)	(1.836)	(1.832)	(1.863)	(1.862)	(1.858)
U_U_IR	21.317***	10.745	31.764***	21.849***	10.469	32.123***
(UGC, undirected, information)	(7.891)	(7.876)	(7.848)	(7.994)	(7.980)	(7.954)
U_D_VA	6.603	2.087	0.170	6.641	1.932	-0.011
(UGC, directed, valence)	(8.883)	(8.864)	(8.832)	(9.009)	(8.992)	(8.962)
U_U_VA	74.311**	36.842	105.436***	76.733**	35.620	106.959***
(UGC, undirected, valence)	(32.819)	(32.755)	(32.642)	(33.224)	(33.166)	(33.058)
M_D_IR	-0.448	0.329	-0.384	-0.437	0.353	-0.397
(MGC, directed, information)	(0.386)	(0.386)	(0.385)	(0.389)	(0.389)	(0.388)
M_U_IR	-15.882	7.452	5.222	-14.209	10.228	4.799
(MGC, undirected, information)	(22.493)	(22.448)	(22.370)	(22.570)	(22.529)	(22.455)
M_D_VA	3.372**	-1.889	-0.111	3.383**	-2.013	-0.038
(MGC, directed, valence)	(1.570)	(1.567)	(1.562)	(1.607)	(1.605)	(1.600)
M_U_VA	76.714	7.488	-34.008	71.473	-6.908	-31.494
(MGC, undirected, valence)	(84.069)	(83.903)	(83.618)	(86.292)	(86.139)	(85.866)
Constant	0.000	0.000	-510.636	-200.030	-27.254	14.672
	(0.000)	(0.000)	(313.979)	(125.129)	(97.639)	(98.789)
Control variables	-included-	-included-	-included-	-included-	-included-	-included-
Number of consumers	796	796	796	398	398	398
Number of observations	52,250	51853	51457	20,406	20,009	19,613
R^2	-	-	-	0.0246	0.0238	0.0236
Wald χ^2	3198.25	3109.81	2064.67	-	-	-

Note: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

In sum, we are confident of the robustness of our findings given that various checks indicated robustness and consistency.

2.6 Discussion and Contribution

2.6.1 Discussion of Findings

Our study which investigates the impact of social media brand community network interaction contents on consumer purchase behavior has several notable findings. First, we empirically show that engagement in social media brand community networks leads to a significant increase in consumer purchases. Second, our in-depth examination of network contents (UGC and MGC) attests to the fact that brand community network contents affect consumer purchase behavior through embedded information as well as persuasion. Besides UGC, MGC in the network also matter, but differently, in influencing consumer purchases. Consumers influence the purchase expenditure of one another through both informative as well as persuasive interactions, whereas marketers influence it only through persuasive communication. Interestingly, consumers' persuasive effect is more than 22 times that of marketer's in terms of marginal effect. The elasticities of demand with respect to UGC's informative effect (directed), informative effect (undirected), and persuasive effect (undirected) are estimated to be 0.006, 3.140 and 0.180 respectively, while that for MGC's persuasive effect (directed) is 0.004. Overall, UGC exhibits a more influential role than MGC in driving purchases.

Finally, evidence affirms directed communication and undirected communication matter differently for UGC and MGC. Specifically, in driving purchases, undirected contents are more effective than directed ones for both informative and persuasive consumer-to-consumer communication, while directed contents are more effective than undirected ones for persuasive marketer-to-consumer communication. For the rejected hypothesis H3, a plausible reason might be due to the manner that posts and comments on Facebook are structured or displayed. UGC undirected communications typically appear as posts on a fan page with the most recent post appearing in the most salient top-most position which can garner the most attention. In contrast, comments are sorted in the opposite manner with the most recent one listed at the bottom.

2.6.2 Theoretical Contributions

Our study offers theoretical contributions in the following ways. First, the predominant emphasis of prior brand community research on consumer engagement and content (i.e., consumer side) (e.g., Algesheimer et al. 2005; Bagozzi and Dholakia 2006; Porter and Donthu 2008) may have unwittingly result in the misconception that businesses can only passively react. By accentuating the role of MGC and its impact (i.e., marketer side), we underscore that marketers can actually transform their role from a passive and reactive one to a proactive and influential one. By actively engaging consumers in the network, marketers can better reap economic values.

Second, by juxtaposing the role of MGC besides that of UGC, we unravel the contention and intricacies between the two, thereby complementing and enriching past works. Our findings suggest that MGC does affect consumer purchase behavior, but in a different way from UGC (e.g., Dhar and Chang 2009; Sonnier et al. 2011; Tumarkin and Whitelaw 2001). Hence, the sole reliance on UGC to explain consumer behavior would overlook and omit the persuasive effect of the marketer's contents. The differential and even contrasting impact of UGC and MGC suggests that consumers not only respond to the information of online contents, but also factor the sources of content into consideration. This provides a foray into better understanding the economic value of content in networks facilitated by social media platforms.

Third, as one of the pioneer efforts to quantify the economic impact of both UGC (or online WOM) and MGC (marketers' proactive marketing activities) in social media-enabled social networks, we augment the discourse on social media marketing with insights on its ROI. Using various identification strategies, we provide a rigorous estimate of the consumer's economic impact of joining brand community networks. Our attempt is also one of the first to empirically quantify the relative effectiveness of UGC and MGC in brand community social network contexts.

Fourth, our research is also amongst the first to propose and validate a model to quantify the economic impact of brand community network contents at the individual consumer level. This approach enables us to control for consumer heterogeneity, selection biases and to address the prior overlooked impact of dyadic communication in terms of the communication modes. Our findings underscore that sharing information alone in brand communities is a necessary but not sufficient condition to generating positive economic outcomes. In addition to contents *per se*, whether contents are communicated in a directed or undirected manner matters.

2.6.3 Practical Implications

Our study has several important practical implications to social media marketers. Consumers (UGC) play both informative and persuasive roles and marketers (MGC) play a persuasive role in social media-enabled social network contexts. This suggests that, a mere reliance on marketers' own marketing activities may not be the most effective way to drive consumer purchases. Similarly, marketers' total reliance on consumers' WOM "buzz" is also suboptimal. An ideal strategy would be the right combination of both

UGC and MGC. Apart from marketers' diligent preparation of their own persuasive content (e.g., use more favorable or positive words and phrases to describe products and services), marketers should conscientiously design campaigns to encourage informative, and especially, persuasive communication among consumer themselves. For instance, marketers can incentivize consumers to share their experiences by using discount coupons and reward points. Not unheard of, there are also marketers who employ a "community manipulation strategy" (Dellarocas 2006) by anonymously behaving as "fellow consumers" to share positive product information within communities.

Second, directed messaging is more effective for persuasive marketer-to-consumer communication, whereas undirected messaging is more effective for informative and persuasive consumer-to-consumer communication, in driving consumer purchases. Hence, when communicating persuasive content to consumers, marketers can choose a directed communication mode for higher ROI impact. In other words, they can generate content to a targeted user or group for better consumer responses. For instance, in the context of Facebook, a marketer can direct marketing communication in the "comment" entries of the fan page to address specific consumers. With regard to informative and persuasive communications among consumer themselves, marketers can encourage consumers to engage more in undirected communications. For instance, marketers can reward consumers who are most active in sharing their content in posts addressed to the fan page members at large.

Third, marketers might want to enhance their analytics by moving beyond the traditional insights from quantitative analysis, such as the identification of the advertising expenditure-sales relationship, to embrace more insights from qualitative analysis as well. Currently available qualitative tools such as the one we adopted can help track, analyze and enlighten the content embedded within UGC in their brand communities. Marketers can then get a more nuanced understanding of consumers' general response, attitude toward and evaluation of the products and marketing campaigns launched.

Finally, our study also presents implications for the design of social media marketing platforms. Many current platforms (e.g., Yelp.com) are popular, and have attracted extensive information sharing in the form of reviews from consumers. However, these platforms do not currently provide much access to marketers' proactive engagements. Indeed, our study suggests that these platforms can actually do better by enabling marketers' engagements. For instance, apart from displaying consumer reviews of a restaurant, social media platforms can also provide free or paid access to marketers from a restaurant to communicate marketing information (e.g., introduction of new cuisines, replies to customers' queries) and to integrate functional aspects of customer relationship management within the social media platform.

2.7 Conclusion

While this research has highlighted several notable findings, we acknowledge some limitations. First, our research context does not entail randomized trials or field experimentations on the UGC and MGC constructs of interest. As such, while we spent considerable efforts in addressing concerns related to selection biases (due to both observables and unobservables), our identification strategies centering on PSM and Heckman selection model only afford us a quasi-counterfactual of a consumer being a brand's fan on Facebook, after accounting for selection on observables and unobservables. Second, apart from textual contents, there were a small number of pictures and videos in our research context. These contents were posted together with some textual descriptions, which at the same time were captured in our sample. Although we were able to account for the impact from all textual contents, we did not account for the other types of content. Third, the data sample for our research context comes from only a single retailer and its consumers as well as brand community members. Nevertheless, the phenomenon of UGC and MGC interactions is not unique to the FFS community on Facebook[17]. Moreover, in terms of the platform used, many

[17] As a quick check of generalizability, we extract UGC and MGC from the Facebook fan page of another well-known apparel brand in the same country/market. We compare (using t-tests) weekly volumes of UGC and

other social media platforms (e.g., MySpace, YouTube) offer similar functionalities for marketers and consumers to engage in social interactions. Reassuringly, the parent retail company of the FFS retailer is well-established as a franchisee of many famous global apparel brands and thus follows both industry recommended practices and brand-guided procedures with regard to social media marketing communications.

Moving forward, we present potential avenues for future research. A meaningful extension to this research is to investigate the role of product type, perhaps in a randomized trial or experimentation setting (Aral and Walker 2011). As discussed, UGC is more consumer-oriented relative to MGC. This may potentially contribute to the stronger role of UGC relative to MGC in our context where experience products (i.e., apparels) were studied. To what extent do our findings apply to search products (e.g., books, plane tickets) context deserves further scrutiny. It might also be worthwhile to study the relative effectiveness of online (UGC and MGC) and offline marketing activities concurrently. Since firms often face limited marketing resources in multichannel marketing settings (Chu et al. 2007; Zhang 2009), assessing their relative effectiveness and identifying the optimal combination of marketing strategies across multiple channels to achieve better sales outcome is vital.

3. STUDY 2: PRODUCT NETWORKS IN ELECTRONIC COMMERCE

3.1 Introduction

E-commerce has experienced a tremendous growth over the last decade. According to Forrester Research, U.S. e-commerce sales have topped $200 billion in 2011 (Mulpuru et al. 2013), and will reach $327 billion by 2016 (Indvik 2012). Capitalizing on its popularity, online retailers have attempted to replicate the “diaper and beer” co-location practice in grocery stores (Srikant and Agrawal 1996). Specifically, on most e-commerce sites, each product is featured on its own designated web page. On each product page, retailers then utilize some recommender systems [18] to explicitly

---

MGC of the FFS retailer’s fan page to those of the other fan page but find no significant difference across these metrics.

[18] In this study, we are interested in online product networks that are created by the use of recommender systems. Thus, we use the terms recommender

recommend additional relevant products, and thus forming a product network. If one analogizes the process of browsing an e-commerce website to walking the aisles of a physical store, the "aisle structure" of the e-commerce website will be defined by this graph of interconnected products, and recommending additional products on a web page is likened to placing additional products on a neighboring shelf. Figure 2-1 illustrates how a directed product network is formed. This phenomenon has attracted some academic research investigating the economic impact (e.g., product sales impact) of product networks (Carmi et al. 2011; Carmi et al. 2010; Oestreicher-Singer and Sundararajan 2012a; Oestreicher-Singer and Sundararajan 2012b). However, critical research gaps still remain to be addressed, which motivate our study.

Figure 2-1 - Product Recommendation Network on Amazon

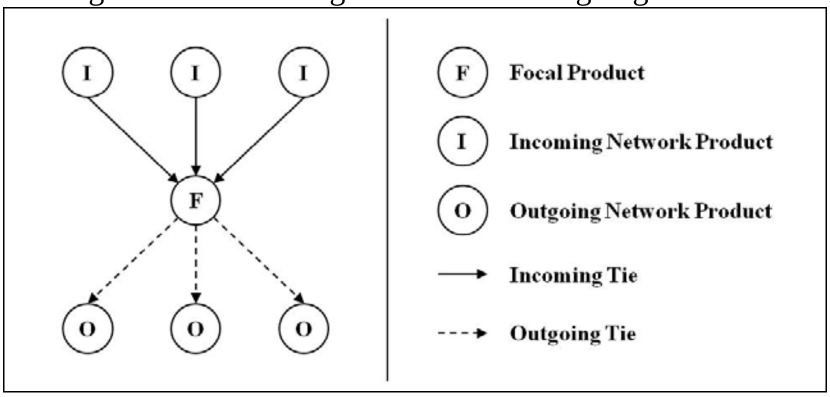
First, past recommendation network studies have overlooked the outgoing product network impact, which might lead to an incomplete and even erroneous analysis. Specifically, past studies (Carmi et al. 2011; Carmi et al. 2010; Oestreicher-Singer and Sundararajan 2012a; Oestreicher-Singer and Sundararajan 2012b) have only examined the demand impact of the incoming network on a focal product. For instance, researchers identified that incoming network connections may increase the exposure and the sales of the focal

---

system, recommendation system, product network, and recommendation network interchangeably.

product (Carmi et al. 2010; Oestreicher-Singer and Sundararajan 2012b). However, past research has overlooked the potential impact of the outgoing network on the demand of the focal product (see Figure 2-2 for an illustration). As a simple example, if a recommended product in the outgoing network better fits a consumer's need, he or she may be tempted to switch from the current product to the better choice. Thus, this suggests that in addition to the impact of the incoming network identified in the past, the concurrent potential impact of the outgoing network should not be ignored. More interestingly, it is not clear whether the overall impact of both the incoming and outgoing networks is positive or negative. As such, whether the introduction of a recommendation system can ultimately generate positive economic impact remains an open question.

Figure 2-2 - Incoming Network and Outgoing Network



Second, past recommendation network studies have invariably treated network structures as given and documented that network structures (e.g., degree centrality) affect product demand (Carmi et al. 2011; Carmi et al. 2010; Oestreicher-Singer and Sundararajan 2012a; Oestreicher-Singer and Sundararajan 2012b). There has been little insight on the programmatic or strategic management of recommendation systems (or network structures[19]) to drive product demand. Specifically, there are plentiful products on a typical e-commerce site, but limited space to display recommendations on a product web page. Thus, retailers first have to make a decision on *which product or category is to be recommended* – a decision that has implications on the

[19] The recommendation system essentially determines the product network structure.

product network's category diversity[20]. Specifically, retailers can recommend a diverse set of categories on the focal product's web page. For instance as illustrated in Figure 2-1, retailers can recommend books on "networks" with diverse topics such as "product networks", "social networks" and "computer networks". The focal product will thus be in a more diverse network. The category diversity in the network is important because consumers may have different needs or preferences (Allenby et al. 1998; Osborne 2011) and are searching for different products or categories online. Thus, offering a diverse set of recommendations may reduce consumers' online product search efforts and improve their search experience. In addition, retailers have to make decision on *how often to replace the current recommendations with new ones* – a decision that has implications on the product network's stability. Updating the recommendations on the focal product's web page is analogous to updating products on the neighboring shelf in a physical store. Changing these product assortments may rejuvenate consumers' curiosity and sense of freshness so that they can be attracted to walk down the aisle (i.e., analogously, click on the recommendations) to continue their product search. Diversity and stability of a focal product's network may affect the demand of the focal product as these attributes allow a diverse and stable set of alternatives, which together with the focal product, to be in the consumers' consideration set to influence consumers' purchase decisions. These issues also parallel the shelf-space management problem in the traditional retail context. Retailers have to

[20] For the network diversity, we focus on the diversity of product categories in the network. Oestreicher-Singer and Sundararajan (2012b) also used a variable of assortative mixing to capture the extent of product category differences in a network and investigate its impact on a focal product's demand. However, our network diversity variable is substantially different from assortative mixing. Conceptually, assortative mixing measures the extent of product category difference between that of the focal product and its network neighbors, whereas network diversity captures the extent of difference among the focal product's network neighbors *per se*. More importantly, network diversity captures more information than assortative mixing. For instance, suppose the focal product is in category A, and the five products in the network are in categories: (1) A, A, B, B and B; or (2) A, A, B, C and D. Then assortative mixing cannot differentiate these two cases but can only show that the category difference is 3/5=0.6, whereas network diversity can differentiate by showing that there are two distinct categories for the first case and four distinct categories for the second case.

strategically arrange products on limited shelf spaces and locations to generate higher profits (Bultez and Naert 1988; Corstjens and Doyle 1981; Van Nierop et al. 2008). The findings could provide important implications to online retailers' strategic management of recommendation systems to achieve product sales, and also the design of e-commerce recommendation systems. However, the issue of how to leverage network structures (or recommendation systems), in terms of network diversity and network stability, to drive product demands has not been investigated and thus remains an interesting and important research question.

