The Dissertation Committee for Dae-Yong Ahn
certifies that this is the approved version of the following dissertation:

A Dynamic Model of Usage Behavior and Network Effects in Social Network Sites

Committee:

Vijay Mahajan, Supervisor

Romana Khan

Raghunath Rao

Garrett Sonnier

Randal Watson

A Dynamic Model of Usage Behavior and Network Effects in Social Network Sites

by

Dae-Yong Ahn, B.A., M.A., M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2009

Dedicated to my family.

Acknowledgments

I wish to thank my family for their love and support. Without them, I would not have strength and courage to pursue my studies. I especially thank my father for being a guiding force in my life. He will always be a person that I aspire to be. I also thank my committee members, Romana Khan, Vijay Mahajan, Raghunath Rao, Garrett Sonnier, and Randal Watson for their comments, suggestions, and corrections. I especially thank Randal Watson for his guidance throughout my doctoral program. As much as he taught me as a scholar, he taught me much more as a person. I also thank Vijay Mahajan for supervising the dissertation committee. I am also thankful to Susan Broniarczyk for her support during my doctoral program. Thanks to her, I had freedom to pursue my studies. I also thank Helen Anderson for her support and friendship.

A Dynamic Model of Usage Behavior and Network Effects in Social Network Sites

Publication No. ___________________

Dae-Yong Ahn, Ph.D.
The University of Texas at Austin, 2009

Supervisor: Vijay Mahajan

This paper structurally estimates a dynamic model of usage behavior and network effects in a social network site using data from MySpace.com. We view a social network as a stock of capital that yields a flow of utilities over time by creating social interactions between the owner and her friends. When one decides to use a social network site, it may have two distinct network effects: (1) one can manage an existing base of friends through social networking and thus prevent depreciation of capital stock (maintenance effect), and (2) one may acquire new friends through social networking, which results in creation of new capital stock (investment effect). Thus, we model social networking as a dynamic process, in which one's current action to use a social network site can influence the evolution of her social network. We found that real-time chat and messaging, features of MySpace.com, positively affect one's usage decision and hence achieve the intended goal of generating site traffic. However, different demographic groups may have idiosyncratic preferences for

these features. Based on parameter estimates, we performed counterfactual simulations with the goal of providing managers with ways to enhance firm performance.

Table of Contents

List of Tables

Chapter 1

Introduction

This dissertation structurally estimates a dynamic model of usage behavior and network effects in social network sites using data from MySpace.com. We view a social network as a stock of capital that yields a flow of utilities over time by creating social interactions between the owner and her friends. When one decides to use a social network site, it may have two distinct network effects: (1) one can manage an existing base of friends through social networking and thus prevent depreciation of capital stock (maintenance effect), and (2) one may acquire new friends through social networking, which results in creation of new capital stock (investment effect). Thus, we model social networking as a dynamic process, in which one's current action to use a social network site can influence the evolution of the social network. The goal of this dissertation is to measure how features of a social network site affect usage behavior and evaluate counterfactual policies that are managerially relevant.

Among US Internet users, 33% of adults and 70% of teenagers used a social network site every month in 2007. eMarketer projects that "50% of online adults and 84% of online teens in US will use social networking" by 2011. To tap into this audience, marketers spent $900 million on US social network sites in 2007 and this figure is expected to climb to $2.7 billion by 2011 (eMarketer, 2007).[1] Industries that advertise in social network sites range from entertainment (25.2%), retail goods and services (17.6%), and telecommunications (16.2%) to financial services (6.3%) and automotive (5.1%) (Nielsen/NetRatings, September 2006).[2] MySpace.com and Facebook.com were largest among US social network sites in 2007. Google signed a three-year, $900 million deal to become an exclusive ad provider to MySpace.com, whereas Microsoft announced a plan to invest $240 million in Facebook.com (BusinessWeek, 2007). eMarketer reports that MySpace.com and Facebook.com accounted for 72% of social network advertising spending in 2007. Media giants such as NBC and Warner Bros. host sites on MySpace.com, while Coca-Cola, CBS, and Chase promote their products on Facebook.com. However, there is a growing trend towards niche social network sites and marketer-sponsored social network sites that attract "a smaller, but passionate audience" rather than the "diverse membership" of MySpace.com and Facebook.com (CNN, 2008).[3]

The benchmark that ranks sizes of social network sites is the number of unique visitors in a month. comScore reports that MySpace.com and

[1]In 2007, advertising spending decreased by 2.7%, 4.7%, and 2.1% for television, newspaper, and radio, whereas it increased by 16.7% for the Internet (TNS media intelligence, 2007). JackMyers.com reports that the Internet's share of advertising budgets will grow 6.2% in 2006 to 10.5% in 2009.

[2]The numbers in parentheses are percentages of the total advertising dollars that each industry spent on social network sites in 2006.

[3]P&G and NBC created petside.com to promote Iams pet foods and the Today show; Campbell's created artofcookie.com for Pepperidge Farm cookies.

Facebook.com each had 72 and 37 million US unique visitors in June 2008. Site traffic closely relates to the advertising revenues of social network sites: Merrill Lynch reports that the advertising rate in MySpace is set at $1.83 per thousand views (BusinessWeek, 2007). As a result, social network sites are in a battle to develop new features that can increase site traffic by attracting Internet users. Facebook.com has a proprietary technology that limits the use of its applications on other sites. In contrast, MySpace.com joined a Google-led alliance "OpenSocial" that allows web developers to invent features that can be used across social network sites (Financial Times, 2007). Developing the right features matters: when Facebook.com tried a new feature that highlights changes made by members to their profiles, it was perceived as an invasion of privacy and sparked protests from members (Washington Post, 2006). Social network sites still actively experiment with new features. In April 2008, Facebook.com unveiled an "instant message" feature, "a chat system that allows members to type back and forth instantly" (CNN, 2008).

We structurally estimate a dynamic discrete choice model to measure how features of social network sites affect one's daily usage decision. We define "usage" as a decision to log in to a social network site on a given day. This definition of usage highly correlates with the number of unique users in a month, a benchmark used in the industry to compare sizes of social network sites. Our approach yields managerial insights and provides a novel way of modeling social networking behavior. First, we estimate parameters that are highly relevant to the industry by modeling one's utility from networking as a function of interactions on social network sites. Social network sites offer various features through which members can interact with each other. For instance, MySpace.com offers two major features through which members can communicate with each other—real-time chat and messaging. Therefore, it is of utmost interest to social network sites to determine how these features affect members' utility of using the Web site. Second, we propose and evaluate counterfactual policies that social network sites may adopt to enhance firm performance. For example, we can measure the impact on site traffic of introducing a "matching service," a collaborative filtering system that matches members based on commonality. Another example is to facilitate meetings among friends who belong to the same social network. Such meetings can be arranged by devising a mechanism that alerts members when their friends log in to the Web site. A priori estimates for the effects of these interventions on site traffic can help managers better design their Web sites.

The contribution of this dissertation can also be seen in the light of the existing "social effects" literature. A seminal paper by Manski (1993) pointed out the difficulty of identifying social effects and proposed a "linear-in-means" parametric estimation for social effects. Brock and Durlauf (2001) provided an analysis of aggregate outcomes among individuals when their utilities are affected by social interactions. They also suggested how to estimate the effects of social interactions on these outcomes using the logistic model. Graham (2008) proposed a semi-parametric estimation method that allows social effects to take a more flexible functional form than the Manski's linear-in-means form.

Hartmann (2008) developed a game-theoretic model for the estimation of social effects, which explicitly allowed for the endogeneity of decisions among individuals. Trusov, Bodapati, and Bucklin (2006) used a variable selection method to identify "important" members in social network sites; these important members are the ones who affect site traffic more than others. These articles assume that social networks are exogenously given and do not account for the fact that social networks may be endogenously chosen or affected by individuals.

Arcidiacono and Nicholson (2005) investigated peer effects among medical students on their academic performance and choices of specialties. They accounted for possible endogeneity on choices of medical schools using a fixed effects model. Mayer and Puller (2008) used data from Facebook.com to study the formation of social networks on university campuses to identify the factors that predict a social tie between two students. Kohler, Behrman, and Watkins (2006) analyzed the determinants of individuals' risk perceptions of HIV infection and accounted for the endogeneity between individuals' risk perceptions and their choice of social networks. Although these papers accounted for the endogeneity in formation or choices of social networks in some fashions, they used a reduced-form model to estimate social effects and thus their methods are not suitable for policy experiments.

