경영학박사학위논문

The Impact of Dynamic Social Network Structure on Product Purchase

동적 소셜네트워크 구조가 제품 구매에 미치는 영향

2014년 8월

서울대학교 대학원

경영학과 경영학전공

이 희 태

The Impact of Dynamic Social Network Structure on Product Purchase

지도교수 김 병 도

이 논문을 경영학박사 학위 논문으로 제출함

2014년 4월

서울대학교 대학원

경영학과 경영학 전공

이 희 태

이희태의 경영학박사 학위논문을 인준함

2014년 6월

위 원 장 주 우 진 (인)

부위원장 김 상 훈 (인)

위 원 김 병 도 (인)

위 원 송 인 성 (인)

위 원 유 시 진 (인)

국문초록

온라인 소셜 미디어 매체를 통해 기업과 마케팅 연구자들은 고객이 소셜네트워크상에서 차지하고 있는 구조적 위치가 소비자 행동에 어떤 영향을 미치는가에 대해 많은 관심을 가져왔다. 소셜네트워크상의 위치 정보는 한 개인이 갖고 있는 사회적 자산을 나타낸다. 그러나 대부분의 연구에서 개인의 네트워크 위치 정보는 시간에 따라 변화하지 않는 (time-invariant or static) 변수로 받아들이고 있다. 소셜네트워크 데이터를 분석한 결과, 연결성과 근접 중심성 등과 같은 소셜네트워크의 위치 변수가 시간에 따라 상당히 변화하고 있다는 것을 발견하였으며, 그와 관련하여 이 연구에서 제기하고 있는 문제와 그 결과는 다음과 같다.

첫째, 기존 마케팅 연구와 소셜네트워크 분석 이론, 그리고 사회학 이론 등을 토대로 하여 네트워크 변수가 구매에 미치는 영향을 분석하였다. 그 결과 외향 연결 정도(out-degree)를 제외한 나머지 네트워크 위치 변수들 □ 내향 연결 정도 (in-degree), 내향 근접 중심성 (in-closeness centrality), 외향 근접 중심성 (out-closeness centrality), 제약성 (constaint) 과 군집 계수 (clustering coefficient) -가 문헌 연구에서 제시한 방향성과 유의성을 충족시켰다.

둘째, 시간에 따라 변화하는 네트워크 위치 변수를 적용한 모델과 기존 문헌에서 가정하고 있는 시간에 따라 변하지 않는 정적인 네트워크 위치 변수를 적용한 모델을 임의 효과 패널 토빗 (Random Effects Panel Tobit) 모형에 각각 적용하여 모형의 적합성을 비교하였다. 모형 비교 결과 동적 성향의 네트워크 변수를 적용한 모델이 그렇지 않은 모델보다 더 우수한 성과를 내는 것으로 나타났다. 또한 동적 (dynamic) 인 네트워크 변수를 설정하지 않고 정적 (static) 인 네트워크 변수를 설정하여 적용할 경우 네트워크 변수의 계수에 편의 (bias) 가 있을 가능성이 크다는 점도 밝혔다. 또한 네트워크의 동적인 성향으로 인해 네트워크 변수에 대한 계수가 시간에 따라 변화하는가를 판단하는 연구를 다층 패널 임의 계수 모형(Multi-level Panel Random Effects Model)을 통해 실행하였다. 그 결과 중개성 (brokerage) 을 측정하는 제약성을 제외한 나머지 네트워크 변수의 계수가 시간에 따라 이질적 (heterogeneous) 이라는 사실을 우도비 검정 (Likelihood Ratio Test) 을 통해 밝혔다.

또한, 두 번째 연구 결과를 바탕으로, 동적인 성향이 작은 제약성을 제외한 나머지 네트워크 변수와 구매 변수를 모두 종속 변수로 하고 그들의 시차 변수(lagged variables)들을 독립변수로 하는 패널 벡터 자기 상관 회귀 모형(panel Vector Autoregression, PVAR) 을 추정하였다. 추정 결과 네트워크 위치 변수의 시차 변수 (lagged variable) 가 현재의 제품 구매에 유의한 영향을 미치고 있음을 발견하였다. 그리고 충격 반응 함수 (Impulse Response Functions) 분석 결과도 마찬가지로 현재 네트워크 변수에 충격이 가해질 경우 내/외향 연결성과 내/외향 근접 중심성이 미래의 구매 가치에 지속적으로 유의하게 영향을 미치는 것으로 나타났다. 반면, 군집 계수는 충격이 가해진 직후에 구매에 긍정적인 영향을 미치지만 이후로는 미래의 구매 가치에 부정적인 영향을 미치는 것을 발견했다. 이를 통해 네트워크 위치 변수가 구매에 지속적인 이월효과 (Carryover effects) 를 나타낸다는 사실을 발견했다. 그리고 네트워크 변수 중 내/외향 근접 중심성에 주는 충격이 미래의 구매 가치에 가장 큰 영향을 준다는 사실을 밝혔다.

또한 구매 변수의 시차 변수를 독립변수로 하는 경우, 과거의 구매 변수는 내/외향 근접 중심성에만 긍정적인 영향을 주는 것으로 나타났다. 충격반응함수 분석에서도 동일한 결과가 나타났는데 현재 구매 변수에 충격을 가하면 내/외향 근접 중심성에 유의하게 긍정적인 영향을 지속적으로 주는 것으로 나타났다.

이 연구의 주요 이론적 및 실무적 시사점은 다음과 같다. 첫째, 기존 연구에서 간과하였던 동적 네트워크 변수를 소비자 구매 모형에 적용하였다. 예를 들어, 소셜네트워크의 연결성으로부터 도출할 수 있는 '허브' 가 될 수 있는 행위자는 시간에 따라 지속적으로 변화한다. 따라서 이러한 정보를 소비자 행동 모형에 적절히 반영하지 않을 경우 편의가 있는 모형을 추정할 가능성이 커지게 된다.

둘째, 네트워크 변수와 소비자 행동 변수 간에 내생성이 있기 때문에 이를 적절히 고려할 수 있는 모형이 필요하다. 특히, 근접중심성의 경우 구매 변수와 상호 영향력이 상당히 크다는 것을 알 수 있었으며 다른 네트워크 변수의 경우 시차 구매 변수의 영향이 유의하지 않은 것으로 나타났다.

그 동안 연구에서는 연결정도 (degree) 와 소비자 행동 변수 간의 관계를 규명하는데 초점을 두었으나 이 연구 결과, 근접 중심성이 가장 중요한 네트워크 변수임을 알아냈다. 즉, 흥미롭게도 소셜네트워크가 Watts and Strogatz (1998) 의 좁은 세상 (Small World) 이 될수록 소비자 행동에 긍정적인 영향을 준다는 점을 발견했다. 소셜네트워크상에서 소비자간 거리가 줄어들수록, 즉 근접 중심성이 높은 고객일수록 소셜네트워크 상에서 아이템과 음악을 더 구매하고 그러한 구매는 다른 사람들과 그들의 거리를 더욱 줄어들게 한다. 그런데 근접중심성은 군집계수를 낮출수록 증가하며, 군집계수를 낮출 수 있는

방법은 새로운 친구를 알 수 있게 할수록 낮아진다. 따라서 기업에서는 군집도가 높은 고객들을 대상으로 페이스북의 '알 수도 있는 친구' 와 같이 신규 친구 추천 서비스를 적극 진행하는 것도 좋은 전략이라고 볼 수 있다.

주요어: 소셜네트워크, 동적 네트워크 위치 변수, 임의 효과 패널 토빗, 시간에 따라, 변화하는 계수, 패널 벡터 자기 상관 회귀 모형, 내생성

Contents

TABLES

FIGURES

Ⅰ. Introduction

There has been a growing research interest in relationships between consumer behavior and social network topological positions (Hartmann et al. 2008; Oestreicher-Singer and Sundararajan 2012a; Oestreicher-Singer and Sundararajan 2012b; Oestreicher-Singer et al. 2013; Van Den Bulte and Wuyts 2007). The empirical marketing research on social networks mainly lies in identifying central nodes (i.e., hubs) and their roles in product adoption or diffusion (Goldenberg et al. 2009; Iyengar, Van den Bulte, and Valente 2011; Katona, Zubcsek, and Sarvary 2011; Trusov, Bodapati, and Bucklin 2010) and in identifying brokers and their roles in product purchase (Lee and Kim, 2013). However, such marketing studies on social networks have focused on static network structure and consumer behavior.[1]

Braha and Bar-Yam (2006) demonstrated that network structure varies significantly over time. The authors state, "The static topology does not capture the dynamics of social networks…. Our conclusions are in sharp contrast to previous complex network research, which emphasizes the importance of aggregate nodal centrality in a static network topology (Braha and Bar-Yam 2006, p. 5)." FIGURE 1-1 shows that degree, which is one of

[1] For example, Yoganarasimhan (2012) estimated a dynamic model. However, he assumed network variables were time-invariant.

the network position variables measuring actors' influential power, fluctuates violently. Therefore, if such a dynamic network effect on an individual's behavior were ignored, the parameters of those models applying static network variables may be underestimated and biased.

There is another issue with the view that network positions are static over time. Network position variables are assumed to be exogenous in the previous marketing research on social network analysis, because network variables are not considered dynamic over time and are thus considered given conditions. For example, in the relationship between product adoption and network position variables (e.g., Iyengar, Van den Bulte, and Valente 2011; Katona, Zubcsek, and Sarvary 2011), it is assumed that network variables cause adoption, but not vice versa. However, product adoption can change a person's network position. In other words, there could be potential endogeneity between his or her social network position and consumer activities such as product adoption.

To address these issues, we intend to investigate 1) whether network topological variables are dynamic over time and the model applying dynamic network position variables is superior to the model using static network variables and 2) whether the coefficients of network variable positions have time heterogeneity (time-varying parameters). Our third primary research interest is to examine 3) whether there are endogenous

relationships between consumer activities and social network position variables in addition to carryover effects among variables.

In sum, this paper makes three major contributions. First, this research investigates relationships between product (online item) purchase and network position variables based on the related previous research. Second, through this research, we can determine how much information that time-varying network position variables entail could improve the model fit and decrease prediction error. Third, this paper enables us to find endogeneity, dynamic response, and interactions between product purchase and network position variables.

The remainder of this paper is organized as follows. Chapter 2 introduces previous literature on network position variables and sociology theory. In Chapter 3, we explain and discuss the model specifications. We describe how the data are collected and how the network position variables are calculated in Chapter 4, and Chapter 5 details the estimation and empirical results. Chapter 6 concludes this paper with contributions and presents limitations and further research directions.

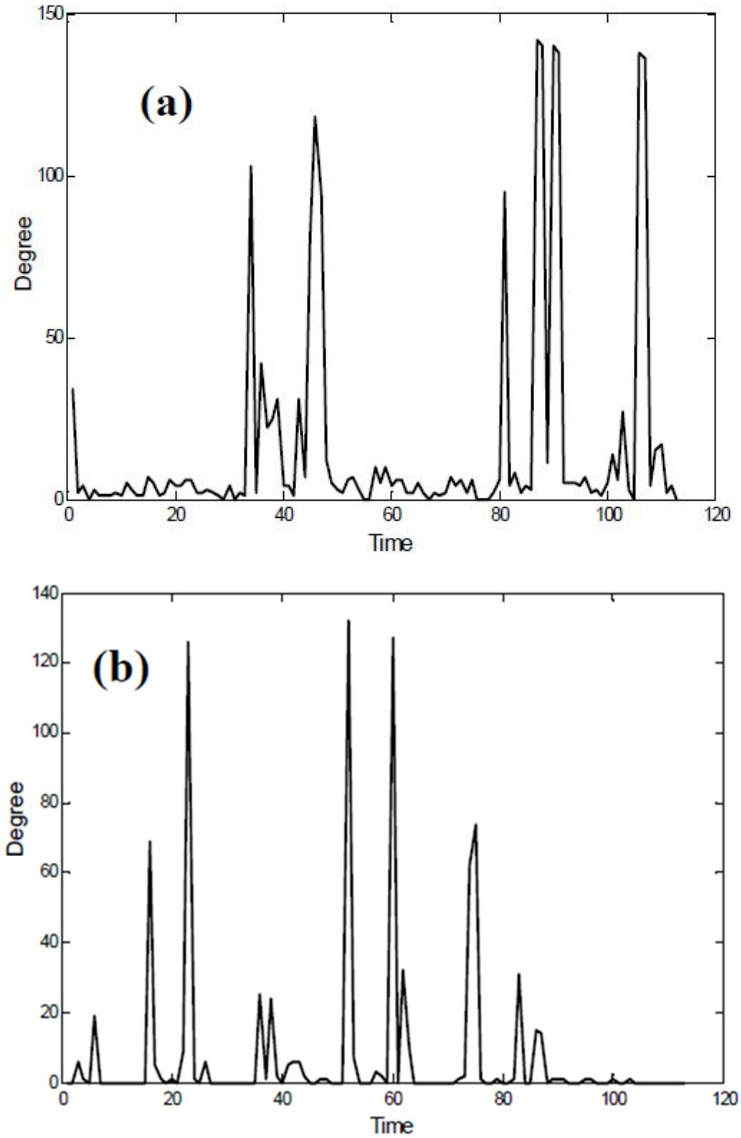


FIGURE 1-1. Degree Variations of Hub over Time (daily).