Lastly, past research on product recommendation networks (Carmi et al. 2011; Carmi et al. 2010; Oestreicher-Singer and Sundararajan 2012a; Oestreicher-Singer and Sundararajan 2012b) has only studied the demand impact of co-purchase recommendation systems, but did not examine the role of co-view recommendation systems, or particularly, the relative effectiveness between these two recommendation mechanisms. Co-view and co-purchase recommendation systems have commonly co-existed on many e-commerce sites (e.g., Amazon.com, Tmall.com). Both co-view and co-purchase recommendation systems provide additional related products to consumers to assist in their purchase decisions. However, co-view recommendations communicate information about other consumers' e-commerce web site browsing behaviors across different products, while co-purchase recommendations communicate information about other consumers' purchase behaviors of products bought typically in the same shopping session. These two different sets of information may potentially generate different demand impacts since consumers may respond differently to these two recommendation mechanisms. Past research however has not shed light on the demand impact of co-view recommendation systems, especially relative to the effect of co-purchase recommendation systems. Hence, identifying their similarities and differences in terms of the impact on product demand would be meaningful.

Based on the aforementioned research gaps, the objective of this research is three-fold. First, we operationalize the product network in a more complete manner by observing both the incoming and outgoing networks of a focal product to investigate and differentiate between the impacts of the incoming and outgoing networks. Second, we further separate the product network into two types of networks (co-view and co-purchase recommendation networks) to examine and differentiate between the impacts of these two diverse recommendation mechanisms. Third, based on the above operationalizations, we propose and examine the economic impact of two important network factors (i.e., network diversity and network stability). We aim to provide significant insights on the economic impact of product recommendation networks through these investigations. In essence, our research questions are: (1) Is the demand of a product influenced by both the incoming network and outgoing network? (2) How is the demand of a product influenced by product network attributes in terms of network diversity and network stability? (3) How do the diversity and stability effects differ between two types of recommendation networks (co-view and co-purchase)?

To answer our research questions, we collect product recommendations and transactions data from a Nikon digital camera store on Tmall.com. Our econometric specification models products' daily demand as a function of network factors, while controlling for relevant factors at the individual product, pricing, product network, product category, and time unit levels. In particular, our identification strategy for the economic impact of network structures is to identify and control for the implicit demand correlations (i.e., substitution and complementarity effects) and to account for the simultaneity between network structures and product demand. We perform robustness checks to validate the consistency of our findings in the presence of potential collinearity, heteroskedasticity, serial correlation, price endogeneity, and across differences in variable operationalizations, time frames, and product categories.

We find several notable results. First, the demand of a product is influenced by both its incoming network and outgoing network. Second, the category diversity in the incoming network (of co-purchases) increases the demand of the focal product, while that for the outgoing network (of co-purchases) decreases it. Moreover, stability of the co-purchase outgoing network has a negative impact. Overall, the co-purchase recommendation network has a stronger impact on product demand than the co-view network.

This research provides the following contributions. First, our study is the first to identify the economic impacts of both the incoming and outgoing product networks. Second, our research is a pioneering empirical effort to document both positive and negative economic impacts of product recommendations on individual product demand[21]. Third, our research serves as the first attempt to propose and validate the network diversity and stability effects of product recommendation networks. Fourth, we shed light on the differential impacts of the co-view and co-purchase product networks. Fifth, we use rigorous empirical approaches to account for the implicit substitution and complementarity effects in product recommendation networks. Finally, our study also provides implications for online retailers' recommendation-based product marketing strategies and the design of recommendation systems.

3.2 Literature Review

Recommender systems are a specific type of information filtering technique that automatically provides recommendations for items (e.g., music, book, or movie) that might be of interest to a user (Yi et al. 2011), using a model built from the characteristics of an item (content-based approach) or the user's social environment (collaborative filtering approach) (Ricci et al. 2011). Recommender systems are increasingly used in various online communities (Sahoo et al. 2011). Particularly, they have been widely adopted by online retailers (e.g., Amazon.com) to recommend related products to a consumer when the consumer is viewing or searching for a product (Huang et al. 2007).

The popularity of recommender systems has attracted some academic interest. Early research work, mostly in the data mining field, focused on developing and evaluating various recommendation algorithms (Aciar et al. 2007; Herlocker et al. 2004; Iaquinta et al. 2008). Among numerous algorithms, the collaborative filtering approach, which determines recommendations by the levels of similarity of preferences of other consumers, is the most popular in e-commerce settings (Schafer et al. 1999). Although

[21] While Oestreicher-Singer and Sundararajan (2012a) reported that the network influence on a product category is associated with both increases and decreases in relative revenue of books depending on their popularity within the category, the authors did not conduct their analysis at the individual product level.

significant efforts have been made to these system design related research, there have been only a few studies investigating the economic impact (e.g., sales impact) of recommender systems. Specifically, Anderson (2006) and Brynjolfsson et al. (2006) reported that recommendations help consumers discover new products and thus increase sales diversity, whereas Mooney and Roy (2000) argued that recommendations only reinforce the position of already-popular products and thus instead reduce diversity. To explain the existence of these two opposite anecdotal views, Fleder and Hosanagar (2009) found that it is possible for individual-consumer level diversity to increase because recommendations can guide each individual consumer to new products, but aggregate market-level diversity to decrease because recommendations often guide similar users toward the same products. In addition to the sales diversity impact of recommender systems, Pathak et al. (2010) also showed that recommendations can directly increase the sales of recommended products, and also their prices (due to the reduction of consumer search costs and quality uncertainty and thus the increase in consumers' willingness to pay a higher price).

As recommendations form a visible product network, some researchers were particularly interested in the impact of recommender systems from the network's perspective. However, the limited attention given to product networks and the paucity of studies on it is surprising. The handful of studies on product networks analyze networks of World Wide Web sites (Katona and Sarvary 2008), blogs (Mayzlin and Yoganarasimhan 2012), news reports (Dellarocas et al. 2010), and videos (Goldenberg et al. 2012). However, these studies did not examine the impact on actual product demand. In the e-commerce contexts, there are some studies investigating the network impact on actual product demand, and thus are more relevant to our research. For instance, Carmi et al. (2010) identified the spread of exogenous demand shocks generated by book reviews featured on the Oprah Winfrey TV show and published in the New York Times through the co-purchase recommendation networks on Amazon.com. Carmi et al. (2011), in the spirit of the PageRank algorithm (Brin and Page 1998), quantified the value of a product to the firm by decomposing the revenue of each product into the intrinsic value portion (i.e., self-generated by the product) and the extrinsic value portion (i.e., driven by the recommendation links pointing from other products to the focal product). Moreover, Oestreicher-Singer and Sundararajan (2012a) associated the average influence of the network centrality on each book category with the inequality in the distribution of its revenue on Amazon.com. Using similar data, Oestreicher-Singer and Sundararajan (2012b) also showed that the explicit visibility of a co-purchase relationship could lead to a three-fold amplification of the influence that complementary products have on each others' demand levels.

Although these product network studies have started exploring the impact of network centrality, the important roles of network diversity and network stability have not yet attracted research interest. The only relevant studies are those in the field of social networks. Social network diversity refers to the diversity among a social member's network neighbors in aspects such as gender, age, education and work experience (Jehn et al. 1999). Most past studies focused on the association between network diversity and work performance. They found that individuals who have a more diverse network are more productive than their peers (Reagans and Zuckerman 2001), would receive higher performance ratings and compensation (Cross and Cummings 2004), are more likely to be recognized as top performers (Burt 2000), and obtain more economic opportunities (Eagle et al. 2010). In addition to work performance, several studies also examined other outcomes such as knowledge sharing (Cummings 2004), health condition (Barefoot et al. 2005), and online content propagation (Yoganarasimhan 2012). Additionally, the stability of an individual's social network is usually defined as the extent of overlap of network neighbors or connections over time (Cummings and Higgins 2006). Ghose et al. (2011) studied the stability of an individual's social network on individual behavior in the mobile Internet setting. They found that users with high network stability have a low intrinsic tendency to engage in content usage and generation on the mobile Internet. Moreover, Tucker (2011) examined the effect of instability in social networks on network externalities, and therefore on the rate of adoption of a video-calling system. She identified that the aggregate effect of network externalities on adoption is augmented by communication network instability, due to individuals' uncertainty of future communication networks.

Despite the research efforts in product networks, the invariable adoption of "unidirectional" view to examine product network effects (i.e., only focusing on the incoming network without consideration of the outgoing network) has probably led to incomplete analyses. Specifically, a fundamental insight of research on products is that the demand for different products can be interrelated (Seetharaman et al. 2005). The marketing literature on demand spillovers has extensively demonstrated the existence of this demand interdependency between a pair of substitutes (Anupindi et al. 1998), complements (Song and Chintagunta 2006), or even less related products (Singh et al. 2005). Particularly, past studies have highlighted the bidirectionality property of demand spillovers (Manchanda et al. 1999). Moreover, prior literature has identified the feature that spillovers are usually asymmetric between a product pair (Knott et al. 2009; Sethuraman and Srinivasan 2002) due to the asymmetry in aspects such as vulnerability (i.e., what is the extent of vulnerability in being hurt by price discounts of other products) (Sethuraman 1995), and product quality (Allenby and Rossi 1991). Therefore, in the product network context where products are connected with both incoming and outgoing networks, the property of product demand interdependency has therefore exposed the deficiencies of the unidirectional perspective.

Our research thus differs from related prior studies by examining the impact of product networks by focusing on both *incoming* and *outgoing* networks, and differentiating between *co-view* and *co-purchase* networks to study the *diversity effect* and *stability effect* in product networks for a more complete and rigorous investigation.

3.3 Hypotheses

Consumers exhibit heterogeneity in terms of their needs for products or categories (Allenby et al. 1998; Kamakura et al. 1996; Osborne 2011). Hence, in e-commerce contexts, consumers may search for different product categories (which can be substitutes or complements to the focal product) that suit their own needs. As such, if the category diversity of a focal product's incoming network increases, this incoming network will attract a larger group of consumers or visitors. As incoming network products provide visible connections to the focal product, the focal product will be accessible by this larger group of potential consumers from the incoming network, and eventually the exposure of the focal product will be increased. Consequently, the demand may increase due to this heightened exposure (Carmi et al. 2010). Past studies have reported that products, which are connected in the network, will have sales correlations. For instance, Oestreicher-Singer and Sundararajan (2012b) studied the network of books on Amazon and identified the incremental correlation in book sales attributable to the product network's visibility. This observation provides further support to our hypothesis. Therefore, we hypothesize that the category diversity in the incoming network has a positive relationship with the demand of the focal product.

However, the impact of the outgoing network category diversity on the demand of the focal product may differ substantially. When browsing the web page of the focal product, consumers will be exposed to the associated recommendations (i.e., outgoing network products) on the same page. Consequently, the outgoing network products may distract consumers' attention, regardless of whether the outgoing network products are substitutes or complements to the focal product, or even when they are irrelevant to consumers' search goals. Specifically, based on the theory of stimulus complexity (Berlyne 1960), web page complexity is defined as the level of diversity of information about the stimulus. This is documented to distract consumers (Deng and Poole 2010; Nadkarni and Gupta 2007). Thus, in line with the effect of web page complexity, we posit that the increase in the category diversity of the outgoing network is more likely to shift consumers' attention away from the focal product, and consequently lower the chance of buying the focal product. As such, different from the impact of the diversity of the incoming network, we hypothesize that the outgoing network diversity has a negative relationship with the demand of the focal product.

Hypothesis 1A (H1A): Diversity of the incoming product network has a positive relationship with the demand of the focal product.

Hypothesis 1B (H1B): Diversity of the outgoing product network has a negative relationship with the demand of the focal product.

In addition to the diversity of a focal product's network at a specific moment, the changes (i.e., instability) of the network connections may also introduce diversity in another manner, i.e., diversity over time. Specifically, an increase in the stability of the focal product's incoming network (i.e., less change in the incoming network) implies that the focal product will form connections with less diverse incoming network products over a certain time period. In other words, over a certain time period, an increase in the incoming network stability will attract a smaller group of consumers or visitors due to the potentially smaller scope of product variety, because consumers have heterogeneous product preferences (Allenby et al. 1998; Kamakura et al. 1996; Osborne 2011). Consequently, the focal product will experience a lower level of exposure to the potential consumers from its incoming links. Hence, we expect that the stability of incoming network has a negative relationship with the demand of the focal product.

As to the stability of the outgoing network, due to the nature of product recommendation systems, i.e., helping consumers find the right ideal product (Schafer et al. 2001), it would be likely for consumers to identify more desirable (compared to the focal product) substitute products in the focal product's outgoing network (i.e., recommendation list). As such, if the stability of a focal product's outgoing network increases, the focal product will more consistently point to certain recommended products over time. From the consumer's perspective, this suggests that those recommended products in the outgoing network may be widely and unanimously preferred by other consumers, in contrast to the case of unstable recommendations which suggests that consumers have quite different preferred substitutes. Thus, the stability may confer a higher degree of perceived product quality or fitness of the outgoing network substitute products to consumers. Consequently, consumers may follow others' preference or behaviors (Duan et al. 2009) to purchase these recommended substitute products in the outgoing network. The likelihood of purchasing the focal product will thus be reduced. This suggests a negative relationship between the outgoing network stability on the focal product's demand, assuming the outgoing links are for substitute products.

However, another important objective of recommendation systems is to suggest additional complementary products to increase cross-selling opportunities (Schafer et al. 2001). In other words, recommendation systems aim to propose complementary products to the focal product for consumers' co-purchase. As such, if the stability of a focal product's outgoing network increases (i.e., less updating involved in the recommendations), there would be less variety in terms of the options for consumers' co-purchase of the focal product and the outgoing network products. Consequently, the demand of the focal product may decrease or remain stable at best. In sum, aggregating across both cases of substitute and complementary products in the outgoing network, we posit that there is a negative demand impact of outgoing network stability. Therefore, we hypothesize a negative relationship between the outgoing network stability and the focal product's demand.

Hypothesis 2A (H2A): Stability of the incoming product network has a negative relationship with the demand of the focal product.

Hypothesis 2B (H2B): Stability of the outgoing product network has a negative relationship with the demand of the focal product.

Finally, we expect that co-view and co-purchase networks may have different impacts on the demand of a focal product. Specifically, it has been widely documented that consumers make purchase decisions by observing other consumers' preferences and behaviors (Yang and Allenby 2003). Co-view product recommendations contain information about other consumers' choice set and online product search or browsing behavior. However, searching or browsing a certain product does not always lead to the consumer's eventual purchase of the same product. In contrast, co-purchase product recommendations indicate other consumers' actual purchase behavior. Thus, from the consumers' perspective, co-purchase product recommendation information may be more salient, persuasive or influential than co-view product recommendation information. In addition, we note that retailers and recommendation system designers typically use co-purchase recommendations to influence consumers to buy additional products (which is more likely to generate additional profits). In contrast, co-view recommendations are usually used to help consumers find the right ideal product (which is less likely to generate incremental profits) (Schafer et al. 2001). Thus, co-purchase recommendations are often exploited more strategically (e.g., by displaying more co-purchase recommendations in a more salient location of a web page) as compared to co-view recommendations. For instance on Amazon, co-purchase recommendations are displayed right below the image of the focal product whereas co-view recommendations are displayed at the bottom of a web page which capture less consumer attention. Therefore, we expect that the demand impact of the co-purchase network would be stronger than that of the co-view network.

Hypothesis 3 (H3): Co-purchase product network has a stronger impact on the demand of the focal product than co-view product network.

3.4 Empirical Method and Analysis

3.4.1 Data Description

Our dataset for empirical validation of the proposed hypotheses is a rich set of data from Tmall.com (formerly Taobao Mall) (www.tmall.com), a Chinese-language business-to-consumer (B2C) e-commerce platform under the Alibaba Group. Tmall was launched in April 2008, but was separated from Taobao's consumer-to-consumer (C2C) marketplace and became an independent business in June 2011. Tmall consists of various online stores, currently featuring more than 70,000 multinational and Chinese brands from more than 50,000 merchants (Alibaba 2012), and is the most visited B2C online retail website in China (Alexa 2012).

Our dataset includes information of all products in an online flagship store selling Nikon digital cameras and various associated components (e.g., lens and battery). Consistent with the retailer's categorization, all products are grouped into six categories, namely (1) accessory, (2) battery, (3) compact camera, (4) flash, (5) lens, and (6) single lens reflex (SLR) camera[22]. Each product is featured on its own designated web page, including all relevant information (e.g., price, inventory, and consumer reviews) (see Figure 2-3 for an illustration) and its detailed transaction records (see Figure 2-4 for an illustration). Moreover, on each product web page, Tmall also utilizes two different product recommendation systems. Recommendations can be listed under sections with the headings "consumers who viewed this item also viewed" (i.e., co-view, see Figure 2-5 for an illustration), and "consumers who bought this item also bought" (i.e., co-purchase, see Figure 2-6 for an

[22] Compact cameras and SLR cameras are categorized separately because compact cameras are highly standardized whereas SLR cameras usually need extra lenses and flashes for different customizations.

illustration). The algorithm Tmall uses to provide recommendations is based on the collaborative filtering approach. Specifically, co-view (co-purchase) recommendation systems first identify the group of consumers who have viewed (purchased) a focal product. Then co-view (co-purchase) recommendation systems further identify what other products these consumers also viewed (purchased) subsequently and then provide as the co-view (co-purchase) recommendations on the focal product's page. Noteworthy, Tmall restricts that recommendation systems in a store only recommend products from the same store. Thus, recommendation links jointly form two networks (i.e., co-view network and co-purchase network) of all the products in the store[23].

Figure 2-3 - Tmall Product Web Page

[23] Tmall co-view and co-purchase recommendation lists have non-overlapping constituent items, indicating that products in the co-view network are completely different from those in the co-purchase network. This also motives us to investigate the relative impact of these two networks.