In contrast, we model social networking in a dynamic, structural framework that solves a utility maximization problem of decision-makers explicitly.[4]

[4] Ryan and Tucker (2007) used a dynamic discrete choice model to incorporate network

We consider two distinct network effects that make social networking inherently dynamic. Maintenance effect implies that one's current action to use a social network site signals friends about her commitment to friendship, which leads to higher involvement of one's friends in social networking in the future. Thus, one can maintain her social capital so that it will yield higher dividends in the future. Investment effect implies that one's action to use a social network site may result in expansion of her social network, which leads to a higher degree of social interactions in the future. Therefore, one accounts for both current and future benefits from social networking in deciding to use a social network site. We base our model on the theory of utility maximization and solve decision-makers' optimization problem explicitly. Thus, we can use the estimates of our model to evaluate counterfactual policies that managers may adopt to increase site traffic.

The estimation of the model requires repeatedly solving a discrete-choice, infinite-horizon optimization problem of one's decision to use a social network site. We follow a two-step estimation technique, first proposed by Hotz and Miller (1993) in a single agent framework and later extended by Bajari, Benkard, and Levin (2007) to a game-theoretic framework. BBL proposes a forward-simulation technique to mitigate computational burden in solving dynamic programming problems. In the first-step, we estimate policy functions and state transitions parametrically and use these estimates to approximate

effects on the adoption of new videoconferencing technology. They analyze one-time technology 'adoption' as an optimal stopping problem, whereas we focus on on-going 'usage' of a social network site.

value functions at each state up to a finite vector of parameters. The estimated value functions are then used as regressors in the second-step regression to recover structural parameters of interest.

We randomly selected a group of college-age members of MySpace.com and collected data from their Web pages on a daily basis for four weeks. We recorded three types of variables for each member of our sample on a daily basis: usage behavior, social interactions, and the evolution of the social network. For usage behavior we recorded whether or not one used MySpace.com for each member each day. For social interactions we recorded real-time chat and messaging, two features of MySpace.com, for each member each day. For the evolution of a social network, we kept track of changes in members' social networks by recording the size of each member's social network each day.

We found that features of MySpace.com positively affected the daily usage decision and hence achieved the intended goal of generating site traffic: real-time chat, measured by the number of friends who are online, and messaging, the number of messages per day, both had positive effects on usage. However, we found some substantial differences in usage behavior across demographic groups. Specifically, the rate of daily site usage decreased with members' ages. US residents also used social network sites at the higher rate than non-US residents. We also found that there exists heterogeneity among demographic groups in their preferences for features of MySpace.com. Notably, we observed statistically significant differences in the effects of messaging on daily site usage across demographic groups. There were differences in site usage with respect to day of the week, implying that members were more likely to use social network sites on weekdays than on weekends. However, these differences with respect to day of the week only had a p-value of about 0.15.

Based on the parameter estimates, we performed a series of policy experiments to evaluate three counterfactual policies: prevention of coordination failure, enhancement of networking experience, and matching service. The first policy is designed to prevent coordination failure among friends by facilitating online meetings. An example of such a policy may be to inform members of MySpace.com when their friends are online through mobile phones or electronic mails. The second policy is designed to enhance the utilities from social networking on MySpace.com. For instance, MySpace.com integrated Skype into its Instant Messenger in 2008 to enable video conferences among its members. Another way to enhance the networking experience may be to upgrade messaging services by allowing members to embed different types of graphics or videos in their messages. The third policy is designed to assist in the acquisition of new friends. An example may be the adoption of a collaborative filtering system used by online retailers such as Amazon.com or survey techniques used by online dating services, such as Eharmony.com and Match.com. We found that adopting these policies resulted in increases in site traffic. However, their effects on site traffic can differ over a period of time. The first two policies were more effective in increasing site traffic immediately after adoption due to higher utilities from increases in social interactions or better networking experiences. In contrast, the third policy had a marginal effect on site traffic initially but was more effective in increasing site traffic in the long run.

This dissertation is organized as follows: Chapter 2 reviews the literature that estimates social effects in various contexts. Chapter 3 describes the model. Chapter 4 describes the data. Chapter 5 describes the estimation method. Chapter 6 discusses the results of the estimation. Chapter 7 describes the policy experiments and discusses the results. Chapter 8 presents the conclusions.

Chapter 2

Literature Review

In this section, we review the existing literature on social effects. Especially, we focus on the recent empirical literature that either provides empirical frameworks for social effects or measures the effects of social interactions in various situations by applying these methodologies. This literature is divided into two categories depending on whether social networks are assumed to be exogenously given or endogenously chosen.

2.1 Social networks as exogenously given

In his seminal paper, Manski (1993) defined social effects as "the propensity of an individual to behave in some way [that] varies with the prevalence of that behavior in some reference group containing the individual." This definition of social effects is inclusive of the terms used in a variety of contexts, such as "social norms," "peer influences," "neighborhood effects," "conformity," "herd behavior," "social interactions," or "interdependent preferences." Manski proposed a linear-in-means model with a continuous dependent variable, in which an individual's behavior varied with the mean of others' in the reference group.

Brock and Durlauf (2001) provided an analysis of aggregate outcomes when an individual's utility was affected by social interactions. They incorporated social effects in a utility maximization framework and studied equilibrium properties under such a scenario. In their setting, each individual makes choices conditional on the expectation of the mean choice level of his group. In contrast to Manski's linear-in-means framework, Brock and Durlauf used the binary choice framework that imposed "a nonlinear relationship between group characteristics and group behaviors." Based on this theoretical framework, they suggested an empirical method to estimate social effects based on the logistic model. Their subsequent paper proposed a multinomial-choice model that accommodated the presence of social interactions (Brock and Durlauf, 2002).

Castronova (2004) estimated the effects of social norms on sexual activity among US high school students. Social norms in this paper took the form of a punishment strategy, where stigma was imposed on students who were sexually active and on those who did not adopt the punishment strategy. Using the Add-Health Survey data, Castronova estimated the effects of social norm with percentage of students who were sexually active in a school and various school policies as proxies for social norms and a punishment strategy respectively. He found that the rate of sexual activity was about 5 percent lower with norm-enforcing equilibria.

Trusov, Bodapati, and Bucklin (2006) used a variable selection method to identify two types of members: ones whose activities influenced others' site usage and those who were susceptible to the influence. They modeled usage rate of members using a Poisson regression, where the rate of usage was assumed to be a function of members' own characteristics as well as the previous usages by their "important" friends. Trusov, Bodapati, and Bucklin used a reduced form approach that did not explicitly capture the endogeneity between one's action and its effect on the state of her social network. Hence, their method cannot be used to perform policy simulations.

Graham (2008) proposed a methodology that was used identify peer effects on academic achievement of students. He pointed out the difficulty of separating the effects of social interactions from those of group-level heterogeneity. For instance, the differences in academic performances across classrooms may come from the variation of teacher quality or that in peer effects. Graham proposed the use of conditional variance restrictions to separate the between-group variance of outcomes into the variance from group-level heterogeneity and that from variation in peer quality across groups.

Hartmann (2008) proposed a game-theoretic model for estimating the effects of social interactions in the context of golf consumption. His approach explicitly allows for the endogeneity of decisions among individuals. Hartmann accounted for the problem of multiple equilibria that may arise in discrete games by imposing the assumption of Pareto dominance: Individuals in the same group prefer the option of playing together to that of none of them playing. Hartmann found significant effects of social interaction such that 65% of value of playing golf was attributable to an individual, while 35% was attributable to the individual's utility from playing with group members.

These articles assume that social networks are exogenously given and do not account for the fact that social networks may be endogenously chosen or affected by individuals. This may be due to lack of richer data sets that contain detailed information about each individual's behavior and the evolution of her social network over time. A natural criticism of this specification is that in many situations individuals may choose social networks that they belong to or influence the state of their social networks through their actions. In other words, social networks are determined endogenously based on individuals' actions rather than given exogenously.