Source: Braha and Bar-Yam (2006)

Ⅱ. Related Literature

The main interest of marketing research on social networks is to examine the influential consumers (e.g., hubs, opinion leaders, lead users, market mavens) in social networks and how they affect their neighbors' and their own consumer behaviors and firm profits. Our paper relates to a large body of literature on social networks and social interaction from a wide variety of disciplines, including economics, marketing, and sociology. In this chapter, we summarize the network topological variables that are mainly used in marketing and other related research fields such as sociology to identify influential actors in social networks and present expected relationships between purchase and network position variables.

2.1. Degree - Hub

The degree is the number of lines showing how many relationships a node or an actor has in his or her social network (Wasserman and Faust, 1994). If a social network type is directed (e.g., communication network[2]), "degree"

[2] On the other hand, a friendship network is an undirected network.

can be categorized into “in-degree” and “out-degree.” From a sociometric point of view, actors or nodes who have even higher degrees than others is called a hub.[3] For example, power Twitterians who have huge numbers of followers and power bloggers are called “hubs” in social media. Hubs have been a main topic and received research interest from various academic spheres related with social networks. Marketing research has also investigated hubs as follows.

Watts and Dodds (2007) investigated the “roles of hubs” in diffusing information. They claim that hubs play restricted roles, and the critical mass of early information adopters is a main key in early diffusion. On the other hand, Goldenberg et al. (2009) empirically proved that hubs play critical roles in the diffusion and adoption of information. They also split hubs into innovative ones that influence the speed of adoption and follower ones that affect market size.

Stephen and Toubia (2010) studied the effects of social network structure positions on the profitability of online sellers in a social commerce platform. Each store formed a large network via links connected to other online stores. The authors found that forming networks among online sellers increased sellers’ revenue. In addition, the authors found that, of the network position variables, in-degree had a significant and positive effect on online sellers’

[3] Goldenberg et al. (2009) define hubs as people with an exceptionally large number of ties to other people (3 standard deviations above the mean).

revenues but out-degree had a significant and negative impact.

Katona, Zubcsek, and Sarvary (2011) empirically proved that while hubs have a higher probability of adopting products or information, having hubs as neighbors lowers the probability of adoption. In other words, having many friends is positively associated with adoption but has a negative impact on neighbors' adoption. Thus, they emphasized that it is important to consider network characteristics of a node's neighbors as well as his or her own network characteristics, which is the main difference of their study from previous related studies.

2.2. Closeness Centrality - Proximity

In social network analysis, "closeness centrality" can measure the "proximity" between a focal node and other nodes in social networks. Closeness shows the shortest distance between actors and can be inversely proportional to distance. The variable concentrates on how close a node or an actor is to all the other nodes or actors in social networks (Wasserman and Faust 1994). From the idea, nodes are central if they are quick to interact with others in their neighborhood.

Nodes that are central in terms of closeness can be efficient if they communicate information (e.g., product information) with other people,

which can influence their activities, such as product adoption (Beauchamp 1965). For example, Stephen and Toubia (2010) found that in-closeness centrality has a strong association with the revenues of online sellers. Thus, it can be expected that closeness centrality (in-closeness centrality and out-closeness centrality) positively affects product purchase.

2.3. Constraint - Brokerage

Constraint is a measure calculating the degree of brokerage (See Chapter 4.2 for a detailed explanation and notation of constraint). Burt (1992, 2004) defined a broker using the concept of a "structural hole" and his or her roles in a social network. He explained that if links between actors or groups are broken or become very weak when an actor or node is removed from a social network, then he or she is likely to be a broker. He also explained that the hole that brokers take is called a "structural hole." Brokerage is a concept that is closely related to "the strength of a weak tie" (Burt 1992; Granovetter 1973). By connecting weak ties, brokers can generate and strengthen "bridging social capital." Granovetter (1973, 1982) stressed that a broker or a bridge plays important roles in distributing information on social networks. According to his studies, brokers diffuse information by shortening the distance among actors. Moreover, he asserted that brokers are

more likely to act as sources of necessary and fresh information.

According to the current research on social networks, brokers play two critical roles in networks. First, they not only deliver fresh information but also change their own behavior. In other words, they tend to be open to new ideas, because they are able to get new information from various sources. Valente and Fujimoto (2010) explain that while hubs having a high degree tend to refuse to accept change and maintain the status quo to keep the public's interest (Becker 1970; Carcian 1979), brokers tend to be open to change and easily persuaded by other people.

From such discussions on brokerage, we can infer that if an actor or node is located in a place for brokers (structural hole), it is probable that he or she is more willing to accept new information and change his or her actions.

2.4. Clustering Coefficient – Density

The clustering coefficient represents the degree to which actors or nodes are interlinked. The density of connections in a network is termed clustering (Watts and Strogatz 1998). Networks that are highly clustered are generally close-knit and distinctive communities (Girvan and Newman 2002). FIGURE 2-1 shows a highly clustered network (a) and a network with low clustering (b).

This measure might be related to circumstances in which social network members have an impact on another actor. According to network closure theory (Burt 2004; Coleman 1988), if two nodes or actors who know the same person are also acquaintances, they are likely to exert a greater influence over that person than if they were unconnected.

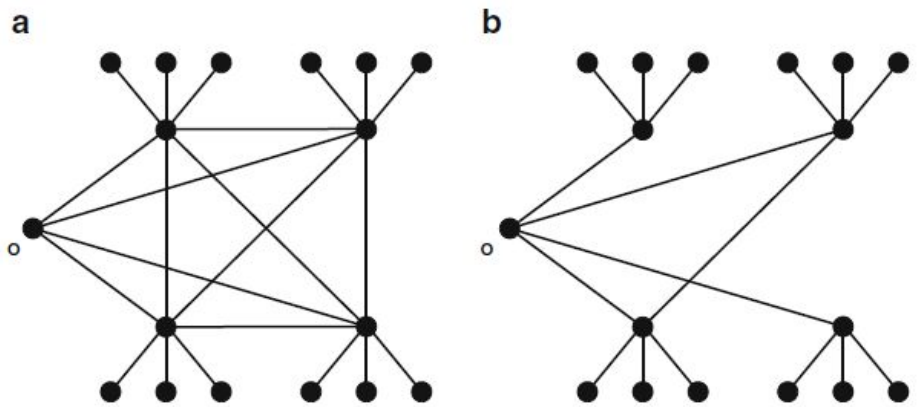


FIGURE 2-1. Networks with High (a) and Low (b) Clustering.
Source: Yoganarasimhan (2012).

For example, if a potential consumer is considering the same purchased product from two acquaintances, the attractiveness of the item will be strengthened if they are also friends with each other. For that reason, the density of relationships among adopted friends may influence the adoption likelihood of potential adopters.

Thus, we expect that clusteredness is positively associated with adoption probability. A higher clustering coefficient represents more powerful

relationships and "bonding social capital[4]" (Katona, Zubcsek, and Sarvary 2011).

Pak, Lee, and Lee (2012) found that people who belong to bonding social capital prefer symbolic products to utilitarian products, because they focus on other peoples' judgments or opinions of themselves. Online items and music, whose purchase quantity we are mainly interested in, are symbolic products. In this context, we can expect that clusteredness is positively correlated with item and music adoption.

TABLE 2-1 presents a summary of all network position variables and the expected sign and impact on purchase.

TABLE 2-1. Summary of Network Variables.

[4] In contrast with bonding social capital which shows "strong ties within groups," bridging social capital represents "weak ties across groups."

Network Variable	Definition/Measure	Expected Sign of coefficients on purchase
wind	Weighted In-degree, Ingoing Influential Power	+
woutd	Weighted Out-degree, Outgoing Influential Power	+
winclose	Weighted In-closeness Centrality, degree of closeness of distance from other nodes to a node	+
woutclose	Weighted Out-closeness Centrality, degree of closeness of distance from a node to other nodes	+
constraint	Constraint is inversely related to degree of brokerage	-
wclustering	Weighted Clustering, degree of nodes' clustering together	+

Sources: Huh (2010), Kim (2003), Wasserman and Faust (1994)

2.5. Modeling Issues

To investigate marketing studies on social networks, marketing researchers have used various technical methodologies, such as simulation (Choi, Kim, and Lee 2008; Watts and Dodds 2007) and empirical methods (Goldenberg et al. 2009; Iyengar, Van den Bulte, and Valente 2011; Katona, Zubcsek, and Sarvary 2011; Trusov, Bodapati, and Bucklin 2010).

TABLE 2-2 shows that previous marketing studies on social networks have applied static and dynamic models with time-invariant network variables. A major reason for such a static network variable specification is the limitation of data. The existing studies have mainly used static network

data, such as friendship network data (Katona, Zubcsek, and Sarvary 2011; Yoganarasimhan 2012). Friendship networks are undirected networks, and thus, only binary relationship information exists in undirected friendship networks that do not contain any longitudinal or dynamic network structure information (Katona, Zubcsek, and Sarvary 2011; Yoganarasimhan 2012). Although Stephen and Toubia (2010) used a directed network variable that can be estimated longitudinally, they calculated network position variables just once and concluded that these network variables are static or time-invariant.

As Braha and Bar-Yam (2006) pointed out, network variables could vary over time. Moreover, if network structural variables are dynamic over time, there might be endogenous relationships between network variables and consumer behaviors (product purchase in this paper). Therefore, we calculated time-varying (monthly) position variables and used them in estimating our model. In sum, we intend to deal with the aforementioned issues through a Random Effects Panel Tobit (REPT) model and a Panel Vector Auto Regression (PVAR) model using dynamic network position variables.

TABLE 2-2. Modeling Specifications of Previous and the Current Research.

Paper	Model	Time-varying (TV) or Time-invariant (TI) Network Variables	Exogeneity (X) or Endogeneity (N) of Network Variables
Watts and Dodds (2007)	Simulation	TI	X
Katona, Zubcsek, and Sarvary (2011)	Complementary Log-Log	TI	X
Hinz et al. (2011)	Random Coefficient Model	TI	X
Lee and Kim (2013)	OLS, Complementary Log-Log	TI	X
Goldenberg et al. (2009)	OLS, Agent-Based Model	TI	X
Stephen and Toubia (2010)	Tobit	TI	X
Iyengar, Van den Bulte, and Valente (2011)	Discrete-Time Hazard	TI	X
Oestreicher-Singer and Sundararajan (2012a)	OLS	TI	X
Oestreicher-Singer and Sundararajan (2012b)	OLS	TI	X
Yoganarasimhan (2012)	Dynamic Panel Model	TI	X
Current paper	**Random Effects Panel Tobit Model**	**TV**	**X**
	Panel Vector Auto Regression Model	**TV**	**N**

Ⅲ. Model Specification

In this study, we use a panel dataset that has both cross-sectional and time-series characteristics. Panel datasets for economic research possess several major advantages over conventional cross-sectional or time-series datasets (Min and Choi 2012). Panel data usually give a large number of data points, increasing the degrees of freedom and reducing the collinearity problem between explanatory variables—thus, improving the efficiency of econometric estimates. Panel data can control for unobserved individual heterogeneity factors. While cross-sectional datasets can estimate static relationships among variables, panel data set can estimate *dynamic* relationships.

To consider dynamic structure of network variables of our dataset, we specify panel data model and PVAR model additionally to examine endogenous interactions between structural network variables and a dependent variable.

3.1. Random Effects Panel Tobit Model (REPT)

In our dataset, for a large portion of the dependent variable, log of purchase is not observed (shown as 0) or left censored. So we specified REPT Model, which we would estimate is as follows (Min and Choi, 2012):

(1) $$y_{it}^{*} = \beta_0 + NP_{it} \cdot \beta_1 + X_{it} \cdot \beta_2 + W_i \cdot \beta_3 + u_i + e_{it} \qquad i = 1, 2, 3, \ldots, N, \quad t = 1, 2, 3, \ldots, T_i$$

$$u_i \sim N(0, \sigma_u^2) \text{ (if } u_i \text{ is random effects)}, \qquad e_{it} \sim N(0, \sigma_e^2)$$

$$y_{it} = \begin{cases} y_{it}^{*}, & y_{it}^{*} > 0 \\ 0, & y_{it}^{*} \le 0 \end{cases}$$

where NP_{it} is a time-varying network position variable vector, and X_{it} represents a time-varying individual variable vector (e.g. homepage visit frequency). W_i is a time- invariant individual characteristic variable vector (e.g. gender), β_0 is a constant scalar, and β_1 β_3 are coefficient parameter vectors to be estimated. σ_u^2 and σ_e^2 are also variances to be estimated. u_i indicates whether individual heterogeneity can be considered in terms of random effects or fixed effects.

3.2. Panel Vector Auto Regression (PVAR) Model

To capture endogeneity, dynamic relationship, and interactions between product purchase and network position variables, we use a PVAR model.