Figure 2-4 - Tmall Product Transaction Record

(buyer) 买家	(product name) 宝贝名称	(product price) 拍下价格	(quantity) 购买数量	(transaction date & time) 成交时间
e**o	Nikon/尼康 COOLPIX S6300 轻便型 数码相机 10倍变焦 防抖 正品 颜色分类:桃红色;套餐:官方标配	1529	1	2012-09-21 20:23:30
r**6	Nikon/尼康 COOLPIX S6300 轻便型 数码相机 10倍变焦 防抖 正品 颜色分类:黑色;套餐:官方标配	1529	1	2012-09-13 17:12:49
爱**理	Nikon/尼康 COOLPIX S6300 轻便型 数码相机 10倍变焦 防抖 正品 颜色分类:金色;套餐:官方标配	1529	1	2012-09-11 21:05:52

上一页 1 2 ... 下一页
(previous) (next)

Figure 2-5 - Tmall Co-View Product Recommendation

Figure 2-6 - Tmall Co-Purchase Product Recommendation

In order to test our hypotheses, we need to capture all products and their corresponding network structure. However, from a researcher's practicality point of view, it is not quite feasible to observe all real-time changes of product information, especially those of the product network structure. As such, we only collect data on product information and product network structure on a daily basis (12:00 a.m.). Consequently, our dataset consists of three parts: (1) daily snapshots of product-related information (e.g., price and consumer reviews), (2) daily snapshots of product network structure, and (3) detailed individual product transaction records with sales quantity and price. The dataset has 257 products in total across 184 days from May to December 2012. Figures 2-7 and 2-8 present the co-view and co-purchase networks of the midpoint (September 1, 2012) of the sample period, and Tables 2-1 and 2-2 summarize the corresponding network metrics.

Figure 2-7 - Co-View Network Structure

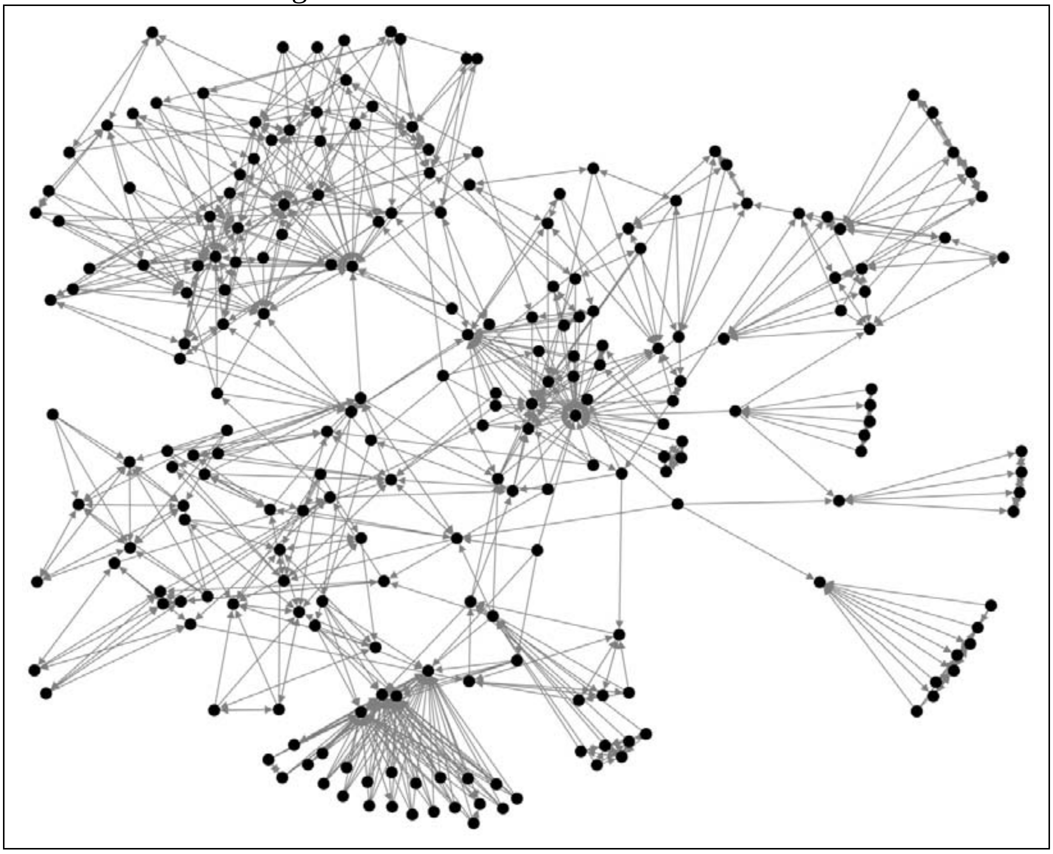


Table 2-1 - Co-View Network Metrics

Metric	Mean	Std. Dev.	Min	Max
In-degree centrality	4.000	5.597	0.000	37.000
Out-degree centrality	4.000	0.000	4.000	4.000
Betweenness centrality	565.644	1,549.084	0.000	17,502.777
Closeness centrality	0.009	0.029	0.001	0.143
Eigenvector centrality	0.004	0.012	0.000	0.074
PageRank	1.000	0.634	0.587	5.365
Clustering coefficient	0.452	0.300	0.000	1.000

Note: Number of nodes = 225; Number of edges = 900. Graph density = 0.018.

Figure 2-8 - Co-Purchase Network Structure

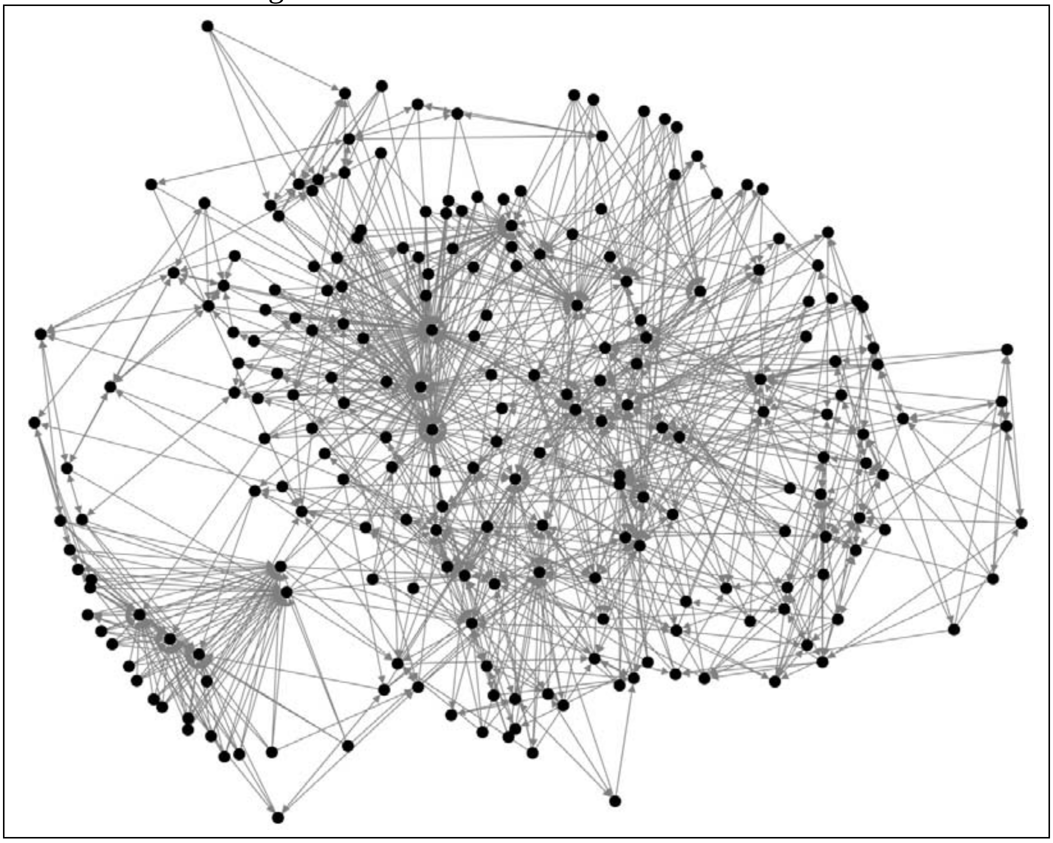


Table 2-2 - Co-Purchase Network Metrics

Metric	Mean	Std. Dev.	Min	Max
In-degree centrality	4.942	9.298	0.000	72.000
Out-degree centrality	4.942	0.269	3.000	5.000
Betweenness centrality	386.347	1,098.046	0.000	9,634.566
Closeness centrality	0.002	0.000	0.001	0.002
Eigenvector centrality	0.004	0.004	0.000	0.034
PageRank	1.000	0.879	0.509	7.415
Clustering coefficient	0.174	0.156	0.000	0.650

Note: Number of nodes = 225; Number of edges = 1,125. Graph density = 0.022.

3.4.2 Empirical Model

Based on our dataset, we operationalize all relevant variables at the product-day level. Let subscript *i* denote each individual product in the Nikon store, and subscript *t* denote each time period (daily). The *dependent variable* in this study is product *i*'s daily sales quantity, $QUAN_{it}$, measured as the total quantity of product *i* sold on day *t*. Next, our *independent variables* include network diversity and network stability. Network diversity is measured as the unique number of product categories in product *i*'s network on day *t*. Network stability is measured as the percentage of overlap of network connections in product *i*'s network across two days (snapshots of *t* and *t*+1), i.e., the number of network connections which still existed on day t+1 over the number of original network connections on day t. Our research differs from prior product network studies in two aspects. First, we simultaneously examine the demand impact of both a focal product's incoming and outgoing network. Second, we incorporate and differentiate the demand impact of two different recommendation mechanisms (i.e., co-view and co-purchase). Therefore, our final set of independent variables include incoming co-view network diversity (ID_CV_{it}), incoming co-purchase network diversity (ID_CP_{it}), outgoing co-view network diversity (OD_CV_{it}), outgoing co-purchase network diversity (OD_CP_{it}), incoming co-view network stability (IS_CV_{it}), incoming co-purchase network stability (IS_CP_{it}), outgoing co-view network stability (OS_CV_{it}), and outgoing co-purchase network stability (OS_CP_{it}). Finally, our *control variables* are gathered from those identified in our literature review and from the available information in our dataset. Specifically, we include control variables at the individual product, product network, product category, and time unit levels: (1) product list price (inclusive of discounts if any) (LP_{it}), (2) volume of product reviews (PV_{it}), (3) rating of product reviews (PR_{it}), (4) past monthly sales quantity[24] (PS_{it}), (5) inventory[25] (IN_{it}), (6) number of web page bookmarks[26] (BM_{it}), (7) in-degree and out-degree co-view and co-purchase network centrality[27] (IC_CV_{it}, IC_CP_{it}, OC_CV_{it}, OC_CP_{it}), (8) average list price of incoming and outgoing co-view and co-purchase network products[28] (ILP_CV_{it}, ILP_CP_{it}, OLP_CV_{it}, OLP_CP_{it}), (9) average review volume of incoming and outgoing co-view and co-purchase network products[29] (IPV_CV_{it}, IPV_CP_{it}, OPV_CV_{it}, OPV_CP_{it}), (10) average review rating of incoming and outgoing co-view and co-purchase network products[30] (IPR_CV_{it}, IPR_CP_{it}, OPR_CV_{it}, OPR_CP_{it}), (11) average sales quantity of

[24] This indicates the sales quantity of product *i* during the past month prior to day *t*.
[25] This indicates the available quantity of product *i* for sale on day *t*.
[26] This indicates the total number of product *i*'s web page bookmarked by consumers on day *t*.
[27] This measures the number of products in *i*'s product network on day *t*.
[28] The average price across all products in *i*'s product network on day *t*.
[29] The average volume of reviews across all products in *i*'s product network on day *t*.
[30] The average rating of reviews across all products in *i*'s product network on day *t*.

incoming and outgoing co-view and co-purchase network products[31] (IQ_CV_{it}, IQ_CP_{it}, OQ_CV_{it}, OQ_CP_{it}), (12) product category dummies (C_i), and (13) time dummies at the daily level (T_t).

Noteworthy, we only observe the daily snapshots of product information and network structure. Thus, in order to derive more accurate measurement of variables, we measure our product-related variables (e.g., price) and network variables (e.g., network diversity) in time period t using the average values in the current (t) and the next (t+1) time periods[32].

We model the influence of network diversity and network stability on product demand. We specify the dependent variable in logarithm[33]. The panel-level regression model is specified in Equation (1):

$$\begin{aligned}\ln\left(QUAN_{it}\right) = {} & \beta_1 ID\_CV_{it} + \beta_2 ID\_CP_{it} + \beta_3 OD\_CV_{it} + \beta_4 OD\_CP_{it} \\ & + \beta_5 IS\_CV_{it} + \beta_6 IS\_CP_{it} + \beta_7 OS\_CV_{it} + \beta_8 OS\_CP_{it} \\ & + \beta_9 LP_{it} + \beta_{10} PV_{it} + \beta_{11} PR_{it} + \beta_{12} PS_{it} + \beta_{13} IN_{it} + \beta_{14} BM_{it} \\ & + \beta_{15} IC\_CV_{it} + \beta_{16} IC\_CP_{it} + \beta_{17} OC\_CV_{it} + \beta_{18} OC\_CP_{it} \\ & + \beta_{19} ILP\_CV_{it} + \beta_{20} ILP\_CP_{it} + \beta_{21} OLP\_CV_{it} + \beta_{22} OLP\_CP_{it} \\ & + \beta_{23} IPV\_CV_{it} + \beta_{24} IPV\_CP_{it} + \beta_{25} OPV\_CV_{it} + \beta_{26} OPV\_CP_{it} \\ & + \beta_{27} IPR\_CV_{it} + \beta_{28} IPR\_CP_{it} + \beta_{29} OPR\_CV_{it} + \beta_{30} OPR\_CP_{it} \\ & + \beta_{31} IQ\_CV_{it} + \beta_{32} IQ\_CP_{it} + \beta_{33} OQ\_CV_{it} + \beta_{34} OQ\_CP_{it} \\ & + C_i\gamma + T_t\mu + \alpha_i + \varepsilon_{it}\end{aligned} \tag{1}$$

where α_i captures unobserved product specific effects. βs, γ, and μ are the model coefficients, and ε_{it} indicates the residual random error term. Table 2-3 reports the descriptive statistics and the correlation matrix is reported in Table 2-4.

[31] The average sales quantity across all products in i's product network on day t.
[32] Essentially, this captures the average value of a variable on day t, as we capture the snapshots of product information and network structure at 12:00 a.m. of t (the beginning of day t) and 12:00 a.m. of t+1 (the end of day t), and then derive the average of these two measures. We also report robustness checks in Section 3.5.2 on the sensitivity of this operationalization by using the actual day t values and find consistent results.
[33] Empirical analyses often fit better with economic variables specified in logarithm (Wooldridge 2006, pp. 197-200). As appropriate, we add one to the variable to avoid logarithms of zeroes.

Table 2-3 - Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
QUAN (Sales quantity)	2.283	6.791	0.000	680.000
ID_CV (Incoming network diversity, co-view)	0.888	0.584	0.000	6.000
ID_CP (Incoming network diversity, co-purchase)	1.180	1.250	0.000	6.000
OD_CV (Outgoing network diversity, co-view)	1.090	0.446	0.000	4.000
OD_CP (Outgoing network diversity, co-purchase)	1.519	1.133	0.000	5.000
IS_CV (Incoming network stability, co-view)	0.787	0.307	0.000	1.000
IS_CP (Incoming network stability, co-purchase)	0.711	0.380	0.000	1.000
OS_CV (Outgoing network stability, co-view)	0.757	0.321	0.000	1.000
OS_CP (Outgoing network stability, co-purchase)	0.727	0.378	0.000	1.000
LP (Product list price)	4,245.943	9,811.471	1.000	69,100.000
PV (Product review volume)	3.875	30.804	0.000	634.000
PR (Product review rating)	1.537	2.276	0.000	5.000
PS (Product past monthly sales quantity)	1.754	12.378	0.000	320.000
IN (Product inventory)	84.742	357.640	1.000	9,765.000
BM (Number of product web page bookmarks)	59.267	598.451	0.000	9,783.000
IC_CV (Network in-degree centrality, co-view)	3.712	4.830	0.000	57.500
IC_CP (Network in-degree centrality, co-purchase)	3.570	7.427	0.000	87.000
OC_CV (Network out-degree centrality, co-view)	3.707	0.931	0.000	4.000
OC_CP (Network out-degree centrality, co-purchase)	3.559	2.032	0.000	5.000
ILP_CV (Incoming network product list price, co-view)	4,147.817	8,256.490	0.000	69,100.000
ILP_CP (Incoming network product list price, co-purchase)	2,885.459	6,062.424	0.000	69,100.000
OLP_CV (Outgoing network product list price, co-view)	3,167.484	4,880.539	0.000	50,550.000
OLP_CP (Outgoing network product list price, co-purchase)	2,578.452	3,199.054	0.000	40,553.750
IPV_CV (Incoming network product review volume, co-view)	2.333	9.811	0.000	327.750
IPV_CP (Incoming network product review volume, co-purchase)	3.506	16.498	0.000	483.500
OPV_CV (Outgoing network product review volume, co-view)	21.294	41.487	0.000	243.250
OPV_CP (Outgoing network product review volume, co-purchase)	25.850	40.051	0.000	295.000
IPR_CV (Incoming network product review rating, co-view)	1.189	1.473	0.000	5.000
IPR_CP (Incoming network product review rating, co-purchase)	1.349	1.585	0.000	5.000
OPR_CV (Outgoing network product review rating, co-view)	2.406	1.590	0.000	5.000
OPR_CP (Outgoing network product review rating, co-purchase)	2.267	1.694	0.000	5.000
IQ_CV (Incoming network product sales quantity, co-view)	0.052	0.652	0.000	57.500
IQ_CP (Incoming network product sales quantity, co-purchase)	0.079	0.641	0.000	43.167
OQ_CV (Outgoing network product sales quantity, co-view)	0.269	0.658	0.000	14.125
OQ_CP (Outgoing network product sales quantity, co-purchase)	0.335	1.210	0.000	70.125

Note: Number of observations = 41,379; Number of products = 257; Number of days = 184. All variables are at the product-day level.

