

Network Effects and Personal Influences: The Diffusion of an Online Social Network

Zsolt Katona, Peter Pal Zubcsek and Miklos Sarvary*

August 7, 2010

*Zsolt Katona is Assistant Professor of Marketing at the Haas School of Business, UC Berkeley, Berkeley, CA 94720. Peter Pal Zubcsek is Assistant Professor of Marketing at University of Florida, Gainesville, FL 32611. Miklos Sarvary is Professor of Marketing at INSEAD, Bd. de Constance, 77305, Fontainebleau, France. E-mail: zskatona@haas.berkeley.edu, pzubcsek@ufl.edu, miklos.sarvary@insead.edu. Peter Pal Zubcsek's work was supported by the Sasakawa Young Leaders Fellowship Fund.

Network Effects and Personal Influences: The Diffusion of an Online Social Network

Abstract

We study the diffusion process in an online social network given the individual connections between members. We model the adoption decision of individuals as a binary choice affected by three factors: (1) the local network structure formed by already adopted neighbors, (2) the average characteristics of adopted neighbors (influencers), and (3) the characteristics of the potential adopters. Focusing on the first factor, we find two marked effects. First, an individual who is connected to many adopters has a higher adoption probability (degree effect). Second, the density of connections in a group of already adopted consumers has a strong positive effect on the adoption of individuals connected to this group (clustering effect). We also record significant effects for influencer and adopter characteristics. Specifically, for adopters, we find that their position in the entire network and some demographic variables are good predictors of adoption. Similarly, in the case of already adopted individuals, average demographics and global network position can predict their influential power on their neighbors. An interesting counter-intuitive finding is that the average influential power of individuals *decreases* with the total number of their contacts. These results have practical implications for viral marketing in a context where, increasingly, a variety of technology platforms are considering to leverage their consumers' revealed connection patterns. In particular, our model performs well in predicting the next set of adopters.

Keywords: Diffusion, Social Networks, Network Marketing.

1 Introduction

While the importance of word of mouth (WOM) in influencing consumer decisions has long been recognized by marketers, it has been difficult to measure its impact and to efficiently use it for commercial purposes until recently. Modern technology, however, has gradually transformed social interactions among people. On average, we communicate more using some sort of technology platform (phone, VoIP, e-mail or chat) and our frequent contacts are readily available to the owner of these communication platforms. With the emergence of Web 2.0 technologies, media such as blogs, instant messaging (e.g. msn.com) and social networking web-sites (e.g. Facebook.com) are becoming ubiquitous and they all provide a “map” of communication paths among their users. Social networking sites, such as MySpace or Facebook, are particularly interesting examples. They represent rich and popular communication interfaces for hundreds of millions of users. On these sites, users exhibit their demographics as well as their preferences by carefully editing (decorating) their profiles. More importantly, they explicitly link to their friends, in this way, revealing their likely communication patterns. In the last five years, social networks have populated the world, often attracting a critical proportion of a country’s population. Cyworld, for example, claims to count over 40% of Koreans among its membership base, whereas Facebook has over 400 million members worldwide and has recently surpassed Google in becoming the most visited website in the US.

The increased importance of technology platforms for social interactions has raised the interest of product marketers who seek to explore these as new advertising/promotion media. Indeed, social networks’ revenue models are primarily based on advertising, although, so far, the use of (mostly) banner-type advertising has produced disappointing results.¹ Increasingly, marketers believe that the efficient way of using social networks for marketing relies on harnessing WOM, by analyzing the network of members’ connections. For example, Google has recently filed a patent for an algorithm that identifies so-called “influencers” on social networks. Several other firms (e.g. Ildiro and Xtract) provide network analysis for the telecom sector as well as for various Web 2.0 platforms in order to assist in viral marketing campaigns. Indeed, social

¹See, for example, “Social graph-iti”, in *The Economist*, 20. October, 2007, pp.77, or “Google: Harnessing the Power of Cliques”, in *Business Week*, 6. October, 2008, pp.50.

networks represent but one area where network analysis might be used for WOM marketing. Other technologies (blogs, telecom, virtual worlds, etc.) that record consumers' communication patterns are also adapted to such techniques. In all these cases, the basic idea is that understanding the network structure of individual consumers can help implement effective viral marketing strategies.