This model integrates the typical VAR methodology, which deals with all the variables in the dynamic system as endogenous, with the panel data, which can control for unobserved individual heterogeneity. We specify a panel VAR model as follows:

$$(2) \quad y_{it} = \sum_{l=1}^{L} A_l y_{i,t-l} + f_i + d_t + e_{tt} \quad e_{it} \overset{i.i.d}{\sim} (0, \Sigma_e)$$
$$i = 1, 2, 3, \ldots, N \ , \quad t = 1, 2, 3, \ldots, T_i$$

where y_{it} is an endogenous dependent variable vector, A is a parameter matrix to be estimated, f_i specifies unobserved heterogeneity, and d_t means time effect. e_{it} indicates an error term vector and needs to meet an orthogonal condition as follows:

$$(3) \quad E[y_{is} e_{it}] = E[f_i e_{it}] = E[d_t e_{it}] = 0$$

This orthogonal condition indicates that lagged variables are qualified as instrumental variables in estimating equation (2). However, unobserved heterogeneity, f_i, should be calculated to estimate parameters of equation (2) by using the orthogonality condition in equation (3).

In applying the VAR procedure to panel data, we need to impose the restriction that the underlying structure is the same for each cross-sectional unit. Since this constraint is likely to be violated in practice, one way to

overcome the restriction on parameters is to allow for "individual heterogeneity" in the levels of the variables by introducing fixed effects, denoted by f_i in the model. If the fixed effects were correlated with the regressors due to lags of the dependent variables, the mean-differencing procedure commonly used to eliminate fixed effects would create biased coefficients. To avoid this problem, we use forward mean-differencing, also referred to as the "Helmert procedure" (see Arellano and Bover (1995) for details). This procedure removes only the forward mean — that is, the mean of all the future observations available for each firm-year (Love and Zicchino 2006). This transformation preserves the orthogonality between transformed variables and lagged regressors, so we can use lagged regressors as instruments and estimate the coefficients by system generalized method of moment (GMM). Our model also allows for individual-specific time dummies, d_t , which can be eliminated by subtracting the means of each variable calculated for each month. As in traditional VAR, PVAR allows us to treat all variables as endogenous, but PVAR also allows estimation for multiple cross sections of data — something not possible in traditional VAR.

The PVAR analysis is supplemented with the analysis of impulse response functions (IRFs) to elucidate the dynamics in the relationships of interest. IRFs show the response of one variable to an exogenous shock (i.e., a one

standard deviation shock) to another variable in the system, while holding all other shocks at zero. Using IRFs, it is possible to visualize the dynamics of the pairwise relationships. In other words, we can isolate the reaction of product purchase to a network position variable shock while other variables remain constant.

Ⅳ. Data

4.1. Data Collection

We obtained data from a social network service in Korea. Data period is 1 year, from Oct. 2011 to Sep. 2012. By snowball sampling, 23,395 nodes were selected as our dataset, including their individual characteristic variables. The network type is a communication network that considers 1) visiting frequency between members (weighted) and 2) the direction of relationships (directed).

The dependent variable is purchase quantities of items or music, which are each generally used to decorate members' individual homepages. We conducted log-transformation of the purchase data. To control for individual characteristics, we used demographic information such as gender and age, as well individual homepage login frequency, gift shop visit frequency, and number of friends, which are aggregated as monthly data.

To set independent variables, we computed network topology variables as per the following chapter. Total possible network observations are 23395*12=28,0740 if missing values are not considered.

4.2. Network Position Variables

We calculated network topology variables as follow. Unlike the previous research (Katona, Zubcsek, and Sarvary 2011; Stephen and Toubia 2010; Yoganarasimhan 2012), we calculated network position variables as time-varying, which is our main differentiation point from other studies.

4.2.1. Degree

Weighted in-degree and out-degree are computed as in equations (4) and (5):

(4) $$Wind_{it} = \sum_{j=1}^{N} W_{ij,t} \cdot L_{ij,t}$$

(5) $$Woutd_{it} = \sum_{j=1}^{N} W_{ji,t} \cdot L_{ji,t}$$

$Wind_{it}$ and $Woutd_{it}$ represent weighted in-degree and out-degree, respectively, of node i in period t. $W_{ij,t}$ means a number of relationships from j to i in period t, $W_{ji,t}$ is a number of relationships from i to j in period t. $L_{ji,t}$ if a relationship from i to j exists, 1 or 0 in period t and $L_{ij,t}$ if a relationship from j to i exists, 1 or 0 in period t.

4.2.2. Constraint

A measure that estimates a degree of brokerage of nodes is related to

"Structural Hole" theory (Burt 1992). Although there is no measure capturing structural hole directly, brokerage can be measured by "constraint." According to Burt (1992), we calculated a "constraint" variable that is negatively correlated with brokerage, which means that the lower is a degree of constraint of a node or an actor, the higher the possibility that he or she becomes a broker. Constraint can be measured as equation (7):

(7) $$Constraint_{it} = \sum_{j} (p_{ij,t} + \sum_{h=1}^{N} p_{ih,t} \cdot p_{hj,t})^2, h \neq i, j$$

where $p_{ih,t}$ signifies how much time and energy i exerts to h out of i's acquaintances as the following equation:

(8) $$p_{ih,t} = \frac{(W_{ih,t} + W_{ih,t})}{\sum_{j} (W_{ij,t} + W_{ji,t})}, i \neq j$$

For more details on constraint, see Burt (1992).

4.2.3. Closeness Centrality

Weighted in-closeness centrality and weighted out-closeness centrality can be calculated as follow:

(6) $$Winclose_{it} = \sum_{i \neq j} W_{ij,t} / d_{ij,t}, \quad i = 1,2,3,...,N$$

$$Woutclose_{it} = \sum_{j \neq i} W_{ji,t} / d_{ji,t}, \quad 22 = 1,2,3,...,N$$

(7)

$\sum_{i \neq j} d_{ij,t}$ is minimum distance from j to i (i ≠ j) in period t, and

$\sum_{j \neq i} d_{ji,t}$ shows minimum distance from i to j (j ≠ i) in period t

4.2.4. Clustering Coefficient

A clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. A weighted clustering coefficient can be calculated using the following procedure.

A graph $G(V, E)$ in which V represents a set of vertices (nodes) $V = \{v_1, v_2, ...\}$ And E shows a set of edges (links) $E = \{e_1, e_2, ...\}$ as the following:

(9) $$e_{ij,t} = \begin{cases} 1, if\,(i,j) \in E \\ 0, otherwise \end{cases}$$

In addition, if we assume a set of neighbors of i in period t as the following:

(10) $$N_{i,t} = \{ j \,|\, i, j \in V, e_{i,j} = 1 \,\&\, (i,j) \in E \}$$

then, weighted clustering coefficient can be measured as the following (11) equation:

$$Wclustering_{i,t} = \frac{\sum_{i,j \in N_{i,t}} e_{ij,t} \cdot \frac{W_{ij,t}}{\sum_{i,j \in N_{i,t}} W_{ij,t}}}{\sum_{i,j \in N_{i,t}} 1}$$

(II)

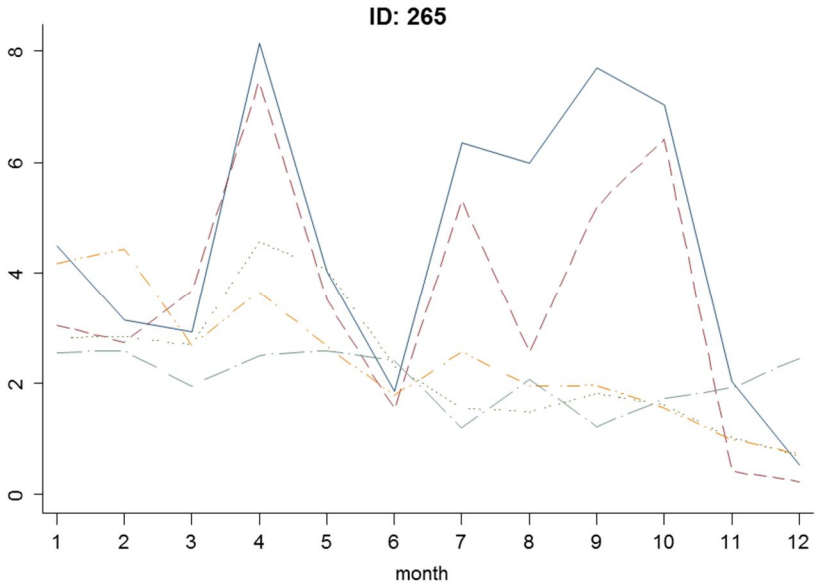
Ⅳ. Empirical Results

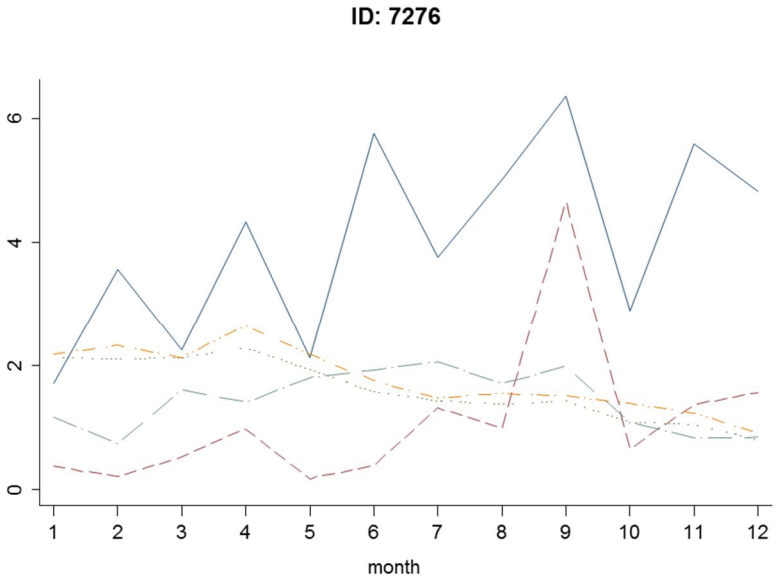
5.1. Descriptive Statistics

The descriptive statistics are shown in TABLE 5-1. All network position variables have substantial standard deviations within (over time) compared to standard deviations between individuals, which shows that there exist considerable variations in each network variable over time. To get additional intuition, we selected three panels and investigated variations of their network position variables. FIGURE 5-1 also shows there are considerable variations in network position variables over time.

TABLE 5-1. Descriptive Statistics

Variable	Variable Definition	Obs.	Mean	Min.	Max.	Std. Dev		
						overall	between	within
wind	weighted in-degree	206,072	118.73	0	65,011	601.43	441.15	**323.51**
woutd	weighted out-degree	206,072	118.73	0	14,082	316.68	216.68	**204.10**
winclose	weighted in-closeness centrality	206,072	0.00	0	0.0069	0.00080	0.00049	**0.00065**
woutclose	weighted out-closeness centrality	206,072	0.00	0	0.0147	0.00085	0.00059	**0.00063**
constraint	brokerage measure	206,072	0.60	0.01	2	0.46	0.37	**0.30**
wclustering	Network density	127,646	0.0008	0	0.07	0.00250	0.00170	**0.00124**
friend	number of friends	278,732	165.86	0	23,237	268.73	246.98	104.36
vigiftshop	frequency of visiting gift shop	128,175	2.18	1	30	2.14	1.19	1.52
vihome	frequency of visiting individual homepage	268,249	19.78	1	31	10.24	7.73	6.77
du	months from registration date	23,395	68.51	5	161	27.19	23.12	3.61
age		23,395	18.77	11	59	5.32		
gender	male: 0, female: 1	23,395	0.54	0	1	0.50		


ID: 265
0
2
4
6
8
1
2
3
4
5
6
7
8
9
10
11
12
month


ID: 7276
0
2
4
6
1
2
3
4
5
6
7
8
9
10
11
12
month

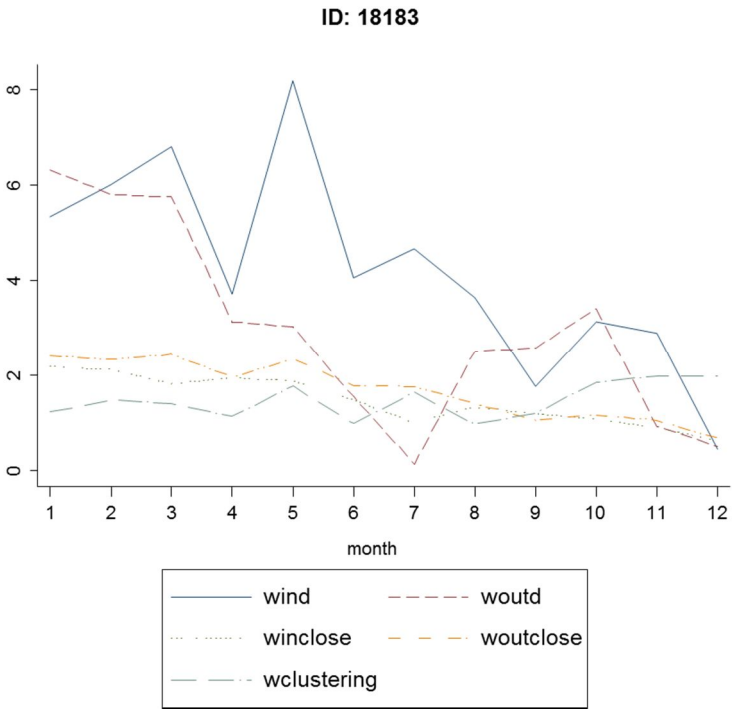


FIGURE 5-1. Network Variable Variations of three Panels

5.2. Multicollinearity Test

Multicollinearity might exist among independent variables. Thus, we conducted a multicollinearity test by showing correlations among variables as in TABLE 5-2 and variance influential factor (VIF) as in TABLE 5-3. If an absolute value of a correlation coefficient is over 0.7, VIF value is over 10, or tolerance is under 0.1, then it is considered that there exists multicollinearity (Dorman et al. 2013; Park 2007; Rawlings, Pantula, and Dickey 1998). The results suggest that there might be little multicollinearity among variables when applying those diagnostic criteria. In other words, all absolute pairwise correlation coefficients are under 0.7, every VIF is under 10, and tolerance is over 0.1. Thus, we used all network variables as independent variables in the analysis.