Table 2-4 - Correlation Matrix

	Variable	1	2	3	4	5	6	7	8	9
1	*QUAN* (Sales quantity)	-								
2	*ID_CV* (Incoming network diversity, co-view)	-0.027	-							
3	*ID_CP* (Incoming network diversity, co-purchase)	0.000	0.478	-						
4	*OD_CV* (Outgoing network diversity, co-view)	-0.068	0.212	0.030	-					
5	*OD_CP* (Outgoing network diversity, co-purchase)	-0.024	0.201	0.472	0.294	-				
6	*IS_CV* (Incoming network stability, co-view)	0.020	-0.184	-0.081	0.060	0.066	-			
7	*IS_CP* (Incoming network stability, co-purchase)	0.011	-0.143	-0.247	0.035	-0.025	0.146	-		
8	*OS_CV* (Outgoing network stability, co-view)	0.013	0.030	0.032	-0.242	-0.009	0.193	0.100	-	
9	*OS_CP* (Outgoing network stability, co-purchase)	-0.000	-0.014	-0.097	-0.226	-0.303	0.026	0.112	0.434	-

Note: Only major variables are reported. The correlations of these variables with other variables are generally small.

3.4.3 Implicit Demand Correlation

A major concern in this study is that two products in the store might have implicit demand correlation (i.e., substitution or complementarity) regardless of visible network connections being present. As such, the demand of a focal product would have been driven jointly by two factors: (1) implicit demand correlation, and (2) explicit network structure. Hence, in order to identify the impact of network structure on product demand, we have to take the implicit demand correlation into account. Oestreicher-Singer and Sundararajan (2012b) provide strategies to identify implicit demand correlation (albeit in terms of complementarity only). Specifically, for each focal product, their study identified a set of products which have implicit demand correlation with the focal product, and then controlled for the demand of this set of products as their main identification strategy. Unfortunately, all strategies used in Oestreicher-Singer and Sundararajan (2012b) were based on a strong assumption, i.e., complementarity is the only factor that drives the implicit demand correlation.

To better address this issue, we propose a different strategy which seeks to capture implicit demand correlation in terms of both product substitution and complementarity. Our strategy goes beyond the assumption used in Oestreicher-Singer and Sundararajan (2012b) to allow a pair of products to be substitutable between, complementary or independent to each other. Specifically, we use the cross-product category price elasticity to determine the extent of substitute or complementary relationships (Leeflang and Parreño-Selva 2012; Manchanda et al. 1999; Niraj et al. 2008). Ideally, substitution and complementarity effects should be estimated at the product level. However, past research (Berry 1994; Song and Chintagunta 2006) has outlined that estimating at the product level would be problematic due to the issue of “parameters explosion” (i.e., too many parameters to estimate due to the huge number of products when we specify a full model to account for product demand interdependency). We thus follow prior approaches to obtain cross-category level price elasticities to determine the extent of substitute or complementary relationships between each pair of products. Specifically, we first group all products into six different product categories as discussed previously based on the retailer’s classification according to the product functionalities, and consider all products within a product category as substitutes for one another. This approach is in line with the definition of substitutes proposed by Henderson and Quandt (1958, p. 29): "…two commodities are substitutes if both can satisfy the same need". This approach has been adopted in past research (Mulhern and Leone 1991; Walters 1991). After the grouping, we aggregate all product characteristic variables (e.g., price) from the product level to the category level[34].

Let subscript m = 1, 2, 3, 4, 5, 6 denote each product category, and subscript t denote each day. The dependent variable is category sales quantity[35], $QUAN_{mt}$, measured as the total quantity of category m sold on day t. The explanatory variables are all the major category-level characteristic variables aggregated from the product level, including: (1) list price (LP_{mt}); (2) volume of product reviews (PV_{mt}); (3) rating of product reviews (PR_{mt}); (4) past monthly sales quantity (PS_{mt}); (5) inventory (IN_{mt}); (6) number of web page bookmarks (BM_{mt}). The descriptive statistics of all these category-level variables are reported in Table 2-5.

[34] Each category-level characteristic variable is the weighted sum of the product-level characteristic variable. Weight for product i on day t is the sales quantity of product i on day t over the total sales quantity of all products in the same category on day t.

[35] We use sales quantity to maintain consistency with the product demand specification in Equation (1).

Table 2-5 - Descriptive Statistics (Product Category)

Variable	Mean	Std. Dev.	Min	Max
QUAN1 (Sales quantity: accessory)	33.863	112.720	0.000	2,200.000
QUAN2 (Sales quantity: battery)	3.349	5.898	0.000	36.000
QUAN3 (Sales quantity: compact camera)	2.313	4.967	0.000	50.000
QUAN4 (Sales quantity: flash)	0.056	0.275	0.000	2.000
QUAN5 (Sales quantity: lens)	1.019	2.105	0.000	26.000
QUAN6 (Sales quantity: SLR camera)	4.514	6.857	0.000	73.000
LP1 (List price: accessory)	302.980	355.428	1.000	2,550.000
LP2 (List price: battery)	282.417	139.319	48.000	978.000
LP3 (List price: compact camera)	1,319.572	389.374	376.000	2,770.000
LP4 (List price: flash)	2,572.396	386.164	848.000	3,578.000
LP5 (List price: lens)	6,551.921	3,943.045	499.000	13,024.500
LP6 (List price: SLR camera)	6,396.445	4,291.084	2,546.000	15,087.420
PV1 (Review volume: accessory)	18.502	19.071	0.000	102.500
PV2 (Review volume: battery)	79.662	157.054	0.000	639.500
PV3 (Review volume: compact camera)	30.967	54.525	0.000	275.500
PV4 (Review volume: flash)	1.000	0.989	0.000	6.000
PV5 (Review volume: lens)	7.964	18.202	0.000	147.000
PV6 (Review volume: SLR camera)	179.917	179.716	0.000	634.000
PR1 (Review rating: accessory)	3.463	1.770	0.000	5.000
PR2 (Review rating: battery)	3.800	1.145	0.000	5.000
PR3 (Review rating: compact camera)	3.323	1.380	0.000	5.000
PR4 (Review rating: flash)	1.991	1.756	0.000	5.000
PR5 (Review rating: lens)	2.691	1.725	0.000	5.000
PR6 (Review rating: SLR camera)	3.710	1.480	0.000	5.000
PS1 (Past monthly sales quantity: accessory)	19.956	22.639	0.000	100.000
PS2 (Past monthly sales quantity: battery)	32.896	60.624	0.000	311.933
PS3 (Past monthly sales quantity: compact camera)	12.070	20.042	0.000	92.833
PS4 (Past monthly sales quantity: flash)	0.426	0.856	0.000	7.500
PS5 (Past monthly sales quantity: lens)	3.206	5.577	0.000	35.000
PS6 (Past monthly sales quantity: SLR camera)	63.832	72.856	0.000	314.000
IN1 (Inventory: accessory)	2,867.024	4,661.136	8.500	38,857.000
IN2 (Inventory: battery)	259.014	269.346	1.500	1,290.500
IN3 (Inventory: compact camera)	523.199	504.865	7.000	3,511.500
IN4 (Inventory: flash)	209.769	306.949	4.500	2,994.500
IN5 (Inventory: lens)	66.008	97.170	2.500	612.500
IN6 (Inventory: SLR camera)	490.337	583.760	1.000	7,053.500
BM1 (Number of web page bookmarks: accessory)	23.005	32.749	0.000	185.213
BM2 (Number of web page bookmarks: battery)	152.111	286.563	0.000	1,248.500
BM3 (Number of web page bookmarks: compact camera)	118.589	158.925	1.000	737.500
BM4 (Number of web page bookmarks: flash)	15.569	8.891	4.000	76.500
BM5 (Number of web page bookmarks: lens)	66.406	93.617	1.000	514.000
BM6 (Number of web page bookmarks: SLR camera)	4,105.217	3,733.216	0.000	12,114.000

Note: Number of observations = 533.

In order to obtain the cross-category price elasticities, we specify a full model including the above six variables for each product category *m*, where we estimate a panel-level linear model shown in Equation (2):

$$\ln(QUAN_{mt}) = \sum_{n=1}^{n=6} \alpha_{mn} LP_{nt} + \sum_{n=1}^{n=6} \beta_{mn} PV_{nt} + \sum_{n=1}^{n=6} \chi_{mn} PR_{nt} + \sum_{n=1}^{n=6} \delta_{mn} PS_{nt} + \sum_{n=1}^{n=6} \varepsilon_{mn} IN_{nt} + \sum_{n=1}^{n=6} \phi_{mn} BM_{nt} + T_t \mu + \omega_m + \xi_{mt} \quad (2)$$

where T_t denotes time dummies at the daily level; ω_m captures unobserved category specific effects; $\alpha_{mn}, \beta_{mn}, \chi_{mn}, \delta_{mn}, \varepsilon_{mn}, \phi_{mn}, \mu$ are the model coefficients, and ξ_{mt} denotes the residual random error term.

The estimated model coefficients based on data from each category of multiple[36] Nikon stores are summarized in Table 2-6. We compute the category-level price elasticities and summarize them in Table 2-7. As discussed above, we use the estimated cross-category price elasticities to determine the potential substitute or complementary relationship between each pair of products. Specifically, let E_{ij} denote the price elasticity of product *i*'s demand with respect to the price of product *j*. Suppose product *i* is in category *m* and product *j* is in category *n*, then we use the estimated category-level elasticity (i.e., entry (*m*, *n*) in Table 2-7, the price elasticity of category *m*'s demand with respect to the price of category *n*) as a proxy for E_{ij}. However, if both products *i* and *j* are in the same category (i.e., $m = n$), we follow the above discussion to consider that products *i* and *j* are substitutes, and use the absolute value of own-category price elasticity[37] (i.e., E_{ij} = |entry (*m*, *n*)|) to indicate the substitution effect of product *j* on product *i*. As such, we can identify the underlying substitute or complementary relationship between each pair of products based on this strategy[38]. Noteworthy, our approach allows for the implicit demand correlation in terms of both substitution and complementarity between the focal product and all the other products in the same store, regardless of a visible network connection being present. In other words, for each focal product, we treat all the other products in the same store

---

[36] Using data from multiple stores provides us with a larger model estimation sample for more precise estimates.

[37] Theoretically, own-price elasticity of demand is always negative. However, the cross-price elasticity between two substitutes should be positive. Thus, we use the absolute value of own-price elasticity.

[38] We report robustness checks in Section 3.5.2 on the sensitivity of this operationalization by assigning zeros to all the insignificant estimated elasticities and find consistent results with those of the main findings reported.

as potential substitutes or complements. We consider that product j is a substitute (complement) to focal product i if $E_{ij} > 0$ ($E_{ij} < 0$).

Table 2-6 - Model Estimation Results (Product Category)

Variable	(1) Accessory		(2) Battery		(3) Compact camera		(4) Flash		(5) Lens		(6) SLR camera	
	Fixed effects	Random effects	Fixed effects	Random effects	Fixed effects	Random effects	Fixed effects	Random effects	Fixed effects	Random effects	Fixed effects	Random effects
LP1	-0.000	-0.000***	-0.000	-0.000**	-0.000*	-0.000***	0.000**	0.000***	0.000**	0.000***	0.000*	0.000
(List price: accessory)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LP2	0.001**	0.001***	-0.001	-0.001*	0.000	0.000*	0.000	0.000	0.001***	0.001***	0.000	0.000*
(List price: battery)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LP3	-0.000	-0.000	-0.000	-0.000***	-0.000**	-0.000***	0.000	0.000**	-0.000	-0.000	-0.000	-0.000
(List price: compact camera)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LP4	-0.000	-0.000*	-0.000	-0.000**	-0.000***	-0.000***	-0.000*	-0.000***	0.000	-0.000	0.000	-0.000
(List price: flash)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LP5	0.000**	0.000***	-0.000	-0.000	0.000**	0.000**	-0.000	-0.000	-0.000***	-0.000***	-0.000	0.000
(List price: lens)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LP6	0.000	0.000	0.000*	0.000	0.000	-0.000	-0.000	-0.000	0.000	0.000	-0.000**	-0.000***
(List price: SLR camera)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Control variables	-included-		-included-		-included-		-included-		-included-		-included-	
Number of observations	533		533		533		533		533		533	
Hausman test	$\chi^2 = 17.78$, $p = 1.00$		$\chi^2 = 30.28$, $p = 1.00$		$\chi^2 = 7.59$, $p = 1.00$		$\chi^2 = 74.26$, $p = 1.00$		$\chi^2 = 2.13$, $p = 1.00$		$\chi^2 = 18.30$, $p = 1.00$	
R^2	0.7670	0.9123	0.5415	0.8889	0.7822	0.8674	0.3349	0.7551	0.8160	0.8663	0.8610	0.9035

Note: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01. For brevity, only price estimates are reported.

Table 2-7 - Price Elasticity Matrix

Category *n* \ Category *m*	(1) Accessory	(2) Battery	(3) Compact camera	(4) Flash	(5) Lens	(6) SLR camera
(1) Accessory	-0.127***	-0.089**	-0.055***	0.043***	0.071***	0.006
	(0.047)	(0.041)	(0.008)	(0.008)	(0.016)	(0.025)
(2) Battery	0.245***	-0.163*	0.032*	0.005	0.151***	0.121*
	(0.076)	(0.094)	(0.018)	(0.012)	(0.012)	(0.073)
(3) Compact camera	-0.209	-0.183***	-0.469***	0.050**	-0.043	-0.034
	(0.318)	(0.070)	(0.135)	(0.023)	(0.030)	(0.024)
(4) Flash	-0.796*	-0.323**	-0.338***	-0.703***	-0.015	-0.144
	(0.458)	(0.153)	(0.092)	(0.077)	(0.148)	(0.125)
(5) Lens	0.183***	-0.012	0.043**	-0.007	-0.832***	0.013
	(0.024)	(0.051)	(0.021)	(0.007)	(0.060)	(0.079)
(6) SLR camera	0.293	0.240	-0.080	-0.022	0.094	-0.881***
	(0.325)	(0.173)	(0.145)	(0.021)	(0.155)	(0.267)

Note: Entry (m, n) represents the elasticity of category m with respect to the price of category n. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Let s = 1, 2, 3,…, S denote a substitute product for focal product i. As different substitutes may have different extent of substitution effects on product i's demand, and price elasticity E_{is} represents the level of substitution between products i and s, we weigh each substitute s's sales quantity on day t ($QUAN_{st}$) according to E_{is}, and then sum these weighted demands[39] to be the "influence" of substitution effect on product i on day t, denoted as SE_{it}, shown in Equation (3) below. Likewise, let c = 1, 2, 3,…, C denote a complementary product for focal product i. As different complements may have different extent of complementarity effects on product i's demand, and price elasticity E_{ic} represents the level of complementarity between products i and c, we weigh each complement c's sales quantity on day t ($QUAN_{ct}$) according to E_{ic}, and then sum these weighted demands to be the "influence" of complementarity effect on product i on day t, denoted as CE_{it}, shown in Equation (4) below.

$$SE_{it} = \sum_{s=1}^{s=S} E_{is} * QUAN_{st}, \quad \textit{for all s with } E_{is} > 0 \tag{3}$$

$$CE_{it} = \sum_{c=1}^{c=C} E_{ic} * QUAN_{ct}, \quad \textit{for all c with } E_{ic} < 0 \tag{4}$$

Essentially, the rationale of our strategy is to include these two variables SE_{it} and CE_{it}[40] as control variables in Equation (1) to account for the substitution and complementarity effects of other products that have intrinsic

[39] Our approach of weighing the demand of a product (can be a substitute or a complement) according to the cross-price elasticity (the extent of substitution or complementarity) is similar to the approach used in Oestreicher-Singer and Sundararajan (2012b) which weighed the demand of a product (complement only) according to its probability of being linked to (the extent of complementarity to) the focal product. However, using price elasticities can allow us to identify both substitution and complementary relationships (based on signs of cross-price elasticities) whereas using the probability of link formation has to assume the absence of substitution relationship. Nevertheless, we report robustness checks in Section 3.5.2 on the sensitivity of this operationalization by using the summation of actual demands without price elasticities as weights and find consistent results with those of the main findings reported.

[40] The correlations of *SE* and *CE* with other sales quantity related variables associated with the network links (i.e., *IQ_CV*, *IQ_CP*, *OQ_CV*, *OQ_CP*) are generally below 0.1, suggesting that both the *SE* and *CE* variables capture implicit demand factors beyond those reflected in the recommended product links.

demand correlations through product usage or exogenous shocks, or may appear as visible network links, or are currently invisible but potentially future network links (Carmi et al. 2010; Oestreicher-Singer and Sundararajan 2012b) which all may implicitly drive the focal product's demand.