2.2 Social networks as endogenously chosen

Arcidiacono and Nicholson (2005) examined peer effects among medical students on whether the abilities and specialty preferences of a medical schools class can affect individual students' academic performance and choices of specialties. One of the issues they faced was the selection problem, or endogenous peer groups, because a student' choice of medical schools may be influenced her unobserved ability. They mitigated this problem by using a fixed-effects model, which yielded an upper bound on peer effects. They did not find any evidence for peer effects along racial lines nor did they find spillover effects between low and high ability students. Peer effects were mainly found along gender lines.

Kohler, Behrman, and Watkins (2006) analyzed the determinants of individuals' risk perceptions of HIV infection and how these perceptions led to the adoption of new behaviors by the individuals. Their central argument was that social networks were not likely to be random. Rather, individuals chose their networks based on their risk perceptions of HIV infection. The networks in turn affected the risk perceptions. Thus, Kohler, Behrman, and Watkins allowed the feedback between the size and composition of social networks and risk perceptions over time. In their case, social networks are the partners with whom the individuals have talked about AIDS. The results indicated that social networks had substantial effects on risk perceptions as well as the adoption of certain behaviors.

Ryan and Tucker (2007) studied the adoption of network technologies while incorporating network effects and forward-looking expectations. In their empirical application, employees adopted a video conferencing technology within a firm. The benefits of such adoption depended on how many others have already adopted as well as how one's decision to adopt the technology will affect the adoption behavior of others. Therefore, Ryan and Tucker modeled this decision problem as an optimal waiting game, where employees were forward looking with respect to the network evolution of the technology adoption within the firm. They used a fully dynamic, structural model of network technology adoption and hence were able to perform policy experiments. They found that targeting the right type of employees for the initial adoption led to a larger network in a shorter period of time than uniformly rolling out the technology across different types.

Mayer and Puller (2008) used data from Facebook.com to study the formation of social networks on university campuses. The main objective of this paper was to identify the factors that led to the formation of a social tie between two students. The authors developed a reduced form, static model where the probability of a social tie between two students was modeled as a function of students' demographics and school outcome characteristics. They found that race was a strong predictor of social ties among demographic variables. If students shared the same major or participated in the same campus activities, they were likely to be friends with each other.

Among the articles mentioned above, only Ryan and Tucker adopted a fully dynamic and structural model approach to model network effects. Because the others used reduced form approaches, their estimation results cannot be used to perform policy experiments.

Chapter 3

Model

We propose a dynamic model of usage behavior and network effects in which one decides whether or not to use a social network site each day. The action space is denoted by $A = \{0, 1\}$ such that $a_{it} = 1$ implies that a member i used a social network site at day t and $a_{it} = 0$ implies otherwise. If a member uses a social network site, or $a_{it} = 1$, she derives per period utility u_{it} from social networking but also incurs a cost c_{it}. The utility from the outside option, or $a_{it} = 0$, is normalized to zero and a member does not incur any cost in that case.

The objective of the individual i at time t is to maximize the discounted sum of expected utilities minus costs through the choice of $\{a_{it}\}$ over the infinite horizon:

$$\underset{\{s_{t+1}, s_{t+2}, \ldots\}}{E} \left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} a_{i\tau} (\tilde{u}_{i\tau} - c_{i\tau}) \,| s_{it} \right], \tag{3.1}$$

where β is a discount factor, s_{it} is the state of social network, and $E(\cdot)$ is the expectation operator. The state space is denoted by $S = \{s_{it}\}$ and consists of factors that affect the state of a social network, which determines the amount of social interactions a member expects receive conditional on site usage. As shown below, the expected utilities at period t are conditional on the state at period t such that $\tilde{u}_{it} = E[u_{it}|s_{it}]$. The expectation operator $E(\cdot)$ is with respect to the time paths of future states with conditional state transition probabilities $p(s_{i,t+1}|s_{it}, a_{it})$. It is important to note that state transitions are conditional on one's actions. One can affect future states of a social network by maintaining the existing stock of social capital and investing in the creation of new social capital.

3.1 Per period utility

Members derive utility from social interactions with friends through features of social network sites. MySpace.com allows members to interact with others in two major ways: real-time chat and messaging. Real-time chat is a system that allows members to type back and forth instantly with friends who are currently online. Messaging is similar to an electronic mail that allows members to post messages to friends' Web pages.

In our model, per period utility has a linear-in-parameters form:

$$u_{it} = \theta_{i1} + x_{it}\theta_{i2} + \tilde{x}_{it}\theta_{i3} + \epsilon_{it}, \tag{3.2}$$

where x_{it} and $\tilde{x}_{it}$ are the amounts of real-time chat and messaging that member i receives at time t respectively. The error term ϵ_{it} is assumed to have zero mean and to be uncorrelated with x_{it} and $\tilde{x}_{it}$. Parameters, θ_{i1}, θ_{i2}, and θ_{i3}, are group-specific based on demographics such that $\theta_{i1} = \theta_{g1}$, $\theta_{i2} = \theta_{g2}$, and $\theta_{i3} = \theta_{g3}$ where $g = 1, \ldots, G$ denotes demographic groups.

In reality, members do not observe the amounts of real-time chat and messaging prior to logging in. Thus, they form expectations about x_{it} and $\tilde{x}_{it}$ based on the states of their social networks s_{it}:

$$\tilde{u}_{it} = E(u_{it}|s_{it}) = \theta_{i1} + E[x_{it}|s_{it}]\theta_{i2} + E[\tilde{x}_{it}|s_{it}]\theta_{i3} + E[\epsilon_{it}|s_{it}], \tag{3.3}$$

where $E[x_{it}|s_{it}]$ is the expected number of friends who are currently online and $E[\tilde{x}_{it}|s_{it}]$ is the expected number of incoming messages.[1] By assumption, $E[\epsilon_{it}|s_{it}] = 0$ and ϵ_{it} can be interpreted as a prediction error.

3.2 State space

The state space consists of all variables that affect the amounts of social interactions that a member expects to receive from using a social network site on a given day. In our context, state variables include network size, last period's action, number of messages seen last period, day of the week, and demographics. Formally, let $s_{it} = (z_{it}^1, z_{it}^2, z_{it}^3, z_{it}^4, z_{it}^5)$, where $z_{it}^1, z_{it}^2, z_{it}^3, z_{it}^4$ and z_{it}^5 are network size, last period's action, the number of messages seen last period, day of the week, and demographics respectively. Specifically, z_{it}^1 is the number of friends in one's social network, $z_{it}^2 = 1$ if one used a social network site last period or 0 otherwise, z_{it}^3 is the number of messages seen if a member used a social network site last period, z_{it}^4 are dummy variables for Monday through Sunday, and z_{it}^5 includes demographic variables such as age, gender

[1] It is implied that members of a social network site do not get utilities directly from state variables, such as network size, only indirectly through its effect on $E[x_{it}|s_{it}]$ and $E[\tilde{x}_{it}|s_{it}]$.

and country of residence.[2][3]

3.3 Network effects

Network effects are embodied in one's expectation about the amounts of social interactions upon site usage; this expectation is conditional on the state of a social network. We model two distinct network effects that may be present in social networking: maintenance effect and investment effect. We view a social network as a stock of capital that yields a flow of utilities over time by creating social interactions between the owner and her friends. Maintenance effect relates to "health" of an existing capital stock. In essence, better managed stock of social capital will yield larger dividends of social interactions. If one uses a social network site this period, she can send a signal to friends about her commitment to friendship. Upon receiving the signal this period, her friends will reciprocate friendship by using a social network site with greater intensity next period. Thus, one can prevent depreciation of existing capital stock through maintenance effect.[4] The investment effect relates to the "size" of the capital stock. In essence, a larger stock of social capital will yield larger dividends of social interactions. If one uses a social

[2]A member does not observe network size unless she used a social network size last period. However, MySpace.com notified each member any changes in her social network via emails. Thus, we implicitly assume that members do keep track of any changes in their social networks even if they did not log in last period.

[3]We assume that messages seen on a given period are replied in that period. Thus, 'messages seen last period' serves as a proxy for 'messages sent last period'. If members sent messages last period, then they would expect replies this period.