TABLE 5-2. Pairwise Correlation Matrix

Variable	Y	N1	N2	N3	N4	N5	N6	X1	X2	X3	X4	W1
ln_pu(Y)												
wind(N1)	.17*											
woutd(N2)	.23*	.27*										
winclose(N3)	.31*	.18*	.49*									
woutclose(N4)	.25*	.40*	.34*	.51*								
constraint(N5)	-.20*	-.16*	-.29*	-.56*	-.56*							
wclustering(N6)	.16*	.45*	.64*	.34*	.35*	-.28*						
friend(X1)	.08*	.26*	.01*	-.01*	.13*	-.07*	.04*					
vigiftshop(X2)	.40*	.13*	.27*	.26*	.17*	-.13*	.13*	.03*				
vihome(X3)	.31*	.12*	.20*	.28*	.27*	-.29*	.14*	.06*	.24*			
du(X4)	,-.04*	.00	-.08*	-.12*	-.02*	.10*	-.05*	.14*	-.12*	-.08*		
age(W1)	-.04*	.02*	-.04*	-.08*	-.02*	.09*	-.03*	.06*	-.07*	-.12*	.33*	
gender(W2)	.07*	.03*	.06*	.04*	.02*	-.05*	.03*	.01*	.13*	.13*	.05*	-.10*

TABLE 5-3. Collinearity Statistics

Variable	VIF	Tolerance
woutd	2.18	0.46
cluster	2.1	0.48
out-close	2.04	0.49
in-close	1.99	0.50
wind	1.63	0.62
constraint	1.26	0.79
vigiftshop	1.18	0.85
du	1.14	0.88
age	1.11	0.90
friend	1.11	0.90
vihome	1.1	0.91
gender	1.05	0.96

5.3. Model Comparison

First, we specified two different network position variables (dynamic vs. static). Then, we applied each specification to equation (1). One is the base model applying an assumption that network position variables are static over time, which has been applied in previous studies (e.g., Katona, Zubcsek, and Sarvary 2011; Stephen and Toubia 2010; Yoganarasimhan 2012). Static demonstrates that the *latest* network positions of nodes are not

different over time. The other is our proposed model, in which we assume that network variables are time varying. By comparing each differently specified model, we can compare the model fit and examine whether there might be parameter bias.

Then, we decided whether we should estimate the model using a random effects model or fixed effects model. Thus, in equation (1), the following null hypothesis should be tested:

H0: σ_u=0

If the null hypothesis is rejected, equation (1) should be estimated by random effects model, or pooled Tobit model otherwise. As in TABLE 5-4, sigma_u (σ_u)=0 hypothesis is rejected. Thus, REPT model was chosen as the model to be estimated. Rho (ρ) represents the portion of variance of u_i showing individual heterogeneity out of total variance. ρ can be calculated as follows:

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \tag{12}$$

In the proposed model, $\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} = \frac{2.13^2}{2.13^2 + 3.7^2} = .25$ means

that the variance of u_i explains 25% of total variance.

The results of the base model and the proposed model are presented in TABLE 5-4. In terms of model fit standards (log likelihood, AIC and BIC), the proposed model is superior to the base model. For example, BIC of the proposed model is lower (319,441) than that of the base model (330,205). Therefore, it can be concluded that the second research question ("Is the model applying dynamic network position variables superior to the model using static network variables?") can be answered positively.

Thus, we intend to present and analyze the results of the proposed model. First, in terms of individual characteristic variables, number of friends, gift shop visit frequency, individual homepage visit frequency, and membership duration each has significant and positive effects on purchase (all p-values < .001). With respect to demographics, female members tend to purchase more than male ones, and age has a positive effect on purchase (p-value < .001).

Results with network position variables are as follows. Except for weighted out-closeness centrality or woutclose and constraint (β=-.14, p-value>.05, however, marginally significant at 10% level), each network variable has a significant and positive impact on purchase (all p-values < .05). The more visits from his or her friends a node or an actor gets (weighted in-degree or wind), the more purchases he or she tends to make.

However, out-degree or woutd does not have a significant impact on purchase. Intuitively, it can be inferred that the more visits from the friends a node gets, the more might he or she intends to decorate his or her individual homepage by purchasing items or music. On the other hand, even if nodes or actors visit their friends' homepage more frequently, they would not purchase more items or music, which makes sense in that those nodes are mainly interested in their friends' homepage rather than their own.

Weighted in-closeness centrality (winclose) and weighted out-closeness centrality (woutclose) each have a positive and significant impact on purchase. As distance coming from friends or distance reaching friends shortens, he or she tends to make more purchases. In other words, nodes located in a central place for communicating information are more likely to buy items or music.

Constraint, which measures a degree of brokerage, has a negative but insignificant impact on purchase at the 5% significance level (however, marginally significant at 10% level). As in the <4.2. Network Position Variables> part, constraint has an inverse relationship with brokerage. If a node is a broker, he or she has more probability of making a purchase of items, though this effect is marginal.

In addition, weighted clustering or wclustering has a significant and positive effect on purchase, as expected (β =-.14, p-value<.05). As we

explained previously, clustering coefficient is strongly related to the concept of the bonding social capital. A high clustering coefficient means network members are tightly knit, thus linking to bonding each other. Pak et al. (2012) found that bonding social capital-type people would prefer symbolic products over utilitarian products because they are under tightly knit environments and focus on other peoples' judgments or opinions on themselves. Online items and music are symbolic products. For that reason, clustering coefficient affects purchase positively.

TABLE 5-4. The Results of Equation (1).

Variable	Base Model	Proposed Model
wind	.0002***(5.74)	.00009***(3.38)
woutd	-.0002***(-2.87)	.0002(.30)

winclose	.003**(2.32)	.41***(12.46)
woutclose	.017***(21.92)	.54***(18.72)
constraint	0.19*(1.86)	-.14*(-1.79)
wclustering	-.002(-.54)	.02**(2.07)
friend	0.0003***(6.88)	0.0003***(6.44)
vigiftshop	.66***(92.66)	.58***(74.76)
vihome	.11***(45.56)	.09***(34.70)
du	.005***(8.3)	.003***(3.17)
age	.02***(5.68)	.01***(2.77)
gender	.06*(1.76)	.17***(3.70)
constant	-3.39***(-33.69)	-3.24***(-26.7)
Log Likelihood	-165024	-159642
AIC	330076	319312
BIC	330205	319441
sigma_u	2.03	2.13
sigma_e	3.8	3.7
rho	.22	.25
Likelihood Ratio Test(sigma_u=0)	4014.3(.000)	6429.5(.000)
No. of obs.	75266	75266
No. of Individuals	17206	17206

Notes: 1) Dependent Variable □ log(purchase), ln_pu

2) *, **, *** - significant at the 10%, 5% and 1% level, respectively

3) Base Model fits equation (1) using a static network position specification,
Proposed Model fits equation (1) using a static network position specification.

TABLE 5-5 summarizes hypothesis test results.

TABLE 5-5. Hypothesis Test Results

Network Variable	Expected Sign	Actual Sign	Result
wind	+	+	Supported
woutd	+	+	reject
winclose	+	+	Supported
woutcloseness	+	+	Supported
constraint	-	-	Marginally Supported
wclustering	+	+	Supported

From the results of estimated coefficients, it can be assured that the proposed model performance is superior to that of the base model. Out of the estimated coefficients of network position variables, the significance and sign of weighted out-degree (woutd), constraint, and weighted clustering (wclustering) are different from the expected sign and significance based on the existing theory and empirical findings from the previous research. More specifically, in the base model, woutd has a significant and negative impact on purchase. Stephen and Toubia (2010) found that woutd has a significantly negative impact on online sellers' revenues. However, were network variables specified as static, the result might be biased.

With respect to weighted clustering (wclustering), the sign of the coefficient is negative and not significant. Yoganarasimhan (2012) expected that a high clustering coefficient was positively associated with the video

viewership diffusion. However, the result was a negative impact of clustering on viewership. It might also be a biased result because their network position variables are assumed static. Constraint has a marginally positive impact on purchase at the 10% significance level. However, we can infer that the result is biased under sociology theory (Burt 1992, 2004).

To corroborate that the random effects panel Tobit with dynamic network variable model (REPTD) is superior to the base model with static network variable model (REPTS)[5] in terms of performance, the in-sample and out-of-sample fit of REPTD and REPTS were compared. We used the final month's network variables as static network variables, which makes sense in that the previous research takes the latest network position as actors' given topology (Katona, Zubcsek, and Sarvary 2011; Stephen and Toubia 2010; Yoganarasimhan 2012).

The dataset was split into 13,395 panels for the estimation dataset and 10,000 panels for the holdout dataset. By using Equation (1), we fitted REPTD and REPTS in TABLE 5-6 with the estimation dataset. In terms of log-likelihood, AIC and BIC, REPTD demonstrates better fits over REPTS. Hence, it can be inferred that the, proposed, model applying dynamic network variables is superior to the model applying static network variables, confirming the results of TABLE 5-4.

[5] The specification of REPTS is the same as the one of base model in TABLE 5-4.

TABLE 5-6. Comparison of the model fit REPTS vs. REPTD

Variable	REPTS	REPTD
wind	.0003***(3.69)	.00008**(2.43)
woutd	-.0002**(-2.15)	-.00002(-0.26)
winclose	.005**(1.98)	.41***(9.44)
woutclose	.02***(10.86)	.54***(14.43)
constraint	.24(1.34)	-.14(-1.30)
wclustering	-.03(-.41)	.02*(1.83)
friend	.0003***(3.67)	0.0005***(5.82)
vigiftshop	.65***(64.91)	.59***(57.41)
vihome	.11***(30.6)	.09***(26.66)
age	.02***(3.9)	.02***(2.99)
gender	.11*(1.72)	.21***(3.28)
du	.004***(3.54)	.002*(1.74)
constant	-3.28***(-19.58)	-3.35***(-26.4)
Log Likelihood	-94,703	-91,664
AIC	189,433	183,357
BIC	189,555	183,488

Notes: 1) Dependent Variable □ log(purchase), ln_pu
2) *, **, *** - significant at the 10%, 5% and 1% level, respectively
3) REPTS: random effects panel Tobit model with static network variables
4) REPTD: random effects panel Tobit model with dynamic network variables

TABLE 5-7 compares the in-sample (estimation dataset) fit and out-of-sample (holdout dataset) fit of REPTS and REPTD models in terms of root mean squared error (RMSE) and mean absolute deviation (MAD). Both in sample and out-of-sample, the proposed REPTD model produces a lesser number both in terms of RMSE and MAD, indicating superior fit to the data over REPTS.

TABLE 5-7. In-Sample Fit and Out of Sample Fit Comparison

		AEPTS *	AEPTD **
In-Sample	AMSE	3.43	3.28
	MAD	1.74	1.63
Out-of-Sample	AMSE	3.63	3.49
	MAD	2.62	2.45

5.4. Dynamic Effects of Network Position Variables

From TABLE 5-1 and FIGURE 5-1, it can be inferred that there exist considerable variations in network variables. To examine whether network variable parameters are time- varying, we fitted a multilevel mixed-effects model as the following:

(13) $$y_{it} = \beta_0 + NP_{it} \cdot (\beta_1 + \lambda_t) + W_i \cdot \beta_2 + X_{it} \cdot \beta_3 + \mu_t + e_{it} \quad \mu_t \sim N(0, \sigma_\mu^2) \quad e_{it} \sim N(0, \sigma_e^2)$$

By equation (13), we can investigate whether the parameter vector of network position variables has random effects over time. In other words, we should test whether λ_t out of $\beta_1 + \lambda_t$ is random or not over time; if var (λ_t) is not zero, network variables have random effects over time. To examine whether each network coefficient is random or not over time, we fitted multi-level panel random intercept model as a base model and multi-level

panel random coefficient models in which each network position variable was specified as random as an alternative model. Then, we performed a likelihood-ratio test comparing the multi-level panel random intercept model as a restricted model) and each multi-level panel random coefficient model over time as an unrestricted model (Rabe-Hesketh and Skrondal 2008).

TABLE 5-8 shows the test results. The multi-level panel random intercept model was rejected in favor of all random coefficient models applying time-varying network position variables, except for the time-varying coefficient for constraint (Likelihood Ration Test Chi-square values are significant at 1% level except for constraint). From the results, we can conclude that network position variables have time-heterogeneous effects.