3.5 Estimation and Results

3.5.1 Main Results

We first estimate a random effects (RE) model of product sales quantity on all the control variables. As reported in Table 2-8, Column (1), product own attributes (e.g., review volume, review rating, past monthly sales and inventory) have the expected relationships with product sales quantity. Moreover, various network-level control variables (e.g., degree centrality, review rating and sales quantity of the connected products) have significant relations with product sales quantity as well. Particularly, the substitution effect (*SE*) and the complementarity effect (*CE*) coefficients are statistically significant and with the expected signs, implying that higher substitution effect decreases the demand of a product, whereas higher complementarity effect instead increases it. These significant effects of *SE* and *CE* attest to the effectiveness of our approach in accounting for the implicit demand correlations. These results overall suggest that our set of control variables have good explanatory power.

Table 2-8 - Model Estimation Results

Variable	(1) RE: Control	(2) RE: Incoming	(3) RE: Outgoing	(4) RE: Co-view	(5) RE: Co-purchase	(6) RE: Full	(7) FE: Full
ID_CV		-0.004		-0.005*		0.001	0.005
(Incoming network diversity, co-view)		(0.003)		(0.003)		(0.003)	(0.003)
ID_CP		0.009***			0.010***	0.012***	0.002
(Incoming network diversity, co-purchase)		(0.001)			(0.002)	(0.002)	(0.002)
OD_CV			-0.003	-0.001		-0.007	-0.002
(Outgoing network diversity, co-view)			(0.004)	(0.004)		(0.004)	(0.004)
OD_CP			-0.005***		-0.007***	-0.007***	-0.002
(Outgoing network diversity, co-purchase)			(0.002)		(0.002)	(0.002)	(0.002)
IS_CV		0.007**		0.006*		0.006	0.003
(Incoming network stability, co-view)		(0.003)		(0.003)		(0.003)	(0.003)
IS_CP		-0.002			0.000	-0.002	-0.004
(Incoming network stability, co-purchase)		(0.003)			(0.003)	(0.003)	(0.003)
OS_CV			0.007*	0.000		0.005	0.007*
(Outgoing network stability, co-view)			(0.004)	(0.003)		(0.004)	(0.004)
OS_CP			-0.017***		-0.015***	-0.017***	-0.012***
(Outgoing network stability, co-purchase)			(0.003)		(0.003)	(0.003)	(0.003)
LP	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000**
(Product list price)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
PV	0.002***	0.002***	0.001***	0.001***	0.002***	0.002***	-0.000
(Product review volume)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
PR	0.004***	0.004***	0.003***	0.004***	0.003***	0.003***	-0.001
(Product review rating)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
PS	0.012***	0.012***	0.012***	0.012***	0.012***	0.012***	0.008***
(Product past monthly sales quantity)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
IN	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
(Product inventory)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BM	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
(Number of product web page bookmarks)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 2-8 - Model Estimation Results (Continued)

Variable	(1) RE: Control	(2) RE: Incoming	(3) RE: Outgoing	(4) RE: Co-view	(5) RE: Co-purchase	(6) RE: Full	(7) FE: Full
IC_CV	-0.000	-0.001***		-0.001***		-0.001*	-0.000
(Network in-degree centrality, co-view)	(0.000)	(0.000)		(0.000)		(0.000)	(0.000)
IC_CP	-0.002***	-0.002***			-0.002***	-0.003***	-0.002***
(Network in-degree centrality, co-purchase)	(0.000)	(0.000)			(0.000)	(0.000)	(0.000)
OC_CV	-0.011***		-0.010***	-0.008***		-0.009***	0.001
(Network out-degree centrality, co-view)	(0.001)		(0.002)	(0.002)		(0.002)	(0.002)
OC_CP	0.001		0.001		-0.000	0.001	0.001
(Network out-degree centrality, co-purchase)	(0.001)		(0.001)		(0.001)	(0.001)	(0.001)
ILP_CV	-0.000	0.000		0.000		-0.000	-0.000
(Incoming network product list price, co-view)	(0.000)	(0.000)		(0.000)		(0.000)	(0.000)
ILP_CP	0.000	-0.000			-0.000	-0.000	0.000
(Incoming network product list price, co-purchase)	(0.000)	(0.000)			(0.000)	(0.000)	(0.000)
OLP_CV	0.000		0.000	-0.000		0.000	-0.000
(Outgoing network product list price, co-view)	(0.000)		(0.000)	(0.000)		(0.000)	(0.000)
OLP_CP	0.000		0.000		0.000	0.000	-0.000
(Outgoing network product list price, co-purchase)	(0.000)		(0.000)		(0.000)	(0.000)	(0.000)
IPV_CV	0.000	0.000		-0.000		0.000	-0.000
(Incoming network product review volume, co-view)	(0.000)	(0.000)		(0.000)		(0.000)	(0.000)
IPV_CP	-0.000	-0.000			-0.000	-0.000	-0.000
(Incoming network product review volume, co-purchase)	(0.000)	(0.000)			(0.000)	(0.000)	(0.000)
OPV_CV	0.000		0.000	0.000		0.000**	-0.000
(Outgoing network product review volume, co-view)	(0.000)		(0.000)	(0.000)		(0.000)	(0.000)
OPV_CP	-0.000		-0.000		-0.000**	0.000	-0.000
(Outgoing network product review volume, co-purchase)	(0.000)		(0.000)		(0.000)	(0.000)	(0.000)
IPR_CV	-0.002**	-0.002***		-0.001		-0.002**	-0.002*
(Incoming network product review rating, co-view)	(0.001)	(0.001)		(0.001)		(0.001)	(0.001)
IPR_CP	0.000	-0.003***			-0.003***	-0.002***	-0.002**
(Incoming network product review rating, co-purchase)	(0.001)	(0.001)			(0.001)	(0.001)	(0.001)
OPR_CV	0.002*		0.000	0.000		0.001	0.001
(Outgoing network product review rating, co-view)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
OPR_CP	-0.001		-0.001		0.002*	0.000	-0.002
(Outgoing network product review rating, co-purchase)	(0.001)		(0.001)		(0.001)	(0.001)	(0.001)

Note: Standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 2-8 - Model Estimation Results (Continued)

Variable	(1) RE: Control	(2) RE: Incoming	(3) RE: Outgoing	(4) RE: Co-view	(5) RE: Co-purchase	(6) RE: Full	(7) FE: Full
IQ_CV	0.002	0.002		0.001		0.002	0.002
(Incoming network product sales quantity, co-view)	(0.002)	(0.002)		(0.002)		(0.002)	(0.001)
IQ_CP	0.005***	0.006***			0.005***	0.005***	0.004***
(Incoming network product sales quantity, co-purchase)	(0.002)	(0.002)			(0.002)	(0.002)	(0.002)
OQ_CV	0.004**		0.005**	0.004*		0.004**	0.003
(Outgoing network product sales quantity, co-view)	(0.002)		(0.002)	(0.002)		(0.002)	(0.002)
OQ_CP	0.004***		0.004***		0.003***	0.004***	0.004***
(Outgoing network product sales quantity, co-purchase)	(0.001)		(0.001)		(0.001)	(0.001)	(0.001)
SE	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***	-0.004***
(Substitution effect)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CE	0.009***	0.008***	0.009***	0.009***	0.008***	0.009***	0.007***
(Complementarity effect)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.044***	0.004	0.039***	0.029*	0.019	0.047***	0.113
	(0.015)	(0.014)	(0.015)	(0.015)	(0.014)	(0.016)	(1,301.134)
Category & *time dummies*	-included-	-included-	-included-	-included-	-included-	-included-	-include-
Number of observations	41,379	41,379	41,379	41,379	41,379	41,379	41,379
Hausman test						$\chi^2 = 0.00$, $p = 1.00$	
R^2	0.4925	0.4919	0.4916	0.4913	0.4922	0.4936	0.1017

Note: Standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Based on these control variables, we then estimate a RE model including the independent variables (network diversity and stability) of only the incoming network. As reported in Table 2-8, Column (2), the coefficients of co-purchase network diversity (*ID_CP*) and co-view network stability (*IS_CV*) are statistically significant. Similarly, we estimate a RE model including the independent variables of only the outgoing network and summarize the results in Table 2-8, Column (3). We find that three model coefficients for outgoing network variables (*OD_CP*, *OS_CV*, *OS_CP*) are statistically significant. These results suggest the importance of accounting for both incoming and outgoing networks when assessing the demand impacts of recommendation networks.

We next estimate a RE model including the independent variables (incoming and outgoing network diversity and stability) of only the co-view network. As reported in Table 2-8, Column (4), the coefficients of incoming network diversity (*ID_CV*) and incoming network stability (*IS_CV*) are statistically significant. However, these variables become insignificant in subsequent model estimations which include co-purchase network related variables. This highlights the importance of investigating both co-view and co-purchase recommendation networks for a more reliable and comprehensive examination.

Likewise, we also estimate a RE model including the independent variables of only the co-purchase network. As reported in Table 2-8, Column (5), the coefficients of the incoming network diversity (*ID_CP*), outgoing network diversity (*OD_CP*), and outgoing network stability (*OS_CP*), are all highly significant and with the expected signs. By further comparing the model fit statistics between the co-view RE model in Column (4) (R^2 = 0.4913) and the co-purchase RE model in Column (5) (R^2 = 0.4922), we find that co-purchase network factors, relative to co-view network factors, explain more variations in the demand of a product.

Next, we further estimate a full RE model including both incoming and outgoing network variables, and both co-view and co-purchase network variables. Table 2-8, Column (6), summarizes the results. Three out of the eight independent variables are statistically significant. First, incoming co-purchase network diversity (*ID_CP*) is positive and significant, suggesting an expected positive relationship with the focal product's demand. Second, outgoing co-purchase network diversity (*OD_CP*) is negative and significant, suggesting an expected negative relationship with the focal product's demand. Lastly, outgoing co-purchase network stability (*OS_CP*) also has an expected negative relationship with the focal product's demand.

In addition to the RE model, we further estimate a full fixed effects (FE) model of product sales quantity on all the explanatory variables and summarize the results in Table 2-8, Column (7). The R^2 of the FE model (R^2 = 0.1017) is substantially lower than that of the full RE model (R^2 = 0.4936) in Column (6), suggesting that RE model has a better model fit than the FE one. More importantly, the Hausman test ($\chi^2 = 0.00$, $p = 1.00$) suggests that the RE model estimates are not inconsistent, implying the appropriateness of the RE model over the FE one. Accordingly, we consider the RE model in Column (6) as our preferred model specification.

One may be concerned about the potential simultaneity between network structures (i.e., diversity, stability) and product demand. To address this issue, we report two additional model estimations. Specifically, we estimate the one-day and seven-day lagged effect of network structure on current product sales quantity. The network structure in the past should not have been influenced by the current product sales quantity. Thus, this can rule out the potential simultaneity concern. The estimation results are reported in Table 2-9, Columns (2) to (3). For brevity, from this point onward, we only report the major variables of interest for hypothesis testing. For ease of reference, Table 2-9, Column (1), presents the results from Table 2-8, Column (6). As indicated, those previously identified three significant network diversity and stability estimates are all consistent with those in Column (1).

Table 2-9 - Simultaneity Checks

Variable	(1) Preferred	(2) Lag 1	(3) Lag 7	(4) Reduce bidirectionality (=100%)	(5) Reduce bidirectionality (>50%)	(6) Reduce bidirectionality (>10%)
ID_CV	0.001	-0.002	-0.010***	0.002	0.005	0.009
(Incoming network diversity, co-view)	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)	(0.014)
ID_CP	0.012***	0.008***	0.010***	0.012***	0.015***	0.021**
(Incoming network diversity, co-purchase)	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)	(0.010)
OD_CV	-0.007	-0.001	-0.005	-0.006	-0.007	0.008
(Outgoing network diversity, co-view)	(0.004)	(0.004)	(0.003)	(0.005)	(0.007)	(0.014)
OD_CP	-0.007***	-0.003**	-0.004***	-0.009***	-0.013***	-0.033***
(Outgoing network diversity, co-purchase)	(0.002)	(0.001)	(0.001)	(0.002)	(0.003)	(0.009)
IS_CV	0.006	-0.004	0.001	0.006	0.010*	0.008
(Incoming network stability, co-view)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.013)
IS_CP	-0.002	0.000	0.005	-0.002	-0.002	-0.006
(Incoming network stability, co-purchase)	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)	(0.013)
OS_CV	0.005	0.009**	0.004	0.006	0.011*	0.019
(Outgoing network stability, co-view)	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.013)
OS_CP	-0.017***	-0.011***	-0.009***	-0.018***	-0.027***	-0.060***
(Outgoing network stability, co-purchase)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.014)
Constant	0.047***	0.049***	0.055***	0.049**	0.057**	0.139**
	(0.016)	(0.016)	(0.016)	(0.019)	(0.029)	(0.062)
Control variables	-included-	-included-	-included-	-included-	-included-	-included-
Number of observations	41,379	40,836	38,927	34,070	22,279	8,200
Sample reduced				17.664%	46.159%	80.183%
R^2	0.4936	0.4949	0.5021	0.5025	0.5339	0.6268

Note: Standard errors in parentheses. * *p*<0.1, ** *p*<0.05, *** *p*<0.01.

Moreover, Oestreicher-Singer and Sundararajan (2012b) excluded the products for which all incoming network links are bidirectional for each day to control for potential simultaneity. We use a similar approach by excluding products for which all incoming network links are bidirectional (100% bidirectionality, i.e., the focal product has outgoing network links terminating at all its incoming network products). More than 17% of our observations are eliminated by this operationalization. The estimation results based on this reduced sample are summarized in Table 2-9, Column (4). As indicated, they are consistent with those in Column (1). Likewise, more conservatively, we further exclude products for which more than 50%, and even more than 10%, of the incoming network links are bidirectional (i.e., the focal product has outgoing network links terminating at more than 50%, and even more than 10%, of its incoming network products). These two operationalizations reduce the observations by 46.159% and 80.183% respectively. The model estimates are reported in Table 2-9, Columns (5) and (6) respectively, and are still consistent with those in Column (1). These operationalizations do not eliminate all cycles from our sample data, but our results here do significantly alleviate concerns of potential simultaneity[41].

Additionally, to address the concern of simultaneity, we further conduct a series of Granger causality tests (Granger 1969) to check whether current product network structure (i.e., diversity and stability) would have been affected by past product sales performance (i.e., sales quantity). Specifically, we use different time lag levels from one to seven days to test whether product sales quantity Granger-causes each of the network diversity and stability variables. The test results show that current product network structures have not been affected by past product sales quantity (i.e., simultaneity would not be a concern in this study). Thus, this lends further credence to the impacts of network diversity and stability on product demand. We summarize the test results in Table 2-10.

[41] We further lower the cut-off values of the bidirectionality percentage, and find that the model estimates generally maintain their consistency across different estimation samples.

Table 2-10 - Granger Causality Tests

No. of lags	*ID_CV*		*ID_CP*		*OD_CV*		*OD_CP*		*IS_CV*		*IS_CP*		*OS_CV*		*OS_CP*	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
1	0.638	0.425	0.024	0.876	2.585	0.108	0.511	0.475	2.359	0.125	2.731	0.098	2.404	0.121	3.054	0.081
2	2.902	0.234	0.653	0.722	3.159	0.206	0.570	0.752	3.130	0.209	3.814	0.148	3.335	0.189	4.997	0.082
3	1.934	0.586	0.977	0.807	1.658	0.646	1.392	0.707	4.759	0.190	4.980	0.173	4.240	0.237	6.169	0.104
4	2.253	0.689	2.715	0.607	3.913	0.418	4.265	0.371	8.189	0.085	7.910	0.095	4.472	0.346	6.735	0.151
5	3.626	0.604	4.631	0.463	4.154	0.528	4.011	0.548	7.954	0.159	8.630	0.125	4.349	0.500	7.496	0.186
6	3.623	0.728	5.855	0.440	3.511	0.742	4.183	0.652	7.534	0.274	8.785	0.186	5.460	0.486	8.700	0.191
7	3.700	0.814	7.405	0.388	3.633	0.821	4.300	0.745	8.251	0.311	10.569	0.159	5.640	0.582	9.051	0.249

Lastly, to further corroborate the impact of network diversity and stability on product demand, we conduct the shuffle test (Anagnostopoulos et al. 2008) to rule out the correlation and validate the influence (causality). This test is based on the idea that if influence does not play a role, even though a product's demand could depend on the neighboring products in the network, the timing of such dependency should be independent of the neighboring products. Thus, for each focal product, we randomly shuffle its network connections over the entire sample period for each day to reconstruct similar measures. We obtain the estimated network diversity and stability parameters before and after the shuffling, and then test for the structural difference across these two sets of parameters. The Chow test result (F = 2.96, p = 0.00) shows that the two sets of parameters are significantly different, which suggests the existence of causality rather than correlation.

In summary, after accounting for the implicit demand correlation and potential simultaneity, we identify three significant effects of product network structures (i.e., *ID_CP*, *OD_CP*, *OS_CP*). To summarize our hypothesis testing results, we further report the elasticities for these three significant factors in Table 2-11 based on the preferred model. First, incoming co-purchase network diversity (*ID_CP*, elasticity = 0.014, $p < 0.01$) has a positive relationship with the focal product's demand, thus supporting H1A. Second, outgoing co-purchase network diversity (*OD_CP*, elasticity = -0.011, $p < 0.01$) has a negative relationship with the focal product's demand, thus supporting H1B. Next, outgoing co-purchase network stability (*OS_CP*, elasticity = -0.012, $p < 0.01$) has a negative relationship with the focal product's demand, thus supporting H2B. However, incoming network stability (*IS_CV* and *IS_CP*) has no significant relationship with product demand, thus H2A is not supported.