[4]This network effect captures state dependence in one's actions across time periods.

network site this period, one may acquire friends through social networking and thus create new social capital, which will yield more social interactions in the future.

We account for the maintenance effect by including last period's action as a state variable that affects an expectation about the amount of social interactions that a member expects to receive upon logging in to a social network site. The action to engage in social networking this period serves as a signal to friends about one's commitment to friendship. This signal will in turn encourage friends to reciprocate friendship by logging in to a social network site next period. Hence, if a member logs in to a social network site this period, she will expect to receive a larger amount of social interactions next period than otherwise. We account for the investment effect by including network size as a state variable that affects an expectation about the amount of social interactions that a member expects to receive upon logging in to a social network site. By engaging in social networking this period, a member can make new friends who will interact with her in the future. If a member acquires friends this period by using a social network site, she will expect to receive a larger amount of social interactions next period than if she did not make any friends this period.

There is an important distinction between the maintenance and investment effects: Although both effects result from the same action to use a social network site, the maintenance effect affects social interactions for a single period following the action, whereas the investment effect can affect social interactions for multiple periods as long as the relationship with new friends persists over time.

3.4 Cost

A member incurs a cost each time she uses a social network site. Denote the cost of usage by c_{it}. This cost includes the time spent locating a computer station, the time spent for socializing, and any financial costs incurred using a social network site on a given day. We model the cost of usage to be time-varying between weekdays and weekends and formulate it as

$$c_{it} = c_i' + c_i'' + \nu_{it}, \tag{3.4}$$

where c_i' is the cost paid across all days of the week and c_i'' is a weekend premium paid only on weekends. We assume that c_{it} is continuously distributed and known up to scale; it is independently and identically distributed across members and time periods. Specifically we assume that ν_{it} has a standard normal distribution. The reasoning behind the inclusion of weekend premium is to examine how a relaxed time constraint or accessibility to the Internet on weekends affects usage decisions of members.[5]

[5] We account for weekend premium by including a weekend dummy in per period utility instead. As we explain later, this simplifies the estimation procedure.

3.5 State transition

State transitions represent beliefs about the evolution of state variables between time periods. Among the state variables, transitions in last period's action are determined by one's action in that period. Demographics are time-invariant or at least fixed in the short run. The day of the week is time-variant, but it follows a deterministic path. The key issue in representing state transitions is to model the transitions of messages seen last period and network size conditional on the other state variables and actions. Messages seen last period is conditional on last period's action as well as on the number of incoming messages, which we model probabilistically conditional on the state variables. We model transitions of network size conditional on actions and states. The beliefs about transitions in messages seen last period and network size are represented with a probability distribution $p(s_{i,t+1}|s_{it}, a_{it})$.

3.6 Policy function

We use the dynamic programming approach and rewrite the optimization problem in Equation (3.1) using the Bellman's equation:

$$\begin{aligned} V(s_{it}, c_{it}) \quad = \quad & \max_{a_{it}\in\{0,1\}} [\tilde{u}_{it} - c_{it} + \beta E[V(s_{i,t+1}, c_{i,t+1})|s_{it}, a_{it} = 1], \\ & \beta E[V(s_{i,t+1}, c_{i,t+1})|s_{it}, a_{it} = 0]], \end{aligned} \tag{3.5}$$

where the expectation operator $E(\cdot)$ is taken over state transitions with a conditional distribution $p(s_{i,t+1}|s_{it}, a_{it})$.

We integrate Equation (3.5) over c_{it} to get the modified Bellman equation:

$$\begin{aligned}\tilde{V}(s_{it}) &= F(\bar{c}(s_{it}))(\tilde{u}_{it} - E[c_{it}|c_{it} < \bar{c}(s_{it})] + \beta E[\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 1]) \\ &\quad +(1 - F(\bar{c}(s_{it})))\beta E[\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 0], \qquad (3.6)\end{aligned}$$

where $\bar{c}(s_{it})$ is the cutoff cost such that $a_{it} = 1$ if $c_{it} < \bar{c}(s_{it})$ and $a_{it} = 0$ if $c_{it} \geq \bar{c}(s_{it})$.

Denote a member's strategy or policy function whether to use a social network site or not at any day by σ_{it}. One optimally employs a cutoff strategy such that $\sigma_{it} = 1$ if and only if

$$c_{it} \leq E[\tilde{u}_{it} + \beta\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 1] - E[\beta\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 0]. \qquad (3.7)$$

Equation(3.7) constitutes the necessary and sufficient condition for a member to use a social network site.

Chapter 4

Data

4.1 Data sources

We randomly selected a group of members of MySpace.com and collected data by tracking their Web sites daily for four weeks from mid-January to mid-February of 2008. We focused on members who visited the site for non-business purposes; we excluded artists or companies who used the site for self-promotion or business purposes. Some members kept their profiles private restricting people out of their social networks from viewing their Web sites. Hence, we only had access to members whose profiles were open to the public.[1] Finally, we focused on college-age members, i.e., those who are between 19 and 23 years old.

About 15% of members in the initial sample were dropped from the final sample according to the following criteria. First, we dropped members who switched their profiles from public to private during the data collection period. Second, we dropped members who had less than 51 or more than 200 friends. This was due to the scarcity of observations outside this range. In

[1] This may give rise to some selection issues for our data. Hence, our results only apply to members whose profiles are open to the public.

addition, this criterion allows us to reduce the size of state space. Finally, we dropped members who exhibited extreme behavior, defined as losing or gaining more than 10 friends on any given day. The final data set consists of 111 members.

We recorded three types of variables for each member of our sample on a daily basis: usage behavior, social interactions, and the evolution of a social network. Usage behavior is defined as one's daily decision to use a social network site. This definition of usage behavior closely relates to the advertising revenue of social network sites and thus is of direct managerial relevance. MySpace.com automatically updates the last time each member used the Web site on a real-time basis. Hence, the accuracy of this variable is highly reliable.

Members of MySpace.com can interact with each other in two major ways. First, members, who are simultaneously online, can chat with others on a real-time basis by typing back and forth instantly. We call this feature 'real-time chat.' Second, members may exchange messages with others. We call this feature 'messaging.' Unlike real-time chat, messaging does not require members to be simultaneously online to interact with each other. One can simply leave messages on others' Web sites or receive messages from others while offline. MySpace.com informs each member of which friends are currently online. We recorded the number of friends online for each member of our sample at the same time of each day of the data collection period. This measure serves as a proxy for the opportunity for real-time chat. MySpace.com also posts messages on each member's Web site. We recorded the number of incoming messages for each member of our sample each day of the sample period.[2]

We traced the evolution of a social network by recording the number of friends for each member each day of the sample period. A member may gain or lose friends over time. Thus, one's social network can grow or shrink over time.

Tables A.1-A.2 describe the state space used in our paper. Tables A.3 provides the definitions of site usage, social interactions, and network evolution. Table A.4 summarizes usage behavior, social interactions, and network evolution of our sample. The average rate of daily usage for our sample was about 52%[3] The average number of friends who were online at the time of sample collection was about 4. The table also shows that members do not receive many messages on their Web pages with the average number of daily messages of about 0.26. This may be due to the availability of other ways of exchanging electronic mails.

4.2 Descriptive statistics

Table A.5 shows summary statistics on site usage and social interactions across demographic groups. We observe noticeable differences in site

[2]We do not observe outgoing messages. Hence, we assume that the number of incoming messages is proportional to that of outgoing messages.

[3]According to an internal survey by Facebook in September 2005, about 60% of college students logged in to the site daily.

usage between men and women: women use a social network site more frequently than men do. We also observe noticeable differences in site usage between US residents and non-US residents. US residents use a social network site more frequently than non-US residents do. The average number of friends online at a given point of time ranges from 3 to 4, which does not vary much across demographic groups. We note that the average number of incoming messages per day is relatively few, which is about 0.25 across demographic groups. To sum, we observe noticeable differences in site usage across demographic groups, whereas the amounts of social interactions do not vary much across demographic groups.

Table A.6 shows summary statistics on network evolution, daily changes in network sizes, across demographic groups. We observe that members do not gain or lose friends most of the time. However, the probability of network expansion is about 3% higher than that of network shrinkage across demographic groups. This may be due to the fact that our sample consists of members with more than 50 and less than 200 friends and at the lower end of this spectrum there is more room for network expansion.