TABLE 5-8. The Results of Equation (13)

Variable	AC	AWIND	AWOUTD	AWINCLOSE	AWOUTCLOSE	ACONSTRAINT	AWCLUSTERING
wind	.00009***(5.90)	.0001***(4.74)	.00009***(5.87)	.00009***(5.64)	.00009***(5.91)	.00009***(5.9)	.00009***(6.01)
woutd	-.006(-.16)	.002(.07)	-.002(-.03)	-.01(-.32)	-.002(-.05)	-.006(-.16)	.002(.05)
winclose	.23***(8.79)	.23***(8.94)	.24***(9.2)	.28***(5.2)	.24***(9.2)	.23***(8.72)	.25***(9.66)
woutclose	.46***(22.68)	.47***(22.79)	.47***(22.98)	.48***(23.29)	.51***(11.86)	.46***(22.59)	.48***(23.3)
constraint	-.02(-.41)	-.02(-.37)	-.009(-.18)	.005(.11)	-.007(-.14)	-.03(-.4)	.007(.14)
wclustering	.00002(.00)	-.004(-.73)	-.0007(-.14)	-.0004(-.08)	-.003(-.49)	.0002(.03)	-.002(-.22)
friend	.0003***(11.55)	.0003***(10.98)	.0003***(11.53)	.0003***(11.58)	.0003***(11.11)	.0003***(11.55)	.0003***(11.41)
vigiftshop	.49***(96.25)	.48***(96.12)	.49***(96.26)	.49***(96.15)	.49***(96.12)	.49***(96.24)	.49***(95.97)
vihome	.07***(43.29)	.09***(43.2)	.07***(43.19)	.07***(43.0)	.07***(43.04)	.07***(43.31)	.07***(42.98)
du	.003***(6.1)	.003***(6.15)	.003***(6.74)	.003***(6.22)	.003***(6.21)	.003***(6.09)	.003***(6.26)
age	.01***(6.78)	.01***(6.7)	.01***(6.2)	.01***(6.63)	.01***(6.7)	.01***(6.78)	.01***(6.73)
gender	.07***(2.95)	.07***(2.93)	.07***(2.97)	.07***(2.96)	.07***(3.01)	.07***(2.93)	.07***(2.95)
constant	-96***(-8.52)	-.97***(-8.47)	-.99***(-8.54)	-1.03***(-7.48)	-1.02***(-7.82)	-96***(-8.42)	-1.02***(-8.89)
No. of obs.	75266	75266	75266	75266	75266	75266	75266
No. of time	12	12	12	12	12	12	12
Log Likelihood	-191525	-191515	-191507	-191490	-191513	-191524	-191497
AIC	383081	383062	383047	383013	383038	383081	383026
BIC	383220	383210	383195	383160	383185	383229	383174

LRT chi2	**Base**	21.6***	36.0***	70.5***	45.2***	2.1	56.6***

Notes: 1) Dependent Variable □ log(purchase), ln_pu

2) *, **, *** - significant at the 10%, 5% and 1% level, respectively

3) RC signifies multi-level panel random coefficient model just with constant random

4) RWIND, RWOUTD, RWINCLOSE, RWOUTCLOSE, RCONSTRAINT,

RWCLUSTERING are multi-level panel random coefficient models with both

constant and coefficient of network position variables random, respectively.

5.5. Panel Vector Auto Regression (PVAR) Model

To investigate endogeneity and dynamic response as well as interactions between product purchase and network position variables, equation (2) was estimated. We excluded the constraint variable for two reasons. First, based on the results of TABLE 5-8, there exists little time heterogeneity in the constraint variable. Second, when PVAR estimation was conducted using each network position variable, including constraint; no dynamic endogenous effects between constraint and purchase were found. Equation (2) can be presented as the following procedures using network variables and log of purchase.

To begin with, we conducted our empirical analysis by testing for stationarity versus evolution of variables (Enders 2004). To this end, we performed various panel unit root tests for variable stationarity, such as the tests of Levin and Lin (1992) and Im, Pesaran, and Shin (1997), ADF-Fisher and PP-Fisher. Results of these tests are shown in TABLE 5-9 and indicate that all of the variables are stationary (all p-values <.00). This shows that model estimations can be run with all variables in levels.

TABLE 5-9. Panel Unit Root Test Results

	Levin, Lin and Chu t*		Im, Pesaran and Shin W-stat		ADF		PP	
	test-statistic	p-value	test-statistic	p-value	test-statistic	p-value	test-statistic	p-value
ln_pu	-572	0.00	-248	0.00	110,585	0.00	126,325	0.00
wind	-4,280	0.00	-293	0.00	84,123	0.00	97,133	0.00
woutd	-10,695	0.00	-1,176	0.00	93,196	0.00	10,625	0.00
winclose	-4,970	0.00	-4,843	0.00	39,470	0.00	44,881	0.00
woutclose	-67	0.00	-655	0.00	42,396	0.00	46,130	0.00
wclustering	-6,067	0.00	-367	0.00	48,687	0.00	55,409	0.00

Next, we conducted Granger-causality tests. The results for these tests in are reported in TABLE 5-10 and demonstrate clear evidence of bi-directional causality in each pair of variables except for weighted in-degree or wind and weighted clustering coefficients or wclustering[6] at the 5% significance level. This supports our approach of analyzing the variables as a "full dynamic system" through PVAR analysis (Trusov, Bucklin, and Pauwels 2009).

TABLE 5-10. Granger Causality Results

DV is Granger Caused by	ln_pu	wind	woutd	winclose	woutclose	constraint
ln_pu		.000	.0101	.000	.000	.000
wind	.000		.000	.000	.000	.000
woutd	.000	.000		.000	.000	.000
winclose	.000	.00032	.000		.000	.000
woutclose	.000	.0065	.000	.000		.000
wclustering	.000	.0687	.0037	.000	.000	

Then, we choose the appropriate lag length L using Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) (for details, see Greene 2008), following the standard approach in the VAR literature (e.g., Holtz-Eakin, Newey, and Rosen 1988; Love and Zicchino 2006). The results are shown in TABLE 5-11. According to the results, lag3 has the lowest values in terms of AIC and BIC; thus, it is chosen as an optimal lag

[6] Wind(weighted in-degree) is not granger caused by constraint at 5% significant level

length.

TABLE 5-11. Optimal Lag Length

lag	AIC	BIC
lag1	5.12	5.14
lag2	5.08	5.14
lag3	5.02	5.13
lag4	5.05	5.16
lag5	5.08	5.18
lag6	5.21	5.30
lag7	5.43	5.51

Based on the above results, we conducted PVAR model using equation (2). In more detail, equation (2) can be presented as the following equation (13).

(13)

$$\begin{bmatrix} ln\_pu_{it} \\ wind_{it} \\ woutd_{it} \\ winclose_{it} \\ woutclose_{it} \\ wclustering_{it} \end{bmatrix} = \sum_{l=1}^{3} \begin{bmatrix} \beta_{11}^{l} & \beta_{12}^{l} & \beta_{13}^{l} & \beta_{14}^{l} & \beta_{15}^{l} & \beta_{16}^{l} \\ \beta_{21}^{l} & \beta_{22}^{l} & \beta_{23}^{l} & \beta_{24}^{l} & \beta_{25}^{l} & \beta_{26}^{l} \\ \beta_{31}^{l} & \beta_{32}^{l} & \beta_{33}^{l} & \beta_{34}^{l} & \beta_{35}^{l} & \beta_{36}^{l} \\ \beta_{41}^{l} & \beta_{42}^{l} & \beta_{43}^{l} & \beta_{44}^{l} & \beta_{45}^{l} & \beta_{46}^{l} \\ \beta_{51}^{l} & \beta_{52}^{l} & \beta_{53}^{l} & \beta_{54}^{l} & \beta_{55}^{l} & \beta_{56}^{l} \\ \beta_{61}^{l} & \beta_{62}^{l} & \beta_{63}^{l} & \beta_{64}^{l} & \beta_{65}^{l} & \beta_{66}^{l} \end{bmatrix} \begin{bmatrix} ln\_pu_{i,t-l} \\ wind_{i,t-l} \\ woutd_{i,t-l} \\ winclose_{i,t-l} \\ woutclose_{i,t-l} \\ wclustering_{i,t-l} \end{bmatrix} + \begin{bmatrix} e_{ln\_pu,it} \\ e_{wind,it} \\ e_{woutd,it} \\ e_{winclose,it} \\ e_{woutclose_{it}} \\ e_{wclustering_{it}} \end{bmatrix}$$

The coefficients of the system are estimated after the fixed effects and the time dummy have been removed. The main results from the PVAR analysis equation (13) are reported in TABLE 5-12. Our major interest lies in

examining dynamic and endogenous relationships between purchase and network position variables. In other words, we intend to investigate whether (a) lagged purchase variable has a significant impact on network position variables and (b) vice versa.

As in TABLE 5-12, some lagged network variables have a significant impact on the present purchase. As expected, lagged in-degree just in t-1 period (β =.04, p-value<.05) and out-degree in t-1 period (β =.003, p-value<.01) have a positive and significant effect on purchase of current t period, the results of which are analogous to the results of random effect panel Tobit model.

Lagged in-closeness centrality in t-1 period has a positive and significant effect on the present period purchase (β =.19, p-value<.01), and Lagged out-closeness centrality in t-1 and t-2 periods has a positive and significant impact on the present period purchase (β =.23, p-value<.01 in t-1 period and β =.06, p-value<.05 in t-2 period). Lagged in-closeness and out-closeness centrality variables have greater impact on purchase than the other network position variables, the results of which are analogous to the results of the random effect panel Tobit model. The previous (lagged) as well as the present ability of reaching from other people (in-closeness centrality) or to other people (out-closeness centrality) quickly is the most influential characteristic among network position variables. This is the case even when

comparing the influence of degree centrality. The existing research on influential actors in social network has focused on degree centrality (e.g., Katona, Zubcsek, and Sarvary 2011; Yoganarasimhan 2012). However, based on our research results, closeness centrality is the most influential variable among network variables, though this cannot be generalizable. It would be interesting to investigate which network position variable is the most influential as a future research.

Lagged clustering coefficients are negatively associated with purchase (β =-.03, p-value<.05 in t-2 period and β =-.02, p-value<.05 in t-3 period). This result contrasts with the results of the random effects panel Tobit model. Clustering has a positive effect on an individual's purchase. However, high clustering can interrupt information flow between communities (Watts and Strogatz 1998). Thus, we can infer that the present highly clustering has a positive effect on an individual's purchase behavior, but the lagged highly clustering has negative impact on the present purchase by blocking the flow of information on product to his or her local community.

TABLE 5-12. Panel Vector Auto Regression Results

	Dependent Variable					
	ln_pu	wind	woutd	winclose	woutclose	wclustering
ln_pu(t-1)	.20***(41.67)	.0002(.24)	-.63(-1.23)	.003***(4.81)	.002***(2.74)	.001(.49)
ln_pu(t-2)	.10***(24.66)	.0002(.34)	.21(.53)	.002***(2.64)	.002***(3.04)	.001(.76)
ln_pu(t-3)	.06***(16.06)	-.0007(-1.07)	-1.12***(-3.12)	-.0001(-.21)	.001**(2.52)	-.005***(-2.81)
wind(t-1)	.04**(2.18)	.61***(10.89)	-4.04(-.69)	-.04***(-6.58)	.06***(3.78)	.13(1.5)
wind(t-2)	-.04(-1.52)	.02(.32)	-1.12(-.21)	-.02***(-3.13)	-.04**(-2.3)	-.17***(-2.61)
wind(t-3)	.02(1.34)	.08(1.31)	4.51(.98)	-.002(-.61)	-.003(-.2)	.05(1.19)
woutd(t-1)	.003***(4.80)	-.00001(-.85)	.51***(20.79)	.0001***(12.6)	.000007(.57)	.00002(.25)
woutd(t-2)	-.000008(-.16)	.000003(.19)	.06***(4.21)	-.00003***(-3.79)	-.00004***(-3.87)	.00007(1.16)
woutd(t-3)	-.00005(-1.05)	-.00002*(-1.73)	.06***(4.77)	-.00001(-1.33)	-.00001(-1.53)	-.00008(-1.59)
winclose(t-1)	.19***(7.79)	.0009(.11)	4.03(1.08)	.35***(62.04)	.07***(13.62)	-04***(-2.66)
winclose(t-2)	.03(1.51)	.002(.46)	.47(.18)	.13***(24.59)	-.007*(-1.68)	-.02(-1.52)
winclose(t-3)	.04*(1.86)	-.009*(-1.90)	3.24(1.16)	.1***(20.47)	.006(1.53)	-.02(-1.2)
woutclose(t-1)	.23***(7.98)	.0004(.03)	15.84***(4.44)	.17***(30.63)	.43***(55.14)	.08***(3.32)
woutclose(t-2)	.06**(2.24)	.006(.59)	-4.46(-1.30)	.01***(2.65)	.18***(26.55)	.02(1.16)
woutclose(t-3)	.03(1.27)	.02(1.18)	4.5(1.4)	.03***(6.24)	.13***(23.7)	.03(1.6)
wclustering(t-1)	-0.000003(-.00)	.02*(1.79)	11.4***(2.78)	-.006**(-2.18)	-.0006(-.19)	.64***(24.04)
wclustering(t-2)	-0.03**(-2.31)	-.01**(-2.06)	-.81(-.33)	-.007***(-3.27)	.18***(26.55)	.07***(3.81)
wclustering(t-3)	-0.02**(-2.20)	.009(1.38)	-.27(-.12)	-.005***(-2.71)	-.006***(-3.16)	.08***(5.1)

Notes: *, **, *** - significant at the 10%, 5%, and 1% level, respectively

We computed Impulse Response Functions (IRFs) from the estimated PVAR system parameters. IRFs are easy to interpret with the estimated results (Joshi and Hanssens 2010; Trusov, Bucklin, and Pauwels 2009; Villanueva, Yoo, and Hanssens 2008). To analyze IRFs, we need an estimate of their 95% confidence intervals. Since the matrix of IRFs is constructed from the estimated VAR coefficients, their standard errors need to be taken into account. Thus, we calculate standard errors of the IRFs and generate confidence intervals with Monte Carlo simulations.[7]

The IRFs demonstrate the incremental effect of a one-standard deviation shock in network structure indicators on the future values of purchase. These enable us to examine carryover effects of each variable on purchase while fully accounting for the indirect effects of these variables in a dynamic system. FIGURE 5-2 and FIGURE 5-3 demonstrate the results. FIGURE 5-2 shows the shock impacts of network position variables on future purchase values, and FIGURE 5-3 represents the shock impacts of purchase on future network position values.