Table 2-11 - Elasticities

Network diversity	*ID_CV*	*ID_CP*	*OD_CV*	*OD_CP*
Elasticity	-	0.014***	-	-0.011***
Network stability	*IS_CV*	*IS_CP*	*OS_CV*	*OS_CP*
Elasticity	-	-	-	-0.012***

Note: * p<0.1, ** p<0.05, *** p<0.01.

Finally, in order to compare the relative impact of co-view network and co-purchase network factors, we report in Table 2-12 the standardized regression coefficients by standardizing all the variables in our model (Darlington 1990). As indicated, network diversity and stability variables are significant only in the co-purchase network. This suggests the more influential role of co-purchase network than co-view network in affecting product demand. Additionally, we also conduct likelihood ratio test to compare the fit of the co-view nested model (i.e., the co-view model in Table 2-8, Column (4)) and the full model (i.e., the model in Table 2-8, Column (6)). The test result (χ^2 = 122.18, p = 0.00) shows that the full model significantly fits better than the co-view nested model, implying that co-purchase network variables have important power in explaining product demand. Similarly, we conduct likelihood ratio test to compare the fit of the co-purchase nested model (i.e., the co-purchase model in Table 2-8, Column (5)) and the full model (i.e., the model in Table 2-8, Column (6)). The test result (χ^2 = 18.56, p = 0.18) shows that the full model does not significantly fit better than the co-purchase nested model, implying that co-view network variables do not significantly help explain product demand. Thus, this again suggests that co-purchase network has a stronger impact than co-view network on product demand. Therefore, both approaches indicate that H3 is supported. We summarize all hypothesis testing results in Table 2-13.

Table 2-12 - Standardized Regression Coefficients

Network diversity	*ID_CV*	*ID_CP*	*OD_CV*	*OD_CP*
Std. coefficient	-	0.055***	-	-0.031***
Network stability	*IS_CV*	*IS_CP*	*OS_CV*	*OS_CP*
Std. coefficient	-	-	-	-0.023***

Note: * p<0.1, ** p<0.05, *** p<0.01.

Table 2-13 - Hypothesis Testing Results

Hypothesis		Support
H1A	Network diversity (incoming) $\rightarrow$ Demand (+)	Yes (co-purchase)
H1B	Network diversity (outgoing) $\rightarrow$ Demand (-)	Yes (co-purchase)
H2A	Network stability (incoming) $\rightarrow$ Demand (-)	No
H2B	Network stability (outgoing) $\rightarrow$ Demand (-)	Yes (co-purchase)
H3	Co-purchase > Co-view	Yes

3.5.2 Robustness Checks

We further corroborate our findings by checking its robustness in multiple ways. We first report robustness checks on the sensitivity of our operationalizations by constructing the substitution and complementarity effects (*SE* and *CE*) in Equations (3) and (4) in different ways. For ease of reference, Table 2-14, Column (1), presents the results from our preferred model in Table 2-8, Column (6).

Table 2-14 - Robustness Checks (1)

Variable	(1) Preferred	(2) Significant elasticity only	(3) Without elasticity weight	(4) 1st-stage uncertainty
ID_CV	0.001	0.001	0.006***	0.001
(Incoming, diversity, co-view)	(0.003)	(0.003)	(0.002)	(0.003)
ID_CP	0.012***	0.012***	0.010***	0.012***
(Incoming, diversity, co-purchase)	(0.002)	(0.002)	(0.001)	(0.002)
OD_CV	-0.007	-0.007*	-0.004	-0.006
(Outgoing, diversity, co-view)	(0.004)	(0.004)	(0.003)	(0.004)
OD_CP	-0.007***	-0.007***	-0.005***	-0.007***
(Outgoing, diversity, co-purchase)	(0.002)	(0.002)	(0.001)	(0.002)
IS_CV	0.006	0.006	0.004	0.006
(Incoming, stability, co-view)	(0.003)	(0.003)	(0.003)	(0.003)
IS_CP	-0.002	-0.002	-0.000	-0.001
(Incoming, stability, co-purchase)	(0.003)	(0.003)	(0.002)	(0.003)
OS_CV	0.005	0.005	0.006*	0.005
(Outgoing, stability, co-view)	(0.004)	(0.004)	(0.003)	(0.004)
OS_CP	-0.017***	-0.017***	-0.012***	-0.017***
(Outgoing, stability, co-purchase)	(0.003)	(0.003)	(0.003)	(0.003)
Constant	0.047***	0.048***	0.444***	0.042***
	(0.016)	(0.016)	(0.013)	(0.016)
Control variables	-included-	-included-	-included-	-included-
Number of observations	41,379	41,379	41,379	41,379
R^2	0.4936	0.4926	0.6623	0.4890

Note: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

First, some of the estimated cross-category price elasticities in Table 2-7 are not significant, which might imply that two products or categories do not have demand correlations. Thus, we replace the values of all the insignificant price elasticities with zeros for the constructions of *SE* and *CE*. The results in Table 2-14, Column (2) are quite close to those in Column (1).

Second, we construct *SE* and *CE* using the sum of actual demands without using the estimated price elasticities as weights. Results shown in Table 2-14, Column (3) are generally consistent with those reported in Column (1).

Lastly, another concern in this study is that *SE* and *CE* in the main estimation (i.e., second stage) are constructed using the estimated price elasticities from the previous stage (i.e., first stage). As such, we take into account the first-stage uncertainty of estimated coefficients (as reflected in standard errors) in the second-stage estimation. Specifically, in Equations (3) and (4) where *SE* and *CE* are computed, we further include the first-stage estimated standard errors as weights for the elasticity estimates. We then estimate a RE model with *SE* and *CE* replaced by the 1st-stage uncertainty weighted substitution effect and complementarity effect. The results are summarized in Table 2-14, Column (4). As indicated, the network diversity and stability estimates are consistent with the results of our preferred specification in Column (1).

In addition to the robustness checks on the sensitivity of different constructions of *SE* and *CE*, we also check the robustness of our findings in many other ways. For ease of reference, Table 2-15, Column (1), also presents the results from our preferred model in Table 2-8, Column (6).

Table 2-15 - Robustness Checks (2)

Variable	(1) Preferred	(2) Mean-centering	(3) Standardi-zation	(4) Hetero-skedasticity	(5) Serial correlation	(6) Price endogeneity	(7) Relative difference	(8) Stability 3-day	(9) Stability 7-day	(10) Market share	(11) Actual day	(12) Weekly	(13) PC
ID_CV	0.001	0.001	0.001	0.001	0.003	0.002	0.001	0.000	-0.000	-0.007	-0.001	0.047***	0.001
(Incoming, diversity, co-view)	(0.003)	(0.003)	(0.002)	(0.009)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.119)	(0.003)	(0.013)	(0.004)
ID_CP	0.012***	0.012***	0.015***	0.012*	0.007***	0.014***	0.012***	0.012***	0.012***	0.641***	0.008***	0.071***	0.006**
(Incoming, diversity, co-purchase)	(0.002)	(0.002)	(0.002)	(0.007)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.066)	(0.001)	(0.007)	(0.003)
OD_CV	-0.007	-0.007	-0.003	-0.007*	-0.004	-0.010*	-0.007	-0.003	-0.002	0.034	-0.003	-0.063***	-0.003
(Outgoing, diversity, co-view)	(0.004)	(0.004)	(0.002)	(0.004)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)	(0.170)	(0.003)	(0.020)	(0.007)
OD_CP	-0.007***	-0.007***	-0.008***	-0.007	-0.004**	-0.008***	-0.007***	-0.006***	-0.006***	-0.274***	-0.005***	-0.032***	-0.016***
(Outgoing, diversity, co-purchase)	(0.002)	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.072)	(0.001)	(0.007)	(0.005)
IS_CV	0.006	0.006	0.002	0.006	0.005	0.008	0.006	0.006*	0.003	0.055	0.003	0.039*	0.010
(Incoming, stability, co-view)	(0.003)	(0.003)	(0.001)	(0.004)	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.144)	(0.004)	(0.020)	(0.012)
IS_CP	-0.002	-0.002	-0.001	-0.002	-0.003	-0.002	-0.002	-0.001	0.001	-0.004	-0.001	0.028	0.010
(Incoming, stability, co-purchase)	(0.003)	(0.003)	(0.001)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.121)	(0.003)	(0.017)	(0.009)
OS_CV	0.005	0.005	0.002	0.005*	0.006	0.005	0.005	0.011***	0.010***	0.361**	0.007*	0.050**	0.028*
(Outgoing, stability, co-view)	(0.004)	(0.004)	(0.001)	(0.003)	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.161)	(0.004)	(0.024)	(0.015)
OS_CP	-0.017***	-0.017***	-0.006***	-0.017*	-0.013***	-0.020***	-0.017***	-0.011***	-0.008**	-0.451***	-0.019***	-0.102***	-0.027***
(Outgoing, stability, co-purchase)	(0.003)	(0.003)	(0.001)	(0.009)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.134)	(0.004)	(0.020)	(0.010)
Constant	0.047***	0.039**	0.039**	0.047	0.039**	0.092***	0.047***	0.038**	0.040***	-0.039	0.044***	-0.034	0.149**
	(0.016)	(0.016)	(0.016)	(0.064)	(0.018)	(0.022)	(0.016)	(0.015)	(0.015)	(0.655)	(0.016)	(0.024)	(0.060)
Control variables	-included-	-included-	-included-	-included-	-included-	-included-	-included-	-included-	-included-	-included-	-included-	-included-	-included-
Number of observations	41,379	41,379	41,379	41,379	41,379	29,172	41,379	40,931	40,030	41,379	41,379	7,800	15,616
R^2	0.4936	0.4936	0.4936	0.4936	0.4754	0.5037	0.4936	0.4947	0.4961	0.4037	0.4938	0.6508	0.4597

Note: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

First, one may be concerned that the changes in a focal product's network (leading to instability in a network) may introduce different products into this focal product's network (i.e., increasing the diversity of the network). Thus, the potential collinearity between network diversity and network stability would be a concern. However, Table 2-4 shows that the correlations between network diversity and network stability are generally small, which suggest that collinearity would not be a concern in this study. Nevertheless, we still perform mean-subtracted centralization and standardization to all the independent variables. Then, we estimate Equation (1) based on these variables and summarize the results in Table 2-15, Columns (2) and (3) respectively. The results are consistent with those from our preferred specification.

Second, we check whether our results are robust in the presence of heteroskedasticity. We estimate robust standard errors clustered by product. As indicated in Table 2-15, Column (4), the results are generally consistent.

Third, to account for the existence of potential serial correlation, we estimate a RE model with a first-order autoregressive (AR1) disturbance structure. As indicated in Table 2-15, column (5), the model estimates under an AR1 structure are generally consistent with those of the preferred one in Column (1).

Fourth, we check whether our findings are robust after accounting for price endogeneity. Specifically, we treat product list price (*LP*) as an endogeneous variable and use the instrumental variables estimation method. The choice of instrumental variable is the average list price of the same product on the previous day from three additional Tmall stores which also exclusively sell Nikon products. The price from another Nikon store would have high correlation with the price of the same product in the focal store, since retailers would typically set comparable prices for the same product. However, prices from another store on the previous day are unlikely to shift the current demand of products in the focal store. Hence, we believe price from alternative stores on the previous day could serve as a reasonable instrument. We perform RE two-stage least-squares estimation and Table 2-15, Column (6), summarizes the estimated model coefficients which are relatively consistent with those in Column (1).

Fifth, some consumers may care more about the relative differences between the recommended products and the focal product in terms of some product attributes (e.g., price, review volume and review rating) shown on a product page, rather than the absolute values of those attributes of the recommended products. Thus, we replace the control variables of the average list price, review volume and review rating of incoming and outgoing co-view and co-purchase network products with the differences between these average values and the value of the corresponding attribute of the focal product. As shown in Table 2-15, Column (7), the model estimation results are consistent with those in Column (1).

Sixth, we check the robustness of our findings across differences in variable operationalizations. Specifically, instead of using the daily network overlap as the measure for network stability, we use the network overlap of today and three days ago as an alternative measure. The similar results based on this new measure are summarized in Table 2-15, Column (8). Interestingly, in terms of the magnitude of elasticity, the impact of *OS_CP* drops from 0.012 in the preferred model to 0.008. We further use the network overlap of today and seven days ago as the stability measure and report the results in Table 2-15, Column (9). The impact of *OS_CP* further diminishes (elasticity magnitude = 0.005). This suggests that the network stability (or the update in recommendations) on a daily basis has a larger impact on product demand, compared to that based on a longer time period. Next, we use a product's daily market share within category as an alternative measure for the dependent variable. The model estimates based on this new measure which are summarized in Table 2-15, Column (10), report similar findings. Moreover, instead of using the average values of a variable on day t and day t+1 as the measure for this variable on day t, we use the actual day t values. Consistent results using this operationalization are shown in Table 2-15, Column (11).

Next, we check the robustness of our findings across timing differences. We use the weekly time frame instead of the daily time frame by computing the average values across all days in each week for our model variables. The estimates based on the weekly time frame are reported in Table 2-15, Column (12). The findings under a weekly time frame are similar to those using a daily level of analysis.

Finally, we check whether our findings are robust across product categories. We collect data from another Tmall store which sells personal computers (PC) and associated components. Results in Table 2-15, Column (13) report findings that are similar to those using the digital camera category.

In summary, we are confident of the robustness of our findings given that all the various checks indicate robustness and consistency of our findings in the presence of potential collinearity, heteroskedasticity, serial correlation, price endogeneity, and across differences in variable operationalizations, time frames, and product categories.

3.6 Discussion and Contribution

3.6.1 Discussion of Findings

Our study that investigates the impact of network diversity and network stability on product demand in an e-commerce setting has several notable findings. First, we empirically show that the diversity and stability of a product's network are related to the sales quantity of the product. Specifically, a 1% increase in the category diversity of the incoming (outgoing) co-purchase network of a product increases (decreases) the product's sales quantity by 0.014% (0.011%). Interestingly, diversity of both the incoming and outgoing networks overall exhibits a positive effect on product demand. Moreover, a 1% increase in the stability of the outgoing co-purchase network of a product decreases the product's sales quantity by 0.012%.

Second, by operationalizing a product's network into incoming network and outgoing network for investigation, we identify that the demand of a product is influenced by both the incoming and outgoing networks of the product. Our results show that the incoming network influences product demand only through diversity effects whereas the outgoing network influences it through both diversity and stability effects. This indicates that the outgoing network links of a product in an e-commerce site may have more mechanisms in affecting product demand than those of the incoming network, which highlights the incomplete view of prior studies that only focused on the incoming network impact.

Finally, by further differentiating a product's network into co-view and co-purchase networks, we surprisingly find that only co-purchase network affects product demand. Specifically, co-purchase network exhibits both significant diversity and stability effects on product demand, whereas co-view network diversity and stability effects on product demand are insignificant. Thus, our results show that overall, co-purchase network has a stronger impact on product demand than co-view network.

3.6.2 Theoretical Contributions

Our study offers important theoretical contributions in the following ways. First, prior product network research has made attempts to analyze the economic impact (e.g., sales impact) of product network structures (Carmi et al. 2011; Carmi et al. 2010; Oestreicher-Singer and Sundararajan 2012a; Oestreicher-Singer and Sundararajan 2012b). Unfortunately, past studies have invariably focused only on the impact of incoming network. Our study advances existing product network literature by validating the impact of both incoming and outgoing networks. For instance, the impact of network degree centrality is investigated in both our study and Oestreicher-Singer and Sundararajan (2012b). Although we both identify similar negative impact of in-degree centrality, our estimation of both incoming and outgoing networks find that out-degree centrality, *OC_CV* (elasticity = -0.034), actually has a stronger impact compared with the impact of in-degree centrality (elasticity of *IC_CV* = -0.002, elasticity of *IC_CP* = -0.009). The above results and discussions of our main findings of network diversity and stability impacts have provided further evidences to the observation that the prior view that only focused on the incoming network in product network research may derive incomplete and even erroneous research conclusions. More broadly, our more complete examination offers new insights for investigating effects of networks in other fields (e.g., social networks).

Second, various studies on product recommendation systems (Carmi et al. 2011; Carmi et al. 2010; Oestreicher-Singer and Sundararajan 2012b; Pathak et al. 2010) have identified some positive economic impacts of product recommendation and reported its benefits (e.g., recommendation systems lead to sales increase). However, in addition to the positive impacts, our study also unravels the potential negative economic impacts of product recommendations resulting from outgoing co-purchase network diversity and stability effects.

Our research findings thus challenge conventional wisdom by elucidating the detrimental effect of recommendation systems on product demand.

Third, in addition to the centrality effect on product demand identified in prior product network literature (Carmi et al. 2011; Goldenberg et al. 2012; Oestreicher-Singer and Sundararajan 2012a; Oestreicher-Singer and Sundararajan 2012b), our research is the first to study and validate the diversity effect and stability effect of product networks. More importantly, by comparing the effects between the prior investigated degree centrality factors and our proposed diversity and stability factors, we further find that the effects of all the significant degree centrality factors (except *OC_CV*), including *IC_CV* (elasticity = -0.002) and *IC_CP* (elasticity = -0.009), are smaller than the effects of network diversity, *ID_CP* (elasticity = 0.014) and *OD_CP* (elasticity = -0.011), and network stability, *OS_CP* (elasticity = -0.012). Our findings thus suggest that network diversity and stability are more influential attributes than degree centrality in driving product demand in product recommendation network contexts.