Our model provides two major implications with respect to network effects that we can assess with simple analyses of the data. The maintenance effect implies that by using a social network site today, one sends a signal about her commitment to friendship, which will in turn encourage friends to reciprocate friendship by logging in to a social network site next period. Thus, we expect that one's action to log in to a social network site this period will be positively correlated with the amounts of real-time chat and messaging she will receive next period. The investment effect implies that larger social capital will yield a larger dividend of social interactions. Thus, we expect that network size will be positively correlated with the amounts of real-time chat and messaging.

Table A.7 shows the results of a Poisson regression of real-time chat on state variables. As we expected, the coefficient of last period's action and network size are estimated to be positive and statistically significant. Table A.8 shows the results of a Poisson regression of messaging on state variables. As we expected, the coefficients of last period's action and network size are estimated to be positive and statistically significant.

It is worth noting that although Table A.5 shows lack of differences in real-time chat across demographic groups, demographic variables turned out to be statistically significant in Table A.7. Table A.5 displays unconditional means of site usage and social interactions across demographic groups. Table A.7 tells us that there are differences in real-time chat across demographic groups conditional on network size.

Chapter 5

Estimation

We adopt a two-step estimation algorithm outlined by Bajari, Benkard, and Levin (2007; henceforth BBL).[1] In the first-step, policy functions and state transitions are estimated. In the second step, structural parameters of per period utility and cost are estimated using the optimality conditions. We describe each step in turn.

5.1 First-step estimation

The goal of the first-step estimation is to recover policy functions, σ_{it}, and state transition probabilities, $p(s_{i,t+1}|s_{it}, a_{it})$ as functions of state variables. To recover policy functions, we first estimate the choice probabilities, $p(a_{it} = 1|s_{it})$, as a flexible function of state variables using a probit regression:[2]

$$p(a_{it}|s_{it}) = f(z_{it}^1, z_{it}^2, z_{it}^3, z_{it}^4, z_{it}^5). \tag{5.1}$$

[1]See Rust (1987), Hotz and Miller (1993), Hotz, Miller, Sanders, and Smith (1994), Keane and Wolpin (1994), Magnac and Thesmar (2002) for estimation of dynamic discrete choice models. See Aguirregabiria and Mira (2007), Pakes, Ostrovsky, and Berry (2007), and Pesendorfer and Schmidt-Dengler (2007) for estimation of dynamic games.

[2]The state space consists of network size, day of the week, last period's action, messages seen last period, age , gender, and country of residence. Hence, the size of the state space is $150 \times 7 \times 2 \times 3 \times 5 \times 2 \times 2 = 126,000$.

Let $v(a_{it}, s_{it})$ denote choice-specific value functions net of the current period's cost such that

$$v(1, s_{it}) = E[\tilde{u}_{it} + \beta \tilde{V}(s_{i,t+1}) | s_{it}, a_{it} = 1] \tag{5.2}$$

and

$$v(0, s_{it}) = E[\beta \tilde{V}(s_{i,t+1}) | s_{it}, a_{it} = 0]. \tag{5.3}$$

It follows that

$$p(a_{it} = 1 | s_{it}) = \Phi\left(v(1, s_{it}) - v(0, s_{it})\right), \tag{5.4}$$

where $\Phi(\cdot)$ denotes a cumulative standard normal distribution.[3] Therefore, and we can invert the estimated choice probabilities to recover differences in the choice-specific value functions and hence policy functions across states.

We estimate the transition probabilities between states, $p(s_{i,t+1} | s_{it}, a_{it})$, as a flexible function of state variables. The key issue in representing state transitions is to model transitions of messages seen last period and network size probabilistically conditional on the other state variables and actions. Our data show that one's social network can grow or shrink in size between periods. Hence, we first estimate the probabilities of any nonnegative change in network size, $\Delta z_{it}^1 = z_{i,t+1}^1 - z_{it}^1$, with a probit regression:

$$p(\Delta z_{it}^1 \geq 0 | a_{it}, s_{it}) = g_1(z_{it}^1, z_{it}^2, z_{it}^3, z_{it}^4, z_{it}^5). \tag{5.5}$$

By definition, $p(\Delta z_{it}^1 < 0 | a_{it}, s_{it}) = 1 - p(\Delta z_{it}^1 \geq 0 | a_{it}, s_{it})$.Note that the probability of acquiring new friends depend on one's action as well as state. This is

[3]Since we normalize c_i' to zero and included a weekend dummy in per period utility to account for c_i'', c_{it} has a standard normal distribution.

due to the fact that if one does not socialize by logging in to a social network site, one cannot meet new friends online. Then we estimate the number of friends lost or gained with an ordered probit regression conditional on either nonnegative or negative change in network size:

$$
\begin{aligned}
p(\Delta z_{it}^1 | a_{it}, s_{it}, \Delta z_{it}^1 \geq 0) &= g_2(z_{it}^1, z_{it}^2, z_{it}^3, z_{it}^4, z_{it}^5) \\
p(\Delta z_{it}^1 | a_{it}, s_{it}, \Delta z_{it}^1 < 0) &= g_3(z_{it}^1, z_{it}^2, z_{it}^3, z_{it}^4, z_{it}^5) \qquad (5.6)
\end{aligned}
$$

The transition probabilities between states are the product of the Equations (5.5) and (5.6).

Finally, we estimate state transitions of the number of incoming messages, $\breve{z}_{it}^3$, with an ordered probit regression:

$$
p(\breve{z}_{it}^3 | s_{it}) = g_4(z_{it}^1, z_{it}^2, z_{it}^3, z_{it}^4, z_{it}^5). \qquad (5.7)
$$

Note that the number of incoming messages at any period does not depend on the current period's action. This is due to the fact that messages can arrive in one's MySpace Web page while she is offline. However, messages seen last period, z_{it}^3, is positive only if members logged in to the Web site to see incoming messages. In other words, $z_{it}^2 = 0$ implies $z_{it}^3 = 0$.

5.2 Computing $E(x_{it}|s_{it})$ and $E(\tilde{x}_{it}|s_{it})$

Members do not see realizations of real-time chat (x_{it}) and messaging ($\tilde{x}_{it}$) unless they logs in to a social network site. Thus, they must form expectations about the amounts of social interactions they will receive prior to using a social network site conditional on the state of their social networks. Therefore, we model these expectations $E(x_{it}|s_{it})$ and $E(\tilde{x}_{it}|s_{it})$ as flexible functions of state variables. To do so, we run Poisson regressions for both real-time chat and messaging with state variables as right-hand side variables.[4]

$$\begin{aligned} E(x_{it}|s_{it}) &= h_1(z_{it}^1, z_{it}^2, z_{it}^3, z_{it}^4, z_{it}^5) \\ E(\tilde{x}_{it}|s_{it}) &= h_2(z_{it}^1, z_{it}^2, z_{it}^3, z_{it}^4, z_{it}^5) \end{aligned} \quad (5.8)$$

These estimates are used to compute expectations of per period utilities, $\tilde{u}_{it}(s_{it})$, of using a social network site conditional on state variables.

5.3 Estimating value functions

We need to compute choice-specific value functions in Equations (5.2) and (5.3) to construct the optimality conditions for the second-step estimation. The estimates of value functions are conditional on actions, states and parameter values. We use a forward-simulation procedure outlined by BBL to get the estimates of choice-specific value functions: $\hat{v}(0, s_{it})$ and $\hat{v}(1, s_{it})$. We describe this forward-simulation procedure in detail in Appendix B.

5.4 Second-step estimation

The goal of the second-step estimation is to recover structural parameters of per period utility and the cost function using the optimality conditions

[4]Instead of using day of the week, we use a dummy variable that indicates whether a period of action is a weekday or a weekend.

in Equation (3.7):

$$c_{it} \leq E[\tilde{u}_{it} + \beta\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 1] - E[\beta\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 0].$$

It follows that

$$\begin{aligned} p(a_{it} = 1|s_{it}) &= p(c_{it} \leq E[\tilde{u}_{it} + \beta\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 1] - E[\beta\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 0]) \\ &= \Phi(E[\tilde{u}_{it} + \beta\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 1] - E[\beta\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 0]),\end{aligned} \tag{5.9}$$

where $\Phi(\cdot)$ is a cumulative normal distribution with mean $c_i' + c_i''$ and unit variance.