In FIGURE 5-2, IRFs for the effect of weighted in-degree, weighted out-degree, weighted in-closeness centrality, weighted in-closeness centrality, and weighted clustering coefficient on purchase over time are plotted.

---

[7] In practice, we randomly generate a draw of coefficients A of model using the estimated coefficients and their variance-covariance matrix and re-calculate the impulse-responses. We repeat this procedure 500 times. We generate 5th and 95th percentiles of this distribution that we use as a confidence interval for the impulse-responses.

A one-standard deviation shock in in-degree and out-degree centrality has a positive and significant effect on purchase. While in-degree has carryover effects on purchase for 6 periods (months), out-degree has a carryover impact on purchase for about 7 periods (months).

A shock in in-closeness and out-closeness centrality has a positive impact on the future values of purchase. The significant and positive carryover effects continue over 12 months. On the other hand, a one-standard deviation shock in clustering coefficient has a positive impact on the present purchase value, but a negative impact on the future purchase value consistently over 12 periods.

FIGURE 5-3 shows the results of the incremental effect of a one-standard deviation shock in purchase on the future values of network position. A shock in purchase shows insignificant carryover effects on in-degree, out-degree, and clustering coefficient. On the other hand, a shock in purchase has a positive impact on both in-closeness and out-closeness centrality.

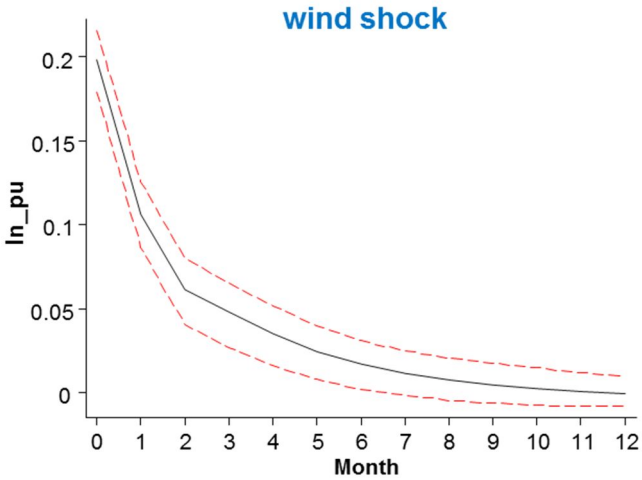
From FIGURE 5-2 and FIGURE 5-3, some interesting results came out. First, we found that the carryover effect of in-closeness and out-closeness centrality on purchase shows the most powerful influence and continues over the longest periods compared with other network variables. In other words, as communication centrality with friends of a node increases, the node subject is likely to purchase items and music. Thus, if companies that

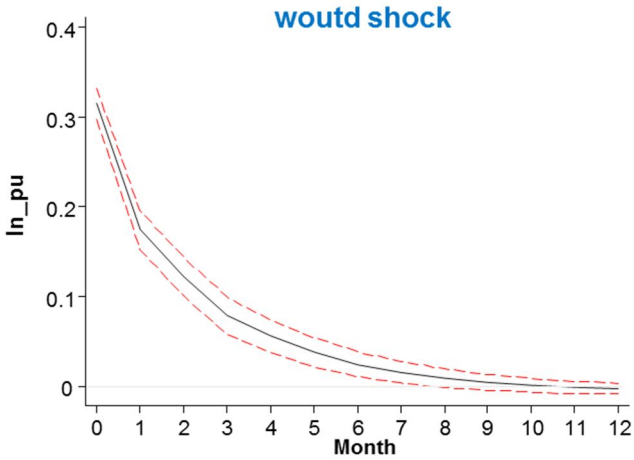
sell online items and music contents intend to increase sales of those contents, one of the most effective ways would be to reduce distance between their target customers in geographical clusters. Second, just future in-closeness and out-closeness centrality out of network position variables are positively influenced by a shock in purchase. That is, buying items or music for decorating nodes' individual homepages makes distance to their friends shorten. Therefore, we found that there exists a positive feedback loop between purchase and closeness centrality, and closeness centrality might be the most important network position variable to be managed properly through time.

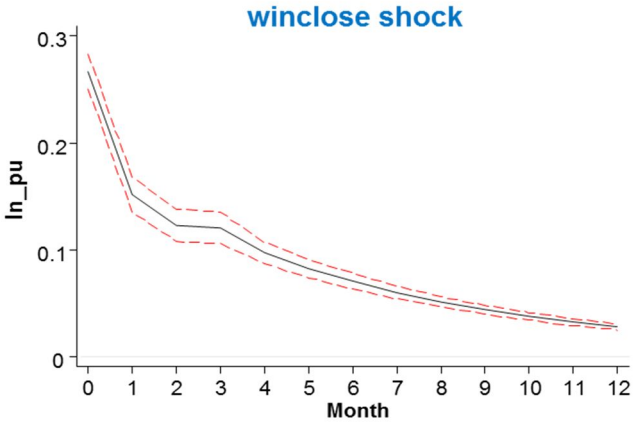
Then, the question will be raised as to how to increase closeness centrality. An important clue is given in Watts and Strogatz (1998). Via simulations, the authors proved that if a level of clustering is lowered by linking two nodes that are unknown to each other cross communities, the shortest path length (closeness) will be shortened. For example, Facebook provides a service that enables consumers to get information on recommended people in FIGURE 5-4. Not only can such a recommendation program increase members of the social network service, but also make the service earn more sales or profits through increasing closeness between nodes by reducing clustering.

FIGURE 5-5 shows that as clustering decreases, closeness increases. In

other words, clustering and closeness centrality are inversely correlated, which is evident from FIGURE 5-6. FIGURE 5-6 indicates that a shock in clustering has a negative and significant effect on future in-closeness and out-closeness centrality.


wind shock
ln_pu
0.2
0.15
0.1
0.05
0
0 1 2 3 4 5 6 7 8 9 10 11 12
Month


woutd shock
ln_pu
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10 11 12
Month


winclose shock
ln_pu
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10 11 12
Month

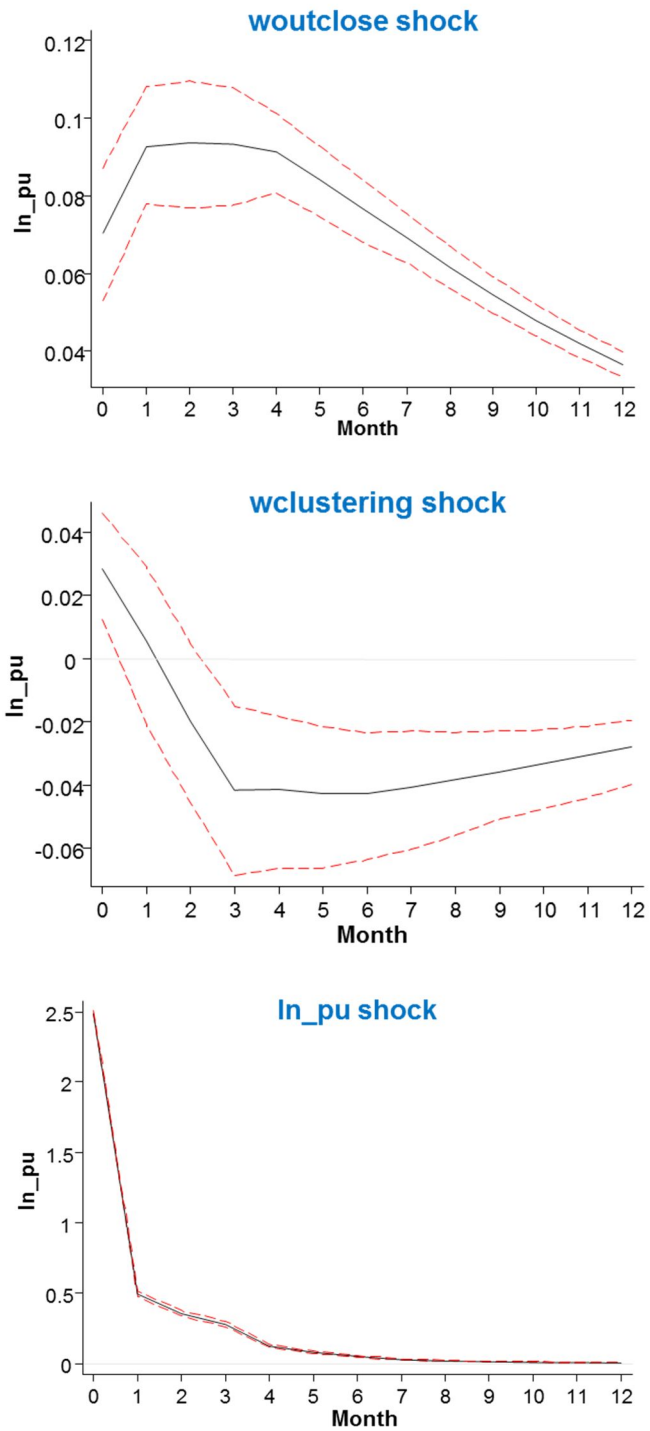
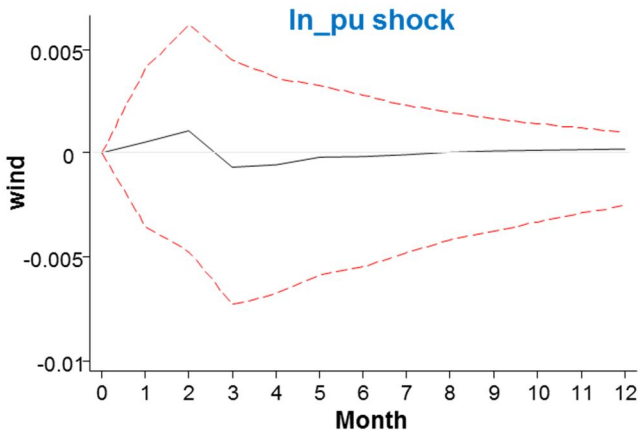
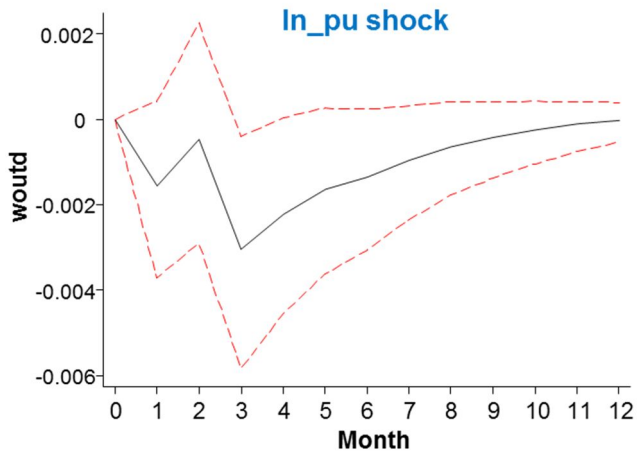
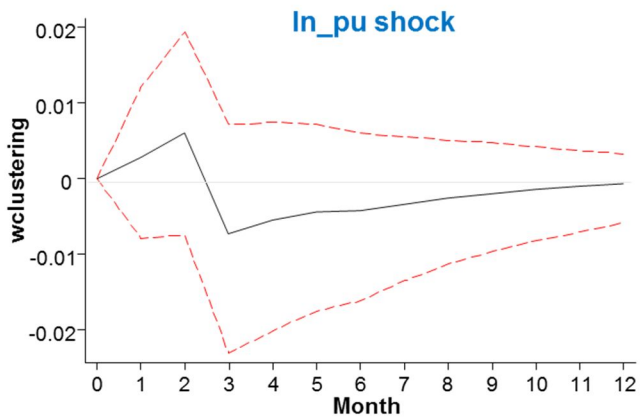
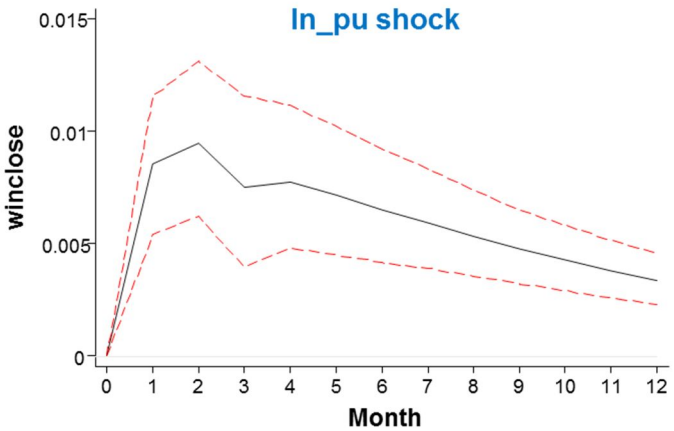