Fourth, by juxtaposing the role of co-view recommendation networks besides that of co-purchase recommendation networks, we unravel the contention and intricacies between the two. Our findings suggest that co-purchase networks affect product demand through diversity and stability effects, while differently, co-view network diversity and stability effects are insignificant. The differential and even contrasting impact of these two types of networks suggests that consumers not only respond to the diversity and stability of product recommendations, but also factor the mechanism of recommendations into consideration. Our findings thus complement and enrich prior work to provide a more comprehensive understanding of the economic impact of product networks.

Fifth, our research also contributes to the product network literature by proposing and demonstrating an empirical strategy to identify the implicit demand correlation in terms of product substitution and complementarity effects that may confound a rigorous exploration of the demand impact of product networks. Compared to the assumptions used in Oestreicher-Singer and Sundararajan (2012b), our approach here is a more direct and econometrically rigorous approach that accounts for the substitution and complementarity effects of other products that have intrinsic demand correlations, or may appear as visible network links, or are currently invisible but potentially future network links. Thus, our approach provides guidance to future work in similar e-commerce settings that investigate similar research questions.

Finally, our research is one of the rare studies to examine product network impacts using actual demand information from e-commerce retailers. Prior research only used some demand proxies such as sales rank on Amazon.com for empirical analysis. Although some studies have documented that "actual demand" can be obtained by a log-linear transformation of sales rank (Brynjolfsson et al. 2003; Goolsbee and Chevalier 2002), the precision of research findings and especially the quantitative insights may have already been compromised via this transformation. Our empirical analysis in this paper thus is able to provide more accurate and reliable conclusions by virtue of the superiority of our demand data.

3.6.3 Practical Implications

Our study also provides important practical implications for e-commerce retailers to drive product sales using recommendation systems. First, retailers can take advantage of the diversity effect to drive product sales. According to our results, the category diversity in the incoming co-purchase network of a product can increase the product's demand. As such, retailers could configure co-purchase recommendation systems to have more diverse incoming links for a product in terms of more heterogeneous product categories. Conversely, the category diversity in the outgoing co-purchase network has a negative relationship with product demand. This suggests that retailers should limit the category diversity of outgoing links in the co-purchase network on a product's page. For instance, retailers can provide co-purchase recommendations restricted to only one or two product categories for each focal product in order to minimize the probability of losing the interest of potential buyers of the focal product.

Second, e-commerce retailers can capitalize on the network stability effect as well to influence product demand. Specifically, the negative stability effect of the outgoing co-purchase network suggests that retailers should frequently change the constituent items in the co-purchase recommendation list in order to positively influence the demand of the focal product.

Third, comparing the effects between co-view and co-purchase recommendation systems, co-purchase recommendation systems exhibit a stronger impact than co-view recommendation systems in driving product demand. Accordingly, retailers could allocate a higher percentage of the limited web page space to more saliently display co-purchase recommendations to achieve better sales performance.

Finally, our study also offers implications for the design of product recommendation systems. As existing product recommendation systems are mainly automated (i.e., automated algorithm execution to generate recommendations) (Sarwar et al. 2000), e-commerce platform or retail site operators may provide interfaces and access to individual brand owners for them to implement the above suggested strategies to drive product sales online. More ideally, recommendation system designers could further develop and integrate these functionalities based on the above insights for retailers' implementation.

3.7 Conclusion

While this research has highlighted several notable findings and important contributions, we acknowledge some limitations. First, while our empirical strategy attempts to control for various observed product attributes such as price, product reviews and inventory, as well as unobserved implicit demand correlations through accounting of both substitution and complementary effects of other network-linked and unlinked products, these approaches may not have fully controlled for all potential sources of endogeneity bias such as endogenous link formations. As such, we do not make causality claims of the impacts of network diversity and stability on product demands. However, as argued by Sundararajan et al. (2013), the lack of a causal mechanism in network studies should not preclude the usefulness and contribution of predictive modeling based on correlation solely. Above all in this study, we conduct many robustness checks such as the Granger causality test whose results all attest to the econometric rigor of our findings here. Second, the impact of recommendation networks could be more precisely quantified if we had observations of individual consumers' clickstream data. Using clickstream data, we could accurately associate the recommendation network structures with consumers' browsing and purchase behaviors for more insightful examinations.

Moving forward, we present potential avenues for future research. To critically and rigorously evaluate the causal impacts of network diversity and stability on demand, large-scale field experimentations of product recommendation network structures could be attempted in cooperation with an e-commerce site operator. Such an effort would shed deep insights into the causal demand effects generated by explicit systematic manipulations of product recommendation links included for a focal product. Another meaningful extension to this research is to investigate the spillover effects of online word-of-mouth (WOM). Due to the use of recommendation systems, WOM information of different products is now commonly connected in networks within a retail site. Therefore, it would be interesting to study the demand impact of a product's WOM on other connected products in the network. This would have important implications for retailers' e-commerce product co-location practices. Another important direction for future research would be to investigate the relative impact and especially the interplay between these automated recommendations (co-view and co-purchase) and the consumer-generated WOM recommendations. One could thus identify whether these two recommendation mechanisms are substitutable, complementary or independent in driving product demand. The findings would have important implications for online retailers' marketing strategies and the design of e-commerce websites.

4. GENERAL CONCLUSION

The aim of this research is to investigate the economic impact of online networks and online recommendations by conducting two empirical studies to examine how user interactions or recommendations in social media-enabled social networks affect consumer purchase behavior in study 1, and then how structures of e-commerce-enabled product recommendation networks influence product demand in study 2.

Our findings from study 1 show that consumers' engagement in social media brand community networks increases their purchase expenditures. Network interaction contents UGC and MGC may affect consumer purchase behavior through embedded information and persuasion. Specifically, the elasticities of demand with respect to UGC information richness are 0.006 (directed communication) and 3.140 (undirected communication), whereas those for MGC information richness are insignificant. The UGC valence elasticity of demand is 0.180 (undirected communication), while that for MGC valence is 0.004 (directed communication). UGC exhibits a stronger impact than MGC on consumer purchase behavior.

Our findings from study 2 show that a 1% increase in the category diversity of the incoming (outgoing) co-purchase network of a product increases (decreases) the product's demand by 0.014% (0.011%). A 1% increase in the stability of the outgoing co-purchase network of a product decreases the product's demand by 0.012%. These results show that the demand of a product is influenced by both the incoming and outgoing networks. Moreover, co-purchase network exhibits a stronger role than co-view network in affecting product demand.

Overall, the notable findings from this research provide significant contributions to the literature on the economic value of online networks and online recommendations, and also offer important guidance to firms' online network-based and recommendation-based business strategies.

To conclude, we further provide a comparison between the two studies and discuss the implications for research. First, the two studies differ in terms of the unit of analysis. The first study linked individual consumers' network interactions to their transaction records to investigate the impact of UGC and MGC on consumer purchase expenditure at the individual consumer level. The detailed individual-level dataset allowed us to account for more potential influencing factors and more accurately identify the causal impact of interest. However, the second study organized the dataset at the aggregated product-daily level, and thus some information (e.g., consumer browsing and purchase behavior) could not be observed. Future research could try to capture more disaggregated-level data (e.g., consumer clickstream data) to link what consumers observe and click under some system-generated product recommendations directly to what they purchase, so that the impact of system recommendations could be more precisely identified.

Second, product demand interdependency is a widespread phenomenon that has been observed in a variety of settings (Carmi et al. 2011). Especially it is widely recognized in marketing research that consumer purchases are correlated among products that are substitutes or complements for one another (Shocker et al. 2004). Therefore, product substitution and complementarity effects should be taken into account in this research. In the first study, due to the data limitation, we could not observe the apparel categories purchased by consumers. Thus in our exploration of the impact of UGC and MGC on consumer purchase expenditure, we were not able to identify the substitution or complementarity nature of the purchased products to isolate this confounding factor. However in the second study, the detailed information from the dataset allowed us to explicitly model and account for the substitution and complementarity effects when we investigate the impact of product network structure on product demand. The substitution and complementarity effects play a more important role in the product recommendation network context as the nature of product recommendation systems is to up-sell, where retailers need to identify substitutes, or cross-sell, where retailers need to identify complements (Schafer et al. 2001). Thus, this nature causes the non-random network connections in the product recommendation network, and clearly requires researchers' efforts to model the mechanism of network connections (i.e., product substitution and complementarity effects).

Lastly, the two studies in this research also differ in terms of product types. The product investigated in the first study is apparel, which is an experience product. Thus consumers would have difficulties in observing its true quality prior to their actual purchase or use of the product. Hence, user recommendations may be more useful than system recommendations because users are more likely to describe products in a more flexible and comprehensive manner than systems which can only convey few fixed and limited product attribute information. This perhaps suggests the more important role of user recommendations in social networks (relative to system recommendations in product networks) for driving consumer purchases of experience products. In contrast, products in the second study are digital cameras and related components. These are typical search products which can be well described based on highly standardized product attributes (e.g., pixels and dimensions). In such a case, user recommendations in numerous ways and styles might instead cause information overload or become less relevant, whereas system-generated recommendations contain information that can provide a better match with the standardized attributes of search products. As such, this might suggest the more important role of system recommendations (relative to user recommendations) for driving consumer purchases of search products. However, these conjectures have not yet been addressed in this research due to data limitation, but could serve as an interesting direction for future exploration.

BIBLIOGRAPHY

Aciar, S., Zhang, D., Simoff, S., and Debenham, J. 2007. "Informed Recommender: Basing Recommendations on Consumer Product Reviews," *IEEE Intelligent Systems* (22:3), pp. 39-47.

Adjei, M.T., Noble, S.M., and Noble, C.H. 2010. "The Influence of C2C Communications in Online Brand Communities on Customer Purchase Behavior," *Journal of the Academy of Marketing Science* (38:5), pp. 634-653.

Akerlof, G.A. 1970. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* (84:3), pp. 488-500.

Albuquerque, P., Pavlidis, P., Chatow, U., Chen, K.Y., and Jamal, Z. 2012. "Evaluating Promotional Activities in an Online Two-Sided Market of User-Generated Content," *Marketing Science* (31:3), pp. 406-432.

Alexa. 2012. "Top Sites in China." Retrieved May 30, 2012, from http://www.alexa.com/topsites/countries/CN

Algesheimer, R., Dholakia, U.M., and Herrmann, A. 2005. "The Social Influence of Brand Community: Evidence from European Car Clubs," *Journal of Marketing* (69:3), pp. 19-34.

Alibaba. 2012. "Company Overview." Retrieved May 30, 2012, from http://news.alibaba.com/specials/aboutalibaba/aligroup/index.html

Allenby, G.M., Arora, N., and Ginter, J.L. 1998. "On the Heterogeneity of Demand," *Journal of Marketing Research* (35:3), pp. 384-389.

Allenby, G.M., and Rossi, P.E. 1991. "Quality Perceptions and Asymmetric Switching between Brands," *Marketing Science* (10:3), pp. 185-204.

Amaldoss, W., and He, C. 2009. "Direct-to-Consumer Advertising of Prescription Drugs: A Strategic Analysis," *Marketing Science* (28:3), pp. 472-487.

Anagnostopoulos, A., Kumar, R., and Mahdian, M. 2008. "Influence and Correlation in Social Networks," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Las Vegas, NV, pp. 7-15.

Anderson, C. 2006. *The Long Tail*. New York: Hyperion Books.

Antweiler, W., and Frank, M.Z. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *Journal of Finance* (59:3), pp. 1259-1294.

Anupindi, R., Dada, M., and Gupta, S. 1998. "Estimation of Consumer Demand with Stock-out Based Substitution: An Application to Vending Machine Products," *Marketing Science* (17:4), pp. 406-423.

Aral, S., and Walker, D. 2011. "Creating Social Contagion through Viral Product Design: A Randomized Trial of Peer Influence in Networks," *Management Science* (57:9), pp. 1623-1639.

Arazy, O., Kumar, N., and Shapira, B. 2010. "A Theory-Driven Design Framework for Social Recommender Systems," *Journal of the Association for Information Systems* (11:9), pp. 455-490.

Austin, P.C. 2010. "Statistical Criteria for Selecting the Optimal Number of Untreated Subjects Matched to Each Treated Subject When Using Many-to-One Matching on the Propensity Score," *American Journal of Epidemiology* (172:9), pp. 1092-1097.

Bagozzi, R.P., and Dholakia, U.M. 2006. "Antecedents and Purchase Consequences of Customer Participation in Small Group Brand Communities," *International Journal of Research in Marketing* (23:1), pp. 45-61.

Barefoot, J.C., Grønbæk, M., Jensen, G., Schnohr, P., and Prescott, E. 2005. "Social Network Diversity and Risks of Ischemic Heart Disease and Total Mortality: Findings from the Copenhagen City Heart Study," *American Journal of Epidemiology* (161:10), pp. 960-967.

Berlyne, D.E. 1960. *Conflict, Arousal, and Curiosity*. New York: McGraw-Hill.

Berry, S.T. 1994. "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics* (25:2), pp. 242-262.

Bertrand, M., Duflo, E., and Mullainathan, S. 2004. "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics* (119:1), pp. 249-275.

Bickart, B., and Schindler, R.M. 2001. "Internet Forums as Influential Sources of Consumer Information," *Journal of Interactive Marketing* (15:3), pp. 31-40.

Brin, S., and Page, L. 1998. "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems* (30:1-7), pp. 107-117.

Brown, J.J., and Reingen, P.H. 1987. "Social Ties and Word-of-Mouth Referral Behavior," *Journal of Consumer Research* (14:3), pp. 350-362.

Brynjolfsson, E., Hu, Y., and Smith, M.D. 2006. "From Niches to Riches: Anatomy of the Long Tail," *Sloan Management Review* (47:4), pp. 67-71.

Brynjolfsson, E., Hu, Y.J., and Smith, M.D. 2003. "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science* (49:11), pp. 1580-1596.

Bultez, A., and Naert, P. 1988. "Sh. A.R.P.: Shelf Allocation for Retailers' Profit," *Marketing Science* (7:3), pp. 211-231.

Burke, M., Kraut, R., and Marlow, C. 2011. "Social Capital on Facebook: Differentiating Uses and Users," in *ACM CHI Conference on Human Factors in Computing Systems*, Vancouver, British Columbia, Canada, pp. 571-580.

Burt, R.S. 2000. "The Network Structure of Social Capital," in *Research in Organizational Behavior*, B. Staw and R. Sutton. New York, NY: JAI Press.

Carmi, E., Oestreicher-Singer, G., Libai, B., and Yassin, O. 2011. "Assessing Value in Product Networks," in *Twenty Third Workshop on Information Systems and Economics*, Shanghai.

Carmi, E., Oestreicher-Singer, G., and Sundararajan, A. 2010. "Is Oprah Contagious? Identifying Demand Spillovers in Online Networks," in *Twenty Second Workshop on Information Systems and Economics*, St. Louis, MI.

Chen, Y., and Xie, J. 2008. "Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix," *Management Science* (54:3), pp. 477-491.

Chevalier, J.A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43:3), pp. 345-354.

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), pp. 944-957.

Chu, J., Chintagunta, P.K., and Vilcassim, N.J. 2007. "Assessing the Economic Value of Distribution Channels: An Application to the Personal Computer Industry," *Journal of Marketing Research* (44:1), pp. 29-41.

Clemons, E.K., Gao, G.G., and Hitt, L.M. 2006. "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems* (23:2), pp. 149-171.

Corstjens, M., and Doyle, P. 1981. "A Model for Optimizing Retail Space Allocations," *Management Science* (27:7), pp. 822-833.

Crosby, L.A., and Stephens, N. 1987. "Effects of Relationship Marketing on Satisfaction, Retention, and Prices in the Life Insurance Industry," *Journal of Marketing Research* (24:4), pp. 404-411.

Cross, R., and Cummings, J.N. 2004. "Tie and Network Correlates of Individual Performance in Knowledge-Intensive Work," *Academy of Management Journal* (47:6), pp. 928-937.

Cummings, J.N. 2004. "Work Groups, Structural Diversity, and Knowledge Sharing in a Global Organization," *Management Science* (50:3), pp. 352-364.

Cummings, J.N., and Higgins, M.C. 2006. "Relational Instability at the Network Core: Support Dynamics in Developmental Networks," *Social Networks* (28:1), pp. 38-55.

Darlington, R.B. 1990. *Regression and Linear Models*. New York: McGraw-Hill.

Darwell, B. 2013. "Facebook Platform Now Encompasses More Than 50m Pages and 10m Apps." Retrieved March 22, 2013, from http://www.insidefacebook.com/2013/02/01/facebook-platform-now-encompasses-more-than-50m-pages-and-10m-apps/

Das, S.R., and Chen, M.Y. 2007. "Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science* (53:9), pp. 1375-1388.

Dellarocas, C. 2003. "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science* (49:10), pp. 1407-1424.

Dellarocas, C. 2006. "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms," *Management Science* (52:10), pp. 1577-1593.

Dellarocas, C., Katona, Z., and Rand, W. 2010. "Media, Aggregators and the Link Economy: Strategic Hyperlink Formation in Content Networks," *NET Institute Working Paper* (10-13).

Deng, L., and Poole, M.S. 2010. "Affect in Web Interfaces: A Study of the Impacts of Web Page Visual Complexity and Order," *MIS Quarterly* (34:4), pp. 711-730.

Dhar, V., and Chang, E. 2009. "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing* (23:November), pp. 300-307.

Dou, Y., Niculescu, M.F., and Wu, D.J. 2013. "Engineering Optimal Network Effects Via Social Media Features and Seeding in Markets for Digital Goods and Services," *Information Systems Research* (24:1), pp. 164-185.