We use a maximum likelihood estimation to recover structural parameters of per period utility and cost by running a probit regression with actual actions as a dependent variable and expected per period utility and net present value of future returns as righthand side variables:

$$LL(\theta) = \sum_i \sum_t \{a_{it} \ln p(a_{it} = 1|s_{it}) + (1 - a_{it}) \ln p(a_{it} = 0|s_{it})\}, \tag{5.10}$$

where $LL(\theta)$ is the log-likelihood function to be maximized and $p(a_{it} = 1|s_{it})$ is given in (5.9).

5.5 Identification

In general, a discount factor, β, is not identified in dynamic discrete choice models.[5] Hence, we set $\beta = 0.98$. The parameters for real-time chat

[5]See Rust (1987).

and messaging are identified from the variation of these variables across state variables. Individual characteristics, such as demographics, cancel out if not interacted with variables that can vary according to the login decision. Thus, we interacted demographics with real-time chat and messaging to identify the effects of individual characteristics on daily site usage. The mean of the cost has two components, c_i' and c_i''. The parameter, c_i', cannot be separately identified from the intercept of per period utility, so we normalize it to zero. We include a weekend dummy variable in per period utility to account for a weekend premium c_i''. This simplifies the forward-simulation procedure, because now we can assume that c_{it} has a normal distribution with a zero mean and a unit variance. Consequently, we can only recover relative differences in costs across states, not absolute magnitudes of the costs across states. Fixing the variance of c_{it} to 1 identifies structural parameters in the second-step estimation.

Chapter 6

Results

The second-step estimation of our model recovers structural parameters in per period utility and cost. The primary goal of this estimation is to measure the effects of features of a social network site on usage behavior of members. The secondary goal of this estimation is to examine any idiosyncrasies of usage behavior across demographic groups.

Table A.9 provides the results of the probit regression that estimates the effects of expected amounts of real-time chat, expected amounts of messaging, and weekend premium on daily site usage.[1] The effects of features of MySpace.com are estimated to be positive and statistically significant at the 1% level:[2] Real-time chat, expected number of friends online, has a positive effect on usage of MySpace.com. Messaging, expected number of messages per day, also has a positive effect on usage of MySpace.com. The results indicate that features of MySpace.com achieve the intended goal of generating site traffic. Weekend premium was estimated to be positive and but not statistically

[1]Expected amounts of real-time chat and messaging are computed conditional on state variables.

[2]Standard errors for parameter estimates were computed based on the second-step estimation only; they are not adjusted for the first-step estimates.

significant at the 10% level.[3] Although the p-value for weekend premium is about 0.15, the positive sign for weekend premium can be explained in two different ways: First, members face a relaxed time constraint on weekends than on weekdays. If a social network site is a normal good in that a relaxed time constraint leads to the higher consumption of it, a weekend premium must be negative. On the other hand, if a social network site is an inferior substitute for face-to-face meetings, then one will engage in traditional friendship rather than online meeting given extra time. Second, some members may have a limited access to the Internet on weekends. If so, weekend premium must be positive. Our results cannot distinguish between the two effects; however, it still has managerially meaningful implications in terms of time of targeted advertising or developing new business policies.

Table A.10 provides the results of the probit regression that includes both the main effects of features of MySpace.com and a weekend dummy and their interactions with demographic variable. We do not interact real-time chat or weekend dummy with demographic variables, because doing so did not improve the fit of the model.[4]

We observe differences in base rates of site usage across demographic groups. The interaction between constant and age are estimated to be negative and statistically significant at the 1% level. The members of our sample are

[3]Positive weekend premium is equivalent to saying usage cost is higher on weekends than on weekdays.

[4]The log-likelihood show little improvement from −1844.93 to −1844.11 when we allowed interaction of real-time chat and weekend with demographic variables.

between 19 to 23 years old. If we assume that our sample mostly consists of college students, we may interpret this result such that the cost of site usage increases as members move towards graduation. For instance, college seniors may have more tasks to perform such as preparing for graduation and looking for jobs than college freshmen. The interaction between constant and country of residence are estimated to be positive and statistically significant at the 1% level, implying that non-US residents may incur lower cost of usage than US residents. Non-US residents in our sample mostly reside in major cities of UK and Australia: given that the rate of internet penetration is higher in Europe than in US, the higher costs for non-US residents is difficult to explain in terms of accessibility to the Internet. Rather, it can be explained by the lower cost for arranging face-to-face meeting due to higher population density and the better public transportation system in Europe than in US.

We also observe differences in the effects of messaging on site usage across demographic groups. The interaction between messaging and age are estimated to be positive and statistically significant at the 1% level. The interaction between messaging and country of residence are estimated to be negative and statistically significant at the 10% level. These results can be interpreted similarly to our reasoning behind different base rates of site usage across demographic groups. More tasks and stricter time constraint may lead to higher utility from messaging for college seniors compared to college freshmen. The lower cost of face-to-face meeting may reduce the desirability of messaging for non-US residents compared to US residents.

In summary, we found that features of MySpace.com, real-time chat messaging, achieve the intended goal of generating site traffic. However, members of MySpace.com exhibit heterogeneity with respect to their demographics. In the next section, we use the parameter estimates to explore counterfactual policies that may help social network sites to develop features that can increase site traffic.

Chapter 7

Policy Experiments

We perform three types of policy experiments with the goal of providing managers with business policies that may enhance firm performance by increasing site traffic. The parameter estimates from the second-step estimation are used to simulate the following counterfactual policies.

1. Prevention of coordination failure
2. Enhancement of networking experience
3. Matching service

The first policy is designed to facilitate online meetings among friends and hence prevent coordination failure. Coordination failure occurs when two parties cannot interact with each other despite the mutual desires to do so. A social network site can mitigate coordination failure by devising a mechanism that alerts members when their friends log in to the Web site. One way to achieve this objective is to update members of their friends' online status through devices such as mobile phones or electronic mails.

The second policy is designed to make social networking online a more pleasurable experience. One way to achieve this objective is to invent features that raise the utility of social networking on the Web site. For instance, a social network site can introduce new video features or provide new graphic icons for real-time chat. As recently as in 2008, MySpace.com introduced capabilities that allow members to engage in video conferences with friends by incorporating Skype into its Internet Messenger service.

The third policy is designed to assist the acquisition of new friends. One way to achieve this objective is to adopt collaborative filtering mechanisms or survey techniques to match members based on some criteria. Many online retailers like Amazon.com use collaborative filtering systems to recommend products to customers based on their purchase histories. Online dating services like Match.com and Eharmony.com use detailed surveys to match members.

We performed policy experiments by simulating the paths of actions and states for 240,000 hypothetical members evenly spread across all states over a period of 52 weeks. The first policy is implemented by multiplying real-time chat, the number of friends online, by 1.20 at each state. The second policy is implemented by multiplying the estimated coefficient for real-time chat by 1.20. The third policy is implemented by increasing the probability of acquiring new friends at each state by 10 percent. There is a caveat to be mentioned here: When real-time chat enters linearly in per period utility, the changes in the first and second policies are proportional. In other words, a 20 percent increase in the first policy will have the same effect on site traffic as a 20 percent increase in the second policy. If the utility function is nonlinear in real-time chat, then these two policies may not have this proportional relationship.

We use two metrics to compare the effectiveness of the three policies on firm performance with that of the baseline policy; the baseline policy represents the status quo of MySpace.com. First, we divide 52 weeks into four quarters and compute the average of daily site traffic for each of the four quarters. Second, we compute the average of network size for each of the four quarters. The first metric is directly related to the revenue source of social network sites—advertising revenue per thousand views. The second metric to some extent measures well-being of members given that social network sites are in part used to make new friends.

Table A.11 shows the average of daily usage rates for the four policies—baseline policy, prevention of coordination failure, enhancement of networking experience, and matching service—for each of the four quarters. After 52 weeks all three counterfactual policies outperformed the baseline policy in daily site usage: Whereas the baseline policy has 55.28 percent of usage rate, the three policies have 56.46, 56.48, and 57.31 percents of daily usage rates respectively.