FIGURE 5-2. IRFs Results: the shock of NP (Network Position) Variables


ln_pu shock
0.005
0
-0.005
-0.01
wind
0 1 2 3 4 5 6 7 8 9 10 11 12
Month


ln_pu shock
0.002
0
-0.002
-0.004
-0.006
woutd
0 1 2 3 4 5 6 7 8 9 10 11 12
Month


ln_pu shock
0.02
0.01
0
-0.01
-0.02
wclustering
0 1 2 3 4 5 6 7 8 9 10 11 12
Month



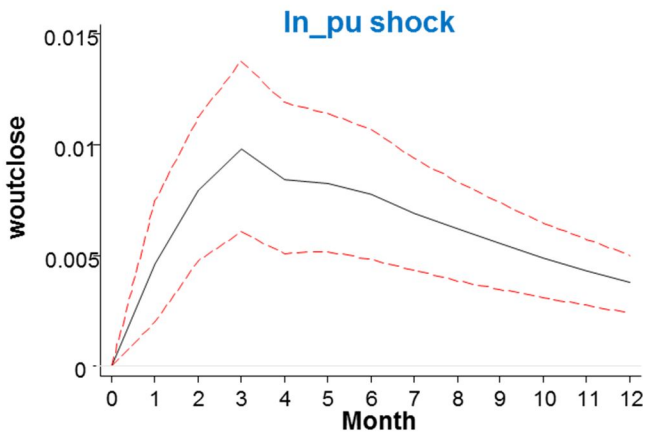


FIGURE 5-3. IRFs Results: the Shock of Purchase on NP (Network Position Variables)

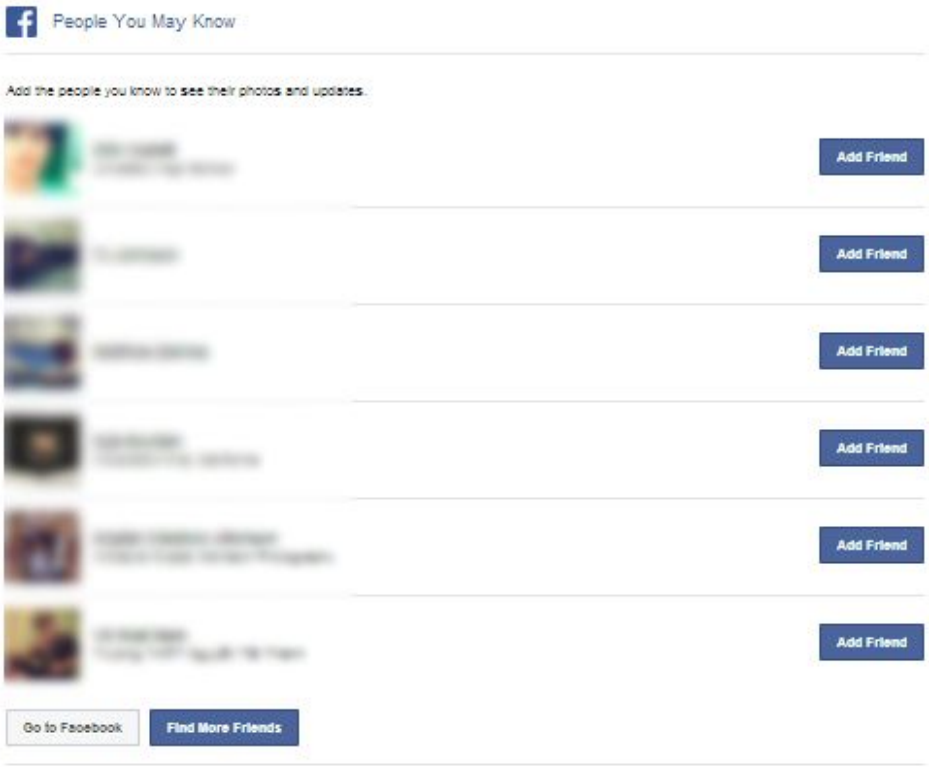


FIGURE 5-4. People You May Know Service in Facebook

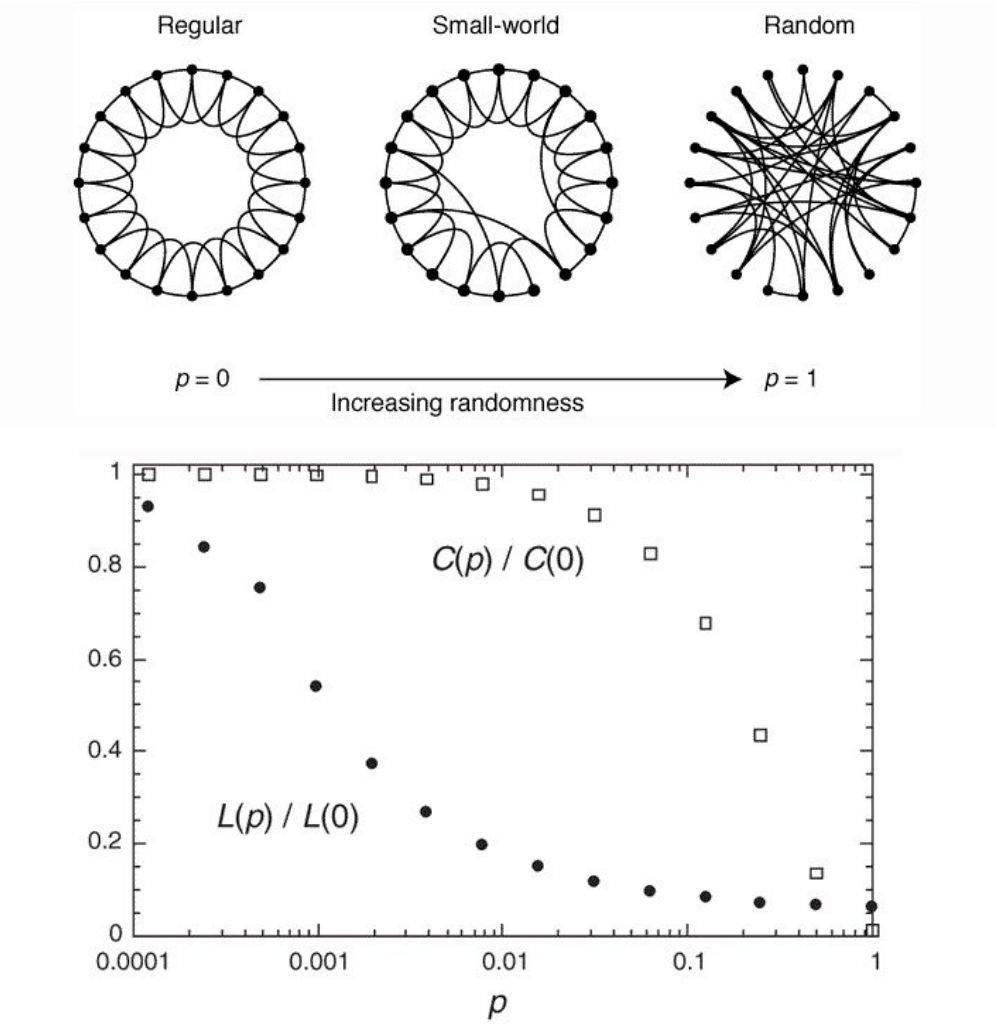
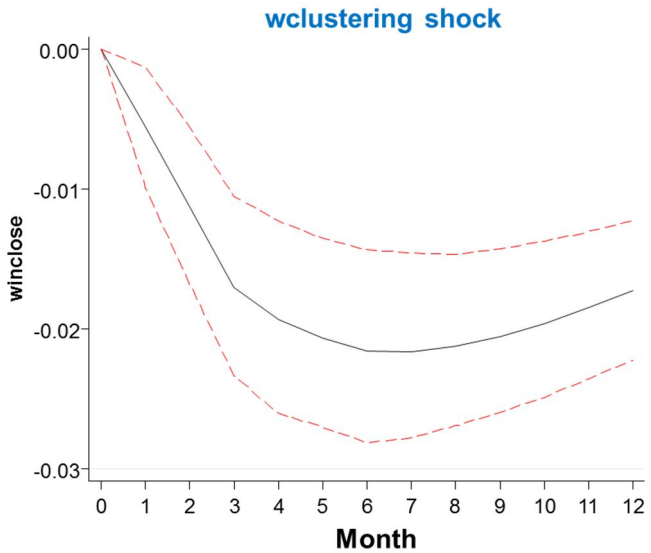


FIGURE 5-5. Dynamic Small World Networks

Source: FIGURE 1 and FIGURE 2 in Watts and Strogatz (1998)



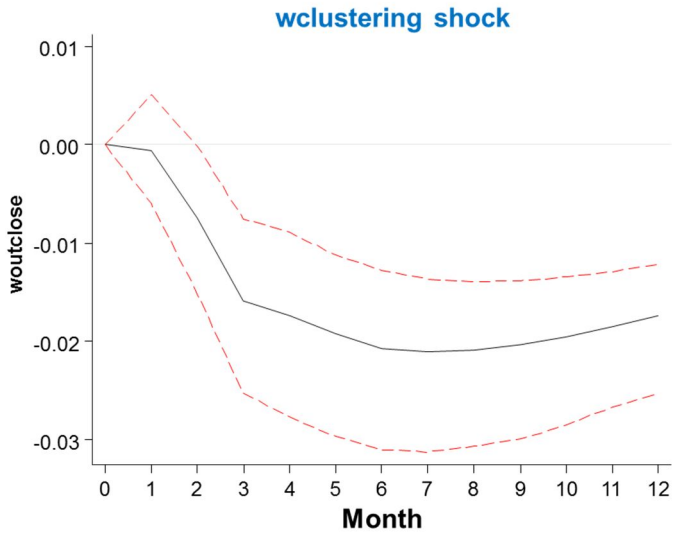


FIGURE 5-6. IRFs Results: the Shock of Clustering on Closeness

5.6. Managerial Implications

The managerial implications through this study are as follow. First, companies should consider dynamic network effects when developing consumer response models using network position variables. For example, a hub, which is measured by counting degree or relationships, can be considerably varied over time. Generally, network positions of actors are given or static over time in traditional models, which can cause biased parameter values.

Second, the present results indicate that there exist dynamic and endogenous relationships between purchase and network position variables. In other words, network positions of actors can contribute to consumer decision making, such as purchase. In particular, in-closeness and out-closeness variables have a more powerful influence on purchase behavior, and purchase behavior can cause closeness variables that can be increased or reduced by controlling the clustering coefficient. If companies can understand these results and apply them to the actual practices, it is possible to get targeting strategies that are more accurate and increase sales.

Ⅵ. Conclusion

6.1. Summary and Contributions

Stephen and Toubia (2010) stated, “It will be profitable if online platforms make online-sellers link each other.” This principle can be applied to other consumer behaviors because consumers make decisions on their behaviors by referring their acquaintances, mainly through online social media. Therefore, it gets more important to understand various network structures of consumers. Thus, much research has studied the relationships between consumer behaviors and network structures. In this vein, we tried to investigate the topic in a way that builds on the related previous research.

From this research, we intended to answer the research questions. The first question is whether network topological variables are dynamic and, relatedly, the model applying dynamic network position variables is superior to the model using static network variables. The second question is whether the coefficients of network variable positions have time heterogeneity. The final question is whether there are such endogenous relationships among consumer activities and social network position variables and carryover effects among variables.

With respect to the first question, we found that network position

variables have considerable variations over time, in as the results of Braha and Bar-Yam (2006). In addition, we compared the model fit between the base model that used time invariant network position variables and the model fit of the proposed model, which applied time-varying network position variables. As a result, the proposed model shows better performance in terms of both the whole data model and in sample fit and out of sample fit. Moreover, the proposed model shows parameter estimation results that are less biased compared to the base model.

By estimating and comparing the multi-level panel random intercept model and the multi-level panel random coefficient model, we found that all network position variables, excluding constraint, have time-varying coefficients, which shows that if applying static network variables, the estimated parameters can be biased.

In addition, we found that there exist endogenous and dynamic effects between purchase and network position variables over time. Not only lagged weighted in-degree and out-degree but also lagged weighted in-closeness and out-closeness centrality have a significant and positive impact on the present purchase. On the other hand, lagged weighted clustering coefficient has a significant and negative effect on the present purchase. However, lagged purchase has a significant and positive impact on the present weighted in-closeness and out-closeness centrality; however, it has an

insignificant effect on the other two degree variables and weighted clustering. From the results, in-closeness and out-closeness variables are the most influential among network variables and they are forming a positive feedback loop with purchase variable. Moreover, closeness can be controlled by clustering coefficient.

This paper can contribute to the existing research in three main respects. First, to the best of our knowledge, this research dealt with the dynamic network variable specifications in estimating models that investigate relationships between consumer behavior (in this paper context, purchase quantity) and network position variables for the first time. The finding that dynamic network positions can explain considerable consumer activities such as purchase behavior is the second posited contribution. The last contribution is that this study found dynamic and endogenous effects between purchase and network position variables by using the PVAR model.