Duan, W., Gu, B., and Whinston, A.B. 2008. "Do Online Reviews Matter? - An Empirical Investigation of Panel Data," *Decision Support Systems* (45:4), pp. 1007-1016.

Duan, W., Gu, B., and Whinston, A.B. 2009. "Informational Cascades and Software Adoption on the Internet: An Empirical Investigation," *MIS Quarterly* (33:1), pp. 23-48.

Eagle, N., Macy, M., and Claxton, R. 2010. "Network Diversity and Economic Development," *Science* (328:5981), pp. 1029-1031.

eMarketer. 2013. "For Driving Purchases, Customer Reviews Trump Pro Critics." Retrieved March 22, 2013, from http://www.emarketer.com/Article/Driving-Purchases-Customer-Reviews-Trump-Pro-Critics/1009625

Escalas, J.E. 2007. "Self-Referencing and Persuasion: Narrative Transportation Versus Analytical Elaboration," *Journal of Consumer Research* (33:4), pp. 421-429.

Facebook. 2013. "Facebook Reports Fourth Quarter and Full Year 2012 Results." Retrieved March 22, 2013, from http://investor.fb.com/releasedetail.cfm?ReleaseID=736911

Fang, X., Hu, P.J.-H., Li, Z., and Tsai, W. 2013. "Predicting Adoption Probabilities in Social Networks," *Information Systems Research* (24:1), pp. 128-145.

Fleder, D., and Hosanagar, K. 2009. "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity," *Management Science* (55:5), pp. 697-712.

Forman, C., Ghose, A., and Wiesenfeld, B. 2008. "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research* (19:3), pp. 291-313.

Ghose, A., Han, S., and Iyengar, R. 2011. "Network Stability and Social Contagion on the Mobile Internet," in *Proceedings of the Thirty Second International Conference on Information Systems*, Shanghai.

Ghose, A., and Ipeirotis, P.G. 2011. "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *IEEE Transactions on Knowledge and Data Engineering* (23:10), pp. 1498-1512.

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.

Gilly, M.C., Graham, J.L., Wolfinbarger, M.F., and Yale, L.J. 1998. "A Dyadic Study of Interpersonal Information Search," *Journal of the Academy of Marketing Science* (26:2), pp. 83-100.

Godes, D., and Mayzlin, D. 2004. "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science* (23:4), pp. 545-560.

Godes, D., and Mayzlin, D. 2009. "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science* (28:4), pp. 721-739.

Goh, K.Y., Hui, K.L., and Png, I.P.L. 2011. "Newspaper Reports and Consumer Choice: Evidence from the Do Not Call Registry," *Management Science* (57:9), pp. 1640-1654.

Goldenberg, J., Oestreicher-Singer, G., and Reichman, S. 2012. "The Quest for Content: How User-Generated Links Can Facilitate Online Exploration," *Journal of Marketing Research* (49:4), pp. 452-468.

Goolsbee, A., and Chevalier, J.A. 2002. "Measuring Prices and Price Competition Online: Amazon and Barnes and Noble," *Quantitative Marketing and Economics* (1:2), pp. 203-222.

Granger, C.W.J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica* (37:3), pp. 424-438.

Gu, B., Konana, P., Rajagopalan, B., and Chen, H.W.M. 2007. "Competition among Virtual Communities and User Valuation: The Case of Investing-Related Communities," *Information Systems Research* (18:1), pp. 68-85.

Häubl, G., and Trifts, V. 2000. "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids," *Marketing Science* (19:1), pp. 4-21.

Healey, J.S., and Kassarjian, H.H. 1983. "Advertising Substantiation and Advertiser Response: A Content Analysis of Magazine Advertisements," *Journal of Marketing* (47:1), pp. 107-117.

Heckman, J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurememt* (5:4), pp. 475-492.

Heckman, J. 1979. "Sample Selection Bias as a Specification Error," *Econometrica* (47:1), pp. 153-161.

Heckman, J., Ichimura, H., and Todd, P. 1998. "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies* (65:2), pp. 261-294.

Henderson, J.M., and Quandt, R.E. 1958. *Microeconomic Theory: A Mathematical Approach*. New York: McGraw-Hill Book.

Herlocker, J.L., Konstan, J.A., Terveen, L.G., and Riedl, J.T. 2004. "Evaluating Collaborative Filtering Recommender Systems," *ACM Transactions on Information Systems* (22:1), pp. 5-53.

Huang, Z., Zeng, D.D., and Chen, H. 2007. "Analyzing Consumer-Product Graphs: Empirical Findings and Applications in Recommender Systems," *Management Science* (53:7), pp. 1146-1164.

Iaquinta, L., de Gemmis, M., Lops, P., Semeraro, G., Filannino, M., and Molino, P. 2008. "Introducing Serendipity in a Content-Based Recommender System," in *Proceedings of the Eighth International Conference on Hybrid Intelligent Systems*, Barcelona, Spain, pp. 168-173.

Indvik, L. 2012. "U.S. Online Retail Sales to Reach $327 Billion by 2016." Retrieved March 2, 2012, from http://mashable.com/2012/02/27/ecommerce-327-billion-2016-study/

Jehn, K.A., Northcraft, G.B., and Neale, M.A. 1999. "Why Differences Make a Difference: A Field Study of Diversity, Conflict and Performance in Workgroups," *Administrative Science Quarterly* (44:4), pp. 741-763.

Kahneman, D., and Tversky, A. 1979. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* (47:2), pp. 263-291.

Kamakura, W.A., Kim, B.D., and Lee, J. 1996. "Modeling Preference and Structural Heterogeneity in Consumer Choice," *Marketing Science* (15:2), pp. 152-172.

Katona, Z., and Sarvary, M. 2008. "Network Formation and the Structure of the Commercial World Wide Web," *Marketing Science* (27:5), pp. 764-778.

Kivetz, R., and Simonson, I. 2000. "The Effects of Incomplete Information on Consumer Choice," *Journal of Marketing Research* (37:4), pp. 427-448.

Knott, A.M., Posen, H.E., and Wu, B. 2009. "Spillover Asymmetry and Why It Matters," *Management Science* (55:3), pp. 373-388.

Kozinets, R.V. 2002. "The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities," *Journal of Marketing Research* (39:1), pp. 61-72.

Leeflang, P.S.H., and Parreño-Selva, J. 2012. "Cross-Category Demand Effects of Price Promotions," *Journal of the Academy of Marketing Science* (40:4), pp. 1-15.

Liu, Y. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing* (70:3), pp. 74-89.

Luca, M. 2011. "Reviews, Reputation, and Revenue: The Case of Yelp. Com," *Harvard Business School Working Paper* (12-016).

Manchanda, P., Ansari, A., and Gupta, S. 1999. "The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions," *Marketing Science* (18:2), pp. 95-114.

Manchanda, P., Xie, Y., and Youn, N. 2008. "The Role of Targeted Communication and Contagion in Product Adoption," *Marketing Science* (27:6), pp. 961-976.

Mayzlin, D. 2006. "Promotional Chat on the Internet," *Marketing Science* (25:2), pp. 155-163.

Mayzlin, D., and Yoganarasimhan, H. 2012. "Link to Success: How Blogs Build an Audience by Promoting Rivals," *Management Science* (58:9), pp. 1651-1668.

McKenna, K.Y.A., Green, A.S., and Gleason, M.E.J. 2002. "Relationship Formation on the Internet: What's the Big Attraction?," *Journal of Social Issues* (58:1), pp. 9-31.

Mishra, D.P., Heide, J.B., and Cort, S.G. 1998. "Information Asymmetry and Levels of Agency Relationships," *Journal of Marketing Research* (35:3), pp. 277-295.

Moe, W.W., and Schweidel, D.A. 2012. "Online Product Opinions: Incidence, Evaluation and Evolution," *Marketing Science* (31:3), pp. 372-386.

Moe, W.W., and Trusov, M. 2011. "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research* (48:3), pp. 444-456.

Mooney, R.J., and Roy, L. 2000. "Content-Based Book Recommending Using Learning for Text Categorization," in *Proceedings of the Fifth ACM Conference on Digital Libraries*, San Antonio, TX, pp. 195-204.

Mulhern, F.J., and Leone, R.P. 1991. "Implicit Price Bundling of Retail Products: A Multiproduct Approach to Maximizing Store Profitability," *Journal of Marketing* (55:4), pp. 63-76.

Mulpuru, S., Johnson, C., and Roberge, D. 2013. "US Online Retail Forecast, 2012 to 2017." Retrieved March 22, 2013, from http://www.forrester.com/rb/Research/us_online_retail_forecast%2C_2011_to_2016/q/id/60672/t/2

Muniz, A.M., and O'Guinn, T.C. 2001. "Brand Community," *Journal of Consumer Research* (27:4), pp. 412-432.

Nadkarni, S., and Gupta, R. 2007. "A Task-Based Model of Perceived Website Complexity," *MIS Quarterly* (31:3), pp. 501-524.

Nam, S., Manchanda, P., and Chintagunta, P.K. 2010. "The Effect of Signal Quality and Contiguous Word of Mouth on Customer Acquisition for a Video-on-Demand Service," *Marketing Science* (29:4), pp. 690-700.

Narayanan, S., Chintagunta, P.K., and Miravete, E.J. 2007. "The Role of Self Selection, Usage Uncertainty and Learning in the Demand for Local Telephone Service," *Quantitative Marketing and Economics* (5:1), pp. 1-34.

Nelson, P. 1970. "Information and Consumer Behavior," *Journal of Political Economy* (78:2), pp. 311-329.

Niraj, R., Padmanabhan, V., and Seetharaman, P.B. 2008. "A Cross-Category Model of Households' Incidence and Quantity Decisions," *Marketing Science* (27:2), pp. 225-235.

Obermiller, C., and Spangenberg, E.R. 1998. "Development of a Scale to Measure Consumer Skepticism toward Advertising," *Journal of Consumer Psychology* (7:2), pp. 159-186.

Oestreicher-Singer, G., and Sundararajan, A. 2012a. "Recommendation Networks and the Long Tail of Electronic Commerce," *MIS Quarterly* (36:1), pp. 65-83.

Oestreicher-Singer, G., and Sundararajan, A. 2012b. "The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets," *Management Science* (58:11), pp. 1963-1981.

Osborne, M. 2011. "Consumer Learning, Switching Costs, and Heterogeneity: A Structural Examination," *Quantitative Marketing and Economics* (9:1), pp. 25-70.

Parks, M.R., and Floyd, K. 1996. "Making Friends in Cyberspace," *Journal of Communication* (46:1), pp. 80-97.

Pathak, B., Garfinkel, R., Gopal, R.D., Venkatesan, R., and Yin, F. 2010. "Empirical Analysis of the Impact of Recommender Systems on Sales," *Journal of Management Information Systems* (27:2), pp. 159-188.

Pavlou, P.A., and Dimoka, A. 2006. "The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation," *Information Systems Research* (17:4), pp. 392-414.

Porter, C.E., and Donthu, N. 2008. "Cultivating Trust and Harvesting Value in Virtual Communities," *Management Science* (54:1), pp. 113-128.

Ratchford, B.T. 1982. "Cost-Benefit Models for Explaining Consumer Choice and Information Seeking Behavior," *Management Science* (28:2), pp. 197-212.

Rau, L.F., Jacobs, P.S., and Zernik, U. 1989. "Information Extraction and Text Summarization Using Linguistic Knowledge Acquisition," *Information Processing & Management* (25:4), pp. 419-428.

Reagans, R., and Zuckerman, E.W. 2001. "Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams," *Organization Science* (12:4), pp. 502-517.

Resnik, A., and Stern, B.L. 1977. "An Analysis of Information Content in Television Advertising," *Journal of Marketing* (41:1), pp. 50-53.

Ricci, F., Rokach, L., and Shapira, B. 2011. "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*. Boston, MA: Springer US, pp. 1-35.

Rosenbaum, P.R., and Rubin, D.B. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* (70:1), pp. 41-55.

Russo, J.E., and Chaxel, A.S. 2010. "How Persuasive Messages Can Influence Behavior without Awareness," *Journal of Consumer Psychology* (20:3), pp. 338-342.

Sahoo, N., Krishnan, R., Duncan, G., and Callan, J. 2011. "The Halo Effect in Multicomponent Ratings and Its Implications for Recommender Systems: The Case of Yahoo! Movies," *Information Systems Research* (23:1), pp. 231-246.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. 2000. "Analysis of Recommendation Algorithms for E-Commerce," in *Proceedings of the 2nd ACM Conference on Electronic Commerce*, Minneapolis, MN, pp. 158-167.

Schafer, J.B., Konstan, J., and Riedi, J. 1999. "Recommender Systems in E-Commerce," in *Proceedings of the First ACM Conference on Electronic Commerce*, Denver, CO, pp. 158-166.

Schafer, J.B., Konstan, J.A., and Riedl, J. 2001. "E-Commerce Recommendation Applications," *Data Mining and Knowledge Discovery* (5:1), pp. 115-153.

Schubert, P., and Ginsburg, M. 2000. "Virtual Communities of Transaction: The Role of Personalization in Electronic Commerce," *Electronic Markets* (10:1), pp. 45-55.

Seetharaman, P.B., Chib, S., Ainslie, A., Boatwright, P., Chan, T., Gupta, S., Mehta, N., Rao, V., and Strijnev, A. 2005. "Models of Multi-Category Choice Behavior," *Marketing Letters* (16:3), pp. 239-254.

Sethuraman, R. 1995. "A Meta-Analysis of National Brand and Store Brand Cross-Promotional Price Elasticities," *Marketing Letters* (6:4), pp. 275-286.

Sethuraman, R., and Srinivasan, V. 2002. "The Asymmetric Share Effect: An Empirical Generalization on Cross-Price Effects," *Journal of Marketing Research* (39:3), pp. 379-386.

Shocker, A.D., Bayus, B.L., and Kim, N. 2004. "Product Complements and Substitutes in the Real World: The Relevance of "Other Products"," *Journal of Marketing* (68:1), pp. 28-40.

Singh, V.P., Hansen, K.T., and Gupta, S. 2005. "Modeling Preferences for Common Attributes in Multicategory Brand Choice," *Journal of Marketing Research* (42:2), pp. 195-209.

Smith, N.C., and Cooper-Martin, E. 1997. "Ethics and Target Marketing: The Role of Product Harm and Consumer Vulnerability," *Journal of Marketing* (61:3), pp. 1-20.

Song, I., and Chintagunta, P.K. 2006. "Measuring Cross-Category Price Effects with Aggregate Store Data," *Management Science* (52:10), pp. 1594-1609.

Sonnier, G.P., McAlister, L., and Rutz, O.J. 2011. "A Dynamic Model of the Effect of Online Communications on Firm Sales," *Marketing Science* (30:4), pp. 702-716.

Srikant, R., and Agrawal, R. 1996. "Mining Quantitative Association Rules in Large Relational Tables," in *ACM SIGMOD International Conference on Management of Data*, New York, NY, pp. 1-12.

Sundararajan, A., Provost, F., Oestreicher-Singer, G., and Aral, S. 2013. "Information in Digital, Economic, and Social Networks," *Information Systems Research* (forthcoming).

Tam, K.Y., and Ho, S.Y. 2005. "Web Personalization as a Persuasion Strategy: An Elaboration Likelihood Model Perspective," *Information Systems Research* (16:3), pp. 271-291.

Tirunillai, S., and Tellis, G. 2012. "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance," *Marketing Science* (31:2), pp. 198-215.

Trusov, M., Bucklin, R.E., and Pauwels, K. 2009. "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing* (73:5), pp. 90-102.

Tsai, J.Y., Egelman, S., Cranor, L., and Acquisti, A. 2011. "The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study," *Information Systems Research* (22:2), pp. 254-268.

Tucker, C. 2011. "Network Stability, Network Externalities and Technology Adoption," *NBER Working Paper*.

Tumarkin, R., and Whitelaw, R.F. 2001. "News or Noise? Internet Postings and Stock Prices," *Financial Analysts Journal* (57:3), pp. 41-51.

Van Nierop, E., Fok, D., and Franses, P.H. 2008. "Interaction between Shelf Layout and Marketing Effectiveness and Its Impact on Optimizing Shelf Arrangements," *Marketing Science* (27:6), pp. 1065-1082.

Von der Fehr, N.H.M., and Stevik, K. 1998. "Persuasive Advertising and Product Differentiation," *Southern Economic Journal* (65:1), pp. 113-126.

Walters, R.G. 1991. "Assessing the Impact of Retail Price Promotions on Product Substitution, Complementary Purchase, and Interstore Sales Displacement," *Journal of Marketing* (55:April), pp. 17-28.

Wooldridge, J.M. 2006. *Introductory Econometrics: A Modern Approach*, (3rd ed.). Mason, OH: Thomson Southwestern.

Wu, C.C., Chen, Y.J., and Wang, C.J. 2009. "Is Persuasive Advertising Always Combative in a Distribution Channel?," *Marketing Science* (28:6), pp. 1157-1163.

Yang, S., and Allenby, G.M. 2003. "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research* (40:3), pp. 282-294.

Yi, P., Wenli, Z., and Gang, K. 2011. "A Hybrid Movie Recommender System," in *Twenty Third Workshop on Information Systems and Economics*, Shanghai.

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.

Zeng, X., and Wei, L. 2013. "Social Ties and User Content Generation: Evidence from Flickr," *Information Systems Research* (24:1), pp. 71-87.

Zhang, X. 2009. "Retailers' Multichannel and Price Advertising Strategies," *Marketing Science* (28:6), pp. 1080-1094.

Zhu, F., and Zhang, X. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing* (74:2), pp. 133-148.