However, the trend of usage rates of these policies over time shows that the short-run effects of the policies may differ from the long-run effects. The first two policies outperformed the third on daily usage rates for the first two quarters. By the end of the third quarter, the third caught up the first two on daily usage rates and in the fourth quarter, the third policy had the highest usage rate among the three counterfactual policies.

The reason behind the discrepancy between the short-run and long-run effects can be explained as follows: The first two policies have immediate effects on usage rate by increasing the amounts of real-time chat and enhancing networking experience for real-time chat at the point of adoption. In contrast, the third policy has a rather gradual effect on daily usage rate by expanding the average size of social networks among members, which in the long run translated into higher site usage. Table A.12 shows that in the fourth quarter, the average network sizes for the first two policies were nearly identical to that of the baseline policy. In contrast, the average network sizes for the third policy was larger by about 15 than that of the baseline policy in the fourth quarter.

A careful cost-benefit analysis must precede the adoption of these policies. The provision of new graphic icons for real-time chat may require one-time investment, whereas mobile alerts or collaborative filtering may incur on-going expenses. In addition, the cost of maintenance may increase for social network sites due to higher site traffic as a result of these policies.

Chapter 8

Conclusions

This dissertation structurally estimates a dynamic model of usage behavior and network effects in social network sites using data from MySpace.com. We incorporate two distinct network effects that make social networking an inherently dynamic process: (1) one can manage an existing base of friends through social networking and thus prevent depreciation of capital stock (maintenance effect), and (2) one may acquire new friends through social networking, which results in creation of new capital stock (investment effect). Our goal is to estimate the effects of features of a social network site on members' daily site usage using the data from MySpace.com and provide managers with business policies that may increase site traffic when implemented. To this end, the estimates from our model are used to evaluate three types of counterfactual policies.

The estimation results quantify how features of MySpace.com affect a member's usage decision. We found that (1) features of MySpace.com positively affect site traffic but (2) their effects may differ significantly across demographic groups. Our counterfactual analysis shows that the three policies that we proposed—prevention of coordination failure, enhancement of networking experience, and matching service—affect daily site usage decisions differently over a time horizon and these results are mitigated by both maintenance and investment effects.

We build on the existing "social effects" literature by structurally modeling social networking as a dynamic process based on the theory of utility maximization. Our approach is in contrast to those by Manski (1993) and Graham (2008) who model social networking as a static process using reduced-form models or to those by Mayer and Puller (2006) and Trusov, Bodapati, and Bucklin (2006) who adopt a dynamic approach in modeling social networking using reduced-form models. Our structural approach yields managerially informative results through counterfactual policies and thus we can provide managers with suggestions for shaping future business policies.

In future research we can focus on other managerially relevant topics related to social network sites. A unique aspect of advertising in social network sites is that consumers now receive information from both advertising messages and their online peers. Thus, it may be of interest to investigate the relative importance of these two information sources on consumers' purchase decisions.

Appendices

Appendix A

Tables

Table A.1: Dimensions of state space

Variable	**Dimension**
Network size	150
Last period's action	2
Messages seen last period	3
Day of the week	7
Age	5
Gender	2
Country of residence	2

Note 1. The maximum number of messages seen last period is truncated at 2, because members rarely received more than 2 messages per day. This truncation reduced the size of state space considerably.
Note 2. Country of residence has two categories: within and outside the US.

Table A.2: Summary statistics on state variables

Variable	**Mean**	**StdDev**	**Min**	**Max**
Network size	120.522	38.677	53	200
Last period's action	0.535	0.499	0	1
Messages seen last period	0.168	0.465	0	2
Day of the week	4.115	2.026	1	7
Age	21.108	0.500	19	23
Gender	0.505	1.365	0	1
Country of residence	0.595	0.491	0	1

Note 1. Day of the week goes from 1 (Monday) to 7 (Sunday).
Note 2. Gender is coded 1 if female and 0 if male.
Note 3. Country of residence is coded 1 if United States and 0 otherwise.

Table A.3: Definitions of site usage, social interactions, and network evolution

Variable	**Definition**
Site usage	1 if used a social network site on a given day, 0 otherwise
Real-time chat	Number of friends online
Messaging	Number of incoming messages
Network evolution	Gain/loss of friends between two days

Table A.4: Summary statistics on site usage, social interactions, and network evolution

Variable	**Mean**	**StdDev**	**Min**	**Max**
Site usage	0.527	0.499	0	1
Real-time chat	3.634	3.241	0	27
Messaging	0.257	0.753	0	11
Network evolution	0.042	0.456	-6	3

Note. The numbers for social interactions are observed values, not expected ones.

Table A.5: Summary statistics on site usage and social interactions across demographic groups

	Gender		Country		Age	
	Male	**Female**	**US**	**Non-US**	**$\leq$ 21**	**21 $<$**
Site usage	49.86%	55.66%	55.71%	48.38%	49.39%	49.89%
	(50.02)	(49.69)	(49.69)	(49.99)	(49.49)	(49.89)
Real-time chat	3.52	3.75	4.02	3.07	3.88	3.33
	(3.51)	(2.94)	(2.69)	(3.85)	(3.52)	(2.84)
Messages	0.24	0.27	0.26	0.26	0.25	0.27
	(0.73)	(0.77)	(0.77)	(0.73)	(0.65)	(0.86)

Note. Standard deviations are in parentheses.

Table A.6: Summary statistics on network evolution across demographic groups

Change in	Gender		Country		Age	
network size	**Male**	**Female**	**US**	**Non-US**	**$\leq$ 21**	**21 $<$**
≤ -3	0.34%	0.00%	0.17%	0.17%	0.13%	0.23%
-2	0.34%	0.56%	0.29%	0.68%	0.38%	0.54%
-1	3.30%	3.08%	2.62%	4.02%	3.53%	2.77%
0	88.74%	88.81%	88.99%	88.46%	88.40%	89.23%
1	6.11%	6.22%	6.29%	5.98%	6.31%	6.00%
2	0.69%	1.12%	1.34%	0.25%	0.88%	0.92%
$3 \leq$	0.48%	0.21%	0.29%	0.43%	0.38%	0.31%

Note 1. The changes in network size are the differences in network sizes between two adjacent days. Note 2. The percentages should be read columnwise or by each demographic group.

Table A.7: Poisson regression of real-time chat on the state variables
Dependent variable: Number of friends online.

RHS variable	**Estimate**	**Std. Error**
Network size	0.00985^{**}	0.00025
Last period's action	0.06464^{**}	0.02109
Messages seen last period	0.02543^{*}	0.01447
Weekend status	-0.05569^{**}	0.02143
Age	-0.12359^{**}	0.00736
Gender	0.10026^{**}	0.02032
Country of residence	0.37088^{**}	0.02144
Constant	-0.21160^{**}	0.03291
Log-likelihood = 6937.6356, $\Pr[>\chi^2] = 0.0000$		

(∗∗) significant at 1% level. (∗) significant at 5% level. (‡) significant at 10% level.
Note. Weekend status is coded 1 if on weekends, 0 if on weekdays.

Table A.8: Poisson regression of messaging on the state variables
Dependent variable: Number of incoming messages.

RHS variable	**Estimate**	**Std. Error**
Network size	0.00270^{**}	0.00096
Last period's action	0.78198^{**}	0.08948
Messages seen last period	0.34237^{**}	0.01890
Weekend status	-0.17034^{*}	0.08490
Age	-0.05872^{*}	0.02811
Gender	-0.08008	0.08035
Country of residence	-0.07695	0.08203
Constant	-2.05493^{**}	0.12456
Log-likelihood = -1796.2108, $\Pr[>\chi^2] = 0.0000$		

(∗∗) significant at 1% level. (∗) significant at 5% level. (‡) significant at 10% level.

Table A.9: Probit regression of site usage on features of MySpace.com
Dependent variable: Daily site usage

RHS variable	**Estimate**	**Std. Error**
Real-time chat	0.0866^{**}	0.0089
Messaging	0.8387^{**}	0.0744
Weekend	-0.0648	0.0477
Constant	-0.5638^{**}	0.0444

Log-likelihood $= -1887.9718$, $\Pr[> \chi^2] = 0.0000$

(∗∗) significant at 1% level. (∗) significant at 5% level. (‡) significant at 10% level.
Note. Real-time chat and messaging are expected values conditional on state variables.