6.2. Limitations and Future Research

Despite important contributions and implications, this research has some limitations, and various future research topics should be explored. First, the proposed models did not consider marketing variables such as price and promotion information. The social network site from which we get data has

provided various promotions and price discounts to its members. To exclude the influence from the company and focus on the interaction effects among members, we removed promotion data. In the future research, it might be a good idea to consider marketing variables in estimating the proposed model.

Second, the dataset for this study contains a sampled network that is not global. Thus, the results from this study may not be generalizable to a global network. However, a typical social network is too large to compute for either researchers or marketing practitioners, and it is usual to use sample data. A snowball sampling method is generally applied to sample social network data. It is widely known that the results from snow balling sampled data are similar to the results from global network data (Frenzen and Davis 1990; Goodman 1961; Henry 2005; Narayan and Yang 2007; Salganik and Heckathorn 2004; Tepper 1994). In the future research, thus, it would be desirable to sample various network data and compare the results (Chen, Yuxin, and Ping 2013).

Third, the data contains 1-year of observations, which makes it hard to remove seasonal effects such as the Christmas season. In addition, the dataset is given as monthly data; thus, we cannot consider weekly or daily effects of network variables. For future research, the proposed model could use weekly or daily panel data as well as monthly data covering over two years.

Fourth, this research does not consider neighbors' influence on an individual's purchase. As per Katona, Zubcsek, and Sarvary (2011)'s research, the future research can consider a model applying neighbors' average influence on an individual.

Finally, we can extend this research to utilitarian products or high involvement products. This research deals with the purchase data of online items or music, which is a symbolic, or low-involvement, product. Therefore, applying utilitarian or high-involvement product purchase data to the proposed model would be of interest.

References

Beauchamp, Murray A. (1965), “An Improved Index of Centrality,” *Behavioral Science,* 10, 161-63.

Becker, Marshall H. (1970), “Sociometric Location and Innovativeness: Reformulation and Extension of the Diffusion Model,” *American Sociological Review*, 35 (2), 267–82.

Borgatti, Stephen P. (2006), “Identifying Sets of Key Players in a Network,” *Computational, Mathematical and Organizational Theory*, 12 (1), 21-34.

——— and Everett, M.G. (2006), “A Graph-theoretic Framework for Classifying Centrality Measures.” *Social Networks* , 28 (4), 466-84.

Braha, Dan and Yaneer Bar-Yam (2006), “From Centrality to Temporary Fame: Dynamic Centrality in Complex Networks,” *Complexity*, 12 (2), 59-63.

Brin, Sergey and Lawrence Page (1998), “The Autonomy of a Large-Scale Hypertextual Web Search Engine,” *Computer Networks ISDN*, 30 (1-7), 107-17.

Burt, R. S. (1980), "Autonomy in a Social Topology." *The American Journal of Sociology*, 85 (4), 892-925.

——— (1992), *Structural Holes: The Social Structure of Competition*, Harvard University Press*,* Cambridge, MA.

——— (2004), “Structural Holes and Good Ideas”**,** *The American Journal of Sociology,* 110, 349-99.

Cancian, F. (1979), *The innovator’s situation: Upper-middle-class conservatism in agricultural communities,* University Press, Palo Alto, CA: Stanford.

Chen, Xinlei, Yuxin Chen, and Ping Xiao (2013), “The Impact of Sampling and Network Topology on the Estimation of Social Intercorrelations,” *Journal of Marketing Research*, 50 (February), 95-110.

Choi, Hanool, Sang-Hoon Kim, Jeho Lee (2008), “Role of Network Structure and Network Effects in Diffusion of Innovations,” *Industrial Marketing Management*, 39, 170-77.

Coleman, J. S. (1988), “Social Capital in the Creation of Human Capital,” *The American Journal of Sociology* , 94, 95-120.

——— (1990), *Foundations of social theory*, Cambridge, Mass.: Harvard University Press.

Dormann, Carsten F., et al. (2013), “Collinearity: a review of methods to deal with it and a simulation study evaluating their performance,” *Ecography*, 36, 27-46.

Enders, Walter (2004), *Applied Econometric Time Series*, 2nd ed., John Wiley & Sons.

Frenzen, J. K. and Harry L. D. (1990), "Purchasing Behavior in Embedded Markets," *Journal of Consumer Research,* 12 (June), 1-12.

Girvan, M. and Newman, M. E. J. (2002), “Community structure in social and biological networks,” *Proceedings of the National Acadeour of Sciences,* 99 (12), 7821–26.

Goldenberg, J., Han, S., Lehmann, D. R., and Hong, J. W. Hong (2009), “The Role of Hubs in the Adoption Process,” *Journal of Marketing*, 73 (March), 1–13.

Goodman, L. A. (1961), "Snowball Sampling," *Annals of Mathematical Statistics,* 32 (1), 148-70.

Granovetter, M. (1973), “The Strength of Weak Ties,” *The American Journal of Sociology*, 78 (6), 1360–80.

——— (1982), “The strength of weak ties: A network theory revisited”,

Sociology theory, 1, 201-33.

Greene, William H. (2008), *Econometric Analysis*, 6th ed., Pearson Education, Upper Saddle River, NJ 07458.

Handcock, Mark S., Adrian E. Raftery, and Jeremy M. Tantrum (2007), "Model-Based Clustering for Social Networks," *Journal of the Royal Statistical Society: Series A*, 170 (2), 301–54.

Hartmann, W. R., Manchanda, P., Nair, H., Bothner, M., Dodds, P., Godes, D., et al. (2008). "Modeling social interactions: identification, empirical methods and policy implications", *Marketing Letters*, 19 (3), 287-304.

Henry, P. C. (2005), "Social Class, Market Situation, and Consumers' Metaphors of (Dis) Empowerment," *Journal of Consumer Research,* 31 (March), 766-78.

Hinz, Oliver, Skiera, Bernd, Barrot, Christian and Becker, Jan (2011), "Seeding Strategies for Viral Marketing: An Empirical Comparison," *Journal of Marketing*, 75 (November), 55-71.

Holme, P. and Goshal G.(2008), "The diplomat's dilemma: Maximal power for minimal effort in social networks", *physics*, 2008/05/26, DOI:0805.3909,arXiv.

Holtz-Eakin, D., Newey, W. and Rosen, H. (1988), "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56 (6), 1371-95.

Huh, Myung-Hoe (2010), *Introduction to Social Network Analysis Using R*, Seoul, Freedomacademy.

Im, K.S., M.H. Pesaran, and Y. Shin (1997), "Testing for Unit Roots in Heterogeneous Panels," manuscript, department of Applied Economics, University of Cambridge, UK.

Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2011), "Opinion Leadership and Social Contagion in New Product

Diffusion," *Marketing Science*, 30 (2), 195-212.

Joshi, Amit and Dominique M, Hanssens (2010), "The Direct and Indirect Effects of Advertising Spending on Firm Value," Journal of Marketing," *Journal of Marketing*, 74 (January), 20-33.

Katona, Zsolt, Peter Pal Zubcsek, and Miklos Sarvary (2011), "Network Effects and Personal Influences: Diffusion of an Online Social Network," *Journal of Marketing Research,* 48 (June), 425-43.

Kim, Yong-Hak (2003), *Social Network Analysis*, Seoul, Parkyoungsa.

Krackhardt, D. 1992. The strength of strong ties: The importance of philos in organizations. in *Networks and organizations: Structure, form, and action*, N. Nohria and R. G. Eccles, eds., Boston: Harvard Business School Press, 216-39.

——— (1998), "Simmelian Tie: Super Strong and Sticky," in *Power and Influence in Organizations.*, edited by Roderick M. Kramer and Margaret Neale, 21-38. Thousand Oaks, CA: Sage.

Lee, Hee Tae and Byung-Do, Kim (2013), "The Role of Brokers in Social Network on the Product Purchase," *Journal of Korean Marketing Association*, 28 (December), 1-22.

Levin, A. and C.F. Lin (1992), "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," Discussion paper No. 92–23, UCSD.

Love, Inessa and Lee Zicchino (2006), "Financial Development and Dynamic Investment Behavior: Evidence from panel VAR," *The Quarterly Review of Economics and Finance*, 46, 190-210.

Min, In Sik and Pilsun Choi (2012), *Advanced Panel Data Analysis*, Jiphil Media

Narayan, Vishal and Sha Yang (2007), "Modeling the Formation of Dyadic Relationships Between Consumers in Online Communities," working paper, Samuel Curtis Johnson Graduate School of Management, New

York University.

Oestreicher-Singer, Gal and Arun Sundararajan (2012a), “Recommendation Networks and the Long Tail of Electronic Commerce,” *MIS Quarterly*, 36 (1), 65-83.

——— and ——— (2012b), “The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets,” *Management Science*, 58 (11), 1963-81.

———, Barak Libai, Liron Sivan, Eyal Carmi, and Ohad Yassin (2013), “The Network Value of Products,” *Journal of Marketing*, 77 (May), 1-14.

Pak, Yong Suhk, KyoungmWe Lee and Jiwon Lee (2012), “The Types of Social Capital Formed on Social Networking Services (SNS) and their Effects on Consumer Preference for a Product Type,” *Korean Management Review*, 41 (6), 1619-41.

Putnam, R. (1993), "The Prosperous Community: Social Capital and Public Life." *The American Prospect*. 13, 35-42.

Rabe-Hesketh, S. and Anders Skrondal (2008), *Multilevel and longitudinal modeling using stata*, Stata Press, College Station, Tex. : StataCorp LP

Salganik, M. J. and Douglas D. Heckathorn (2004), "Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling," *Sociological Methodology,* 34 (1), 193-239.

Stephen, Andrew T. and Oliver Toubia (2010), “Deriving Value from Social Commerce Networks,” *Journal of Marketing Research*, 47 (April), 215-28.

Tepper, Kelly (1994), "The Role of Labeling Processes in Elderly Consumers' Responses to Age Segmentation Cues," *Journal of Consumer Research,* 20 (March), 503-19.

Trusov, Michael, Randolph E. Bucklin, and Koen Pauwels (2009), “Effects

of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing*, 73 (September), 90-102.

———, Anand V. Bodapati and Randolph E. Bucklin (2010), "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research,* 47 (August), 643–58.

Tucker, Catherine (2008), "Identifying Formal and Informal Influence in Technology with Network Externalities," *Management Science*, 54 (12), 2024-38.

Valente, T. W., Fujimoto, K. (2010) "Bridging: Locating critical connectors in a network." *Social Networks,* 32, 212–20.

Van den Bulte, Christophe, and Stefan Wuyts (2007), *Social Networks and Marketing*, Marketing Science Institute.

Villanueva, Julian, Shijin Yoo, and Dominique M. Hanssens (2008), "The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth," *Journal of Marketing Research,* 45 (February), 48-59.

Wasserman, Stanley, and Katherine Faust (1994), *Social Network Analysis*, London: Cambridge University Press.

Watts, J. Duncan and Steven H. Strogatz (1998), "Collective dynamics of 'small-world' networks," *Nature*, 393 (6684), 440–42.

——— and Peter Sheridan Dodds (2007), "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, 34 (4), 441–58.

Yang, Sha and Allenby, Greg M.(2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research,* 40 (August), 282-94.

Yoganarasimhan, Hema (2012), "Impact of Social Network Structure on

Content Propagation: A Study Using YouTube Data," *Quantitative Marketing and Economics*, 10, 111-50.

Abstract

The Effects of Dynamic Social Network Position on Product Purchase

Hee Tae Lee
Byung-Do Kim, Ph.D. Advisor
College of Business Administration
The Graduate School of
Seoul National University

We intend to answer the following research questions through this research. First, are network topological variables dynamic, and is the model applying dynamic network position variables superior to the model using static network variables? Second, do the coefficients of network variable positions have time heterogeneity? Finally, are there endogenous relationships among consumer activities and social network position variables and carryover effects among variables?

With respect to the first question, we found that network position variables have considerable variations over time by comparing the model fit

of the base model that applied time-invariant network position variables and the model fit of the proposed model that applied time-varying network position variables. As a result, the proposed model showed better performance in terms of both the whole data model and in-sample fit and out-of-sample fit. Moreover, the proposed model showed less biased parameter estimation results compared to the base model.

By estimating and comparing the multi-level panel random intercept model and the multi-level panel random coefficient model, we found that all network position variables, excluding constraint variables, had time-varying coefficients, which showed that the estimated parameters can be biased when applying static network variables.

In addition, we found that endogenous and dynamic effects exist between purchase and network position variables over time. That is, lagged weighted in-degree and out-degree as well as lagged weighted in-closeness and out-closeness centrality have a significant and positive impact on the present purchase. On the other hand, the lagged weighted clustering coefficient has a significant and negative effect on the present purchase. However, lagged purchase has a significant and positive impact on the present weighted in-closeness and out-closeness centrality and an insignificant effect on the other two degree variables and weight clustering. The results suggest that in-closeness and out-closeness variables might be the most influential among

the network variables. In addition, we found that the "Small World" phenomenon exists in social networks. Thus, to reduce the distance between consumers and increase the closeness centrality of nodes in social networks, recommending friends of friends to nodes (e.g., "People Who You May Know," a recommendation service on Facebook) would be the best policy for increasing sales.

Based on the above results, managerial implications, limitations, and future research topics were presented.

Keywords: social network, dynamic network position variable, time-varying coefficients, random effects panel Tobit model, panel vector auto regression, endogeneity