Table A.10: Probit regression of site usage on features of MySpace.com interacted with demographic variables
Dependent variable: Daily site usage

RHS variable		**Estimate**	**Std. Error**
Main effects	Real-time chat	0.0500^{**}	0.0107
	Messaging	0.8154^{**}	0.2023
	Weekend	-0.0680	0.0482
	Constant	-0.2903^{**}	0.0837
Interacted with age	Messaging	0.1235^{**}	0.0588
	Constant	-0.1444^{**}	0.0225
Interacted with gender	Messaging	0.0608	0.1667
	Constant	-0.0228	0.0632
Interacted with country	Messaging	-0.3315^{*}	0.1765
	Constant	0.2679^{**}	0.0658

Log-likelihood $= -1844.9314$, $\Pr[> \chi^2] = 0.0000$

(∗∗) significant at 1% level. (∗) significant at 5% level. (‡) significant at 10% level.
Note. Real-time chat and messaging are expected values conditional on state variables.

Table A.11: The effects of counterfactual policies on daily site usage

Policy	**Qtr1**	**Qtr2**	**Qtr3**	**Qtr4**
Baseline policy	51.59	53.42	55.82	54.93
Prevention of coordination failure	53.46	54.86	57.51	56.38
Enhancement of networking experience	53.47	54.85	57.50	56.42
Matching service	51.42	53.73	57.17	56.84

Note 1. The numbers in cells are the rates of daily site usage in percentage.
Note 2. Each quarter consists of thirteen weeks.

Table A.12: The effects of counterfactual policies on network size

Policy	**Qtr1**	**Qtr2**	**Qtr3**	**Qtr4**
Baseline policy	75.71	76.09	76.54	76.99
Prevention of coordination failure	75.87	76.59	77.28	77.97
Enhancement of networking experience	75.88	76.60	77.28	77.96
Matching service	77.93	82.74	87.10	91.25

Note. The numbers in cells are average network sizes in number of friends.

Appendix B

Forward Simulation

This appendix describes the forward-simulation procedure that we use to estimate value functions. The goal of forward-simulation is to approximate modified value functions:

$$E[\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 1], \quad E[\tilde{V}(s_{i,t+1})|s_{it}, a_{it} = 0].$$

Step 1. We fix the initial state $s_{i0} = s$ for individual i at period $t = 0$. We need to compute modified value functions only for the states that can be reached in one period from the states that we observe in the data. In order for approximations to be accurate, we need to forward simulate actions and states for many simulation draws over many time periods. In practice, we set the number of simulated individuals to be 100 and the number of simulation periods to be 40.

Step 2. We simulate actions a_{i0} for individual i at period $t = 0$ using the estimated policy functions. As explained above if $v(a_{it}, s_{it})$ denote choice-specific value functions, then $a_{i0} = 1$ if and only if

$$c_{it} < v(1, s_{it}) - v(0, s_{it}) = \Phi^{-1}(p(a_{it} = 1|s_{it})),$$

where c_{i0} is the simulated from a standard normal distribution.

Step 3. If $a_{i0} = 1$, then we save $E(x_{it}|s_{it})$ and $E(\tilde{x}_{it}|s_{it})$ and $E[c_{i0}|c_{i0} < \bar{c}(s_{i0})]$. Since c_{it} has a standard normal density, $E[c_{i0}|c_{i0} < \bar{c}(s_{i0})])$ is the mean of a truncated standard normal density. Thus, we have

$$\begin{aligned} E[c_{i0}|c_{i0} < \bar{c}(s_{i0})] &= -E[-c_{i0}| - \bar{c}(s_{i0}) < -c_{i0}] \\ &= -\frac{\phi(-\bar{c}(s_{i0}))}{1 - \Phi(-\bar{c}(s_{i0}))} \\ &= -\frac{\phi(\bar{c}(s_{i0}))}{1 - (1 - \Phi(\bar{c}(s_{i0})))} \\ &= -\frac{\phi(\bar{c}(s_{i0}))}{\Phi(\bar{c}(s_{i0}))}. \end{aligned}$$

We can recover $\bar{c}(s_{i0})$ by inverting the probability of using a social network site at each state.

Step 4. We draw a new state s_{i1} using the estimated state transition probabilities $p(s_{i,t+1}|s_{it}, a_{it})$.

Step 5. We repeat Steps 1–4 for many periods.

By averaging discounted sum of utilities minus expected costs over simulated individuals for parameter values θ_i, we can compute $\hat{V}(s_{it}, \theta_i)$, an estimate of $\tilde{V}(s_{it}, \theta_i)$.

Bibliography

[1] Victor Aguirregabiria and Pedro Mira. Sequential estimation of dynamic discrete games. *Econometrica*, 75(1), 2007.

[2] Peter Arcidiacono and Sean Nicholson. Peer effects in medical school. *Journal of Public Economics*, 89(2-3), 2005.

[3] Patrick Bajari, Lanier Benkard, and Jonathan Levin. Estimating dynamic models of imperfect competition. *Econometrica*, 75(5), 2007.

[4] William A. Brock and Steven N. Durlauf. Discrete choice with social interactions. *Review of Economic Studies*, 68(2), 2001.

[5] William A. Brock and Steven N. Durlauf. A multinomial-choice model of neighborhood effects. *The American Economic Review: Papers and Proceedings*, 92(2), 2002.

[6] Edward Castronova. Social norms and sexual activity in u.s. high schools. *The Journal of Human Resources*, 39(4), 2004.

[7] Bryan S. Graham. Identifying social interactions through conditional variance restrictions. *Econometrica*, 76(3), 2008.

[8] Wesley R. Hartmann. Demand estimation with social interactions and the implications for targeted marketing. *Marketing Science*, forthcoming,

2008.

[9] V. Joseph Hotz and Robert A. Miller. Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies*, 60(3), 1993.

[10] V. Joseph Hotz, Robert A. Miller, Seth Sanders, and Jeffrey Smith. A simulation estimator for dynamic models of discrete choice. *Review of Economic Studies*, 61(2), 1994.

[11] Michael P. Keane and Kenneth I. Wolpin. The solution and estimation of discrete dynamic programming models by simulation and interpolation: Monte carlo evidence. *Review of Economics and Statistics*, 76(4), 1994.

[12] Hans-Peter Kohler, Jere R. Behrman, and Susan C. Watkins. Social networks, hiv/aids and risk perceptions. *Demography*, 44(1), 2006.

[13] Thierry Magnac and David Thesmar. Identifying dynamic discrete decision processes. *Econometrica*, 70(2), 2002.

[14] Charles F Manski. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3), 1993.

[15] Adalbert Mayer and Steven L. Puller. The old boy (and girl) network: Social network formation on university campuses. *Journal of Public Economics*, 92(1-2), 2008.

[16] Ariel Pakes, Michael Ostrovsky, and Steven Berry. Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *RAND Journal of Economics*, 38(2), 2007.

[17] Martin Pesendorfer and Phillip Schmidt-Dengler. Asymptotic least squares estimators for dynamic games. *Review of Economic Studies*, forthcoming, 2007.

[18] John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5), 1987.

[19] Stephen Ryan and Catherine Tucker. Heterogeneity and the Dynamics of Technology Adoption. unpublished manuscript, 2007.

[20] Michael Trusov, Anand V. Bodapati, and Randolph E. Bucklin. Determining Influential Users in Internet Social Networks. unpublished manuscript, 2006.

Vita

Dae-Yong Ahn was born in Seoul, Republic of Korea, the son of Byung-Jik Ahn and Hea-Kyung Rhee. He received the Bachelor of Arts degree in Economics from the University of Colorado at Boulder, the Master of Arts in Economics and the Master of Science in Statistics from the University of Iowa. He continued his education at University of Texas at Austin as a doctoral student in the field of Marketing.

Permanent address: Jugong Apartment 1010-402
Gwacheonsi, Kyunggido, Republic of Korea

This dissertation was typeset with LaTeX[†] by the author.

[†] LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.