The assumption underlying all these “network marketing” techniques is that network information can help identify influencers and predict consumers' adoption probabilities. The goal of this paper is to verify this assumption and to identify how the network structure drives adoption. Specifically, we develop and empirically test an individual level diffusion model that explicitly takes into account the micro-structure of interpersonal connections between potential adopters. Our empirical application studies the growth of a country-specific social network site, where membership can only be acquired by receiving an invitation from an existing member. In this context, we study the adoption process of network memberships at an early stage of the site's development. We assume that, gradually, the social network site replicates people's real-life social connection patterns. Thus, recording the site's membership network at a much later point in time, we can assume that it reflects people's real-life social networks quite accurately. Then, looking back in time, we observe a diffusion process of membership on this “ultimate” or “final” network.²

Our primary research objective is to uncover the effects that differences in individuals' connection patterns have on the diffusion process. Motivated by conceptual as well as practical considerations, we distinguish three factors that may affect a potential adopter's adoption decision. The first factor, that we call “network effects”, relates to the influence of the *structure* of connection patterns of the potential adopter's already adopted neighbors. The second factor, that we call “influencer effects”, refers to the (average) *individual characteristics* of already adopted network members on their not-yet-adopted neighbors. As opposed to the previous factor that concentrates on the impact of the structure, under influencer effects, we essentially measure the (average) influential power of every adopter. The average characteristics describing influencers

²It is important to realize that the present paper does not study a network growth or network formation process but, rather, a diffusion process on an already existing network of potential adopters. Of course, the results depend on the validity of the assumption that we have indeed captured the relevant social network by observing the membership network at a late stage. In our empirical application, we test this assumption in various ways (see Section 4.4.3).

may be demographics, such as age for example, but they may also include measures that describe the global network position of the influencers, i.e. their position in the “final” social network observed. Finally, the third factor, called “adopter effects” captures the effects of the adopter’s individual characteristics. As for influencers, these characteristics include both demographics as well as characteristics describing the adopter’s global network position. Depending on data availability and legislative constraints, marketers can use each of these categories of variables to identify potential marketing targets in order to influence the adoption process. They can target members of the network who have already adopted the promoted product or potential adopters who have yet to adopt. In both cases, potential marketing targets can be selected based on their individual characteristics or, alternatively, based on the local network structure surrounding them. Our model helps identify potential adopters by predicting the next set of adopters substantially better than models based solely on demographic variables.

In our choice of network variables, we rely on diffusion theory, as well as previous research in sociology. In particular, individuals who are related to many already adopted members may have higher adoption probability since their related partners can provide more information about the service or innovation in question, ultimately exercising greater joint influential power. We call this the *degree effect*, after the number of connections (degree) an individual has. In addition to the number of connections, the density of connections in a group of already adopted users may also affect the adoption of individuals being linked to the members of this group. A more tightly connected group should have a stronger influence on its members. We call this effect the *clustering effect*. We find strong empirical evidence supporting the presence of both the degree effect and the clustering effect, as well as a positive interaction between them.

Beyond network effects we also find strong influencer and adopter effects on the adoption probability of potential network members. For example, some demographic characteristics (e.g. age, gender) can predict adopters’ influential power on their neighbors. Similarly, demographic variables can also predict potential adopters’ propensity to adopt. More interestingly, variables that describe individuals’ position in the social network (e.g. the number of their contacts, how connected these neighbors are or the extent to which they are interconnecting parts of the network) are also good predictors of both influential power and adoption propensity. From both a

theoretical and practical perspective, a particularly interesting empirical result is that the average influential power of individuals is lower the larger their social network is. This suggests that a high number of friends dilutes the influential power that an individual has on each of his/her friends. We also find some evidence for that the same influential power is higher the more the actors occupy a “brokering” position among their contacts.

The rest of the paper is organized as follows. The next section relates our work to the literature and highlights our points of departure. Section 3 sets up the stochastic network-based diffusion model and introduces the network measures used. Section 4 presents the empirical analysis, including extensive validity tests. Finally, Section 5 discusses the results, concludes and highlights a few limitations.

2 Related Literature

Our work draws on two broad research streams: (1) the marketing literature on new product diffusion and (2), the sociology literature on social network analysis.

The large body of quantitative research on new product diffusion is based on models that, for the most part, ignore connection patterns between individuals. Like most of its generalizations, the famous Bass-model (1969) implicitly assumes that every consumer is connected to every other consumer, and estimates a uniform interpersonal influence (interpreted as WOM) on this (assumed-to-be) complete network. This typically also applies to diffusion models that attempt to take into account consumer heterogeneity.³ While these models allow for heterogeneous WOM effects, they still ignore the network structure of the adopter population. Given the central role of WOM communication for the diffusion process, there has been a clear call in the literature to incorporate the fine-grain structure of interpersonal connections in diffusion models (see Mahajan, Muller and Bass 1993, 1990). While various models have been developed to address this call (e.g., Goldenberg, Libai and Muller (2001, 2002) or Shaikh, Rangaswamy and Balakrishnan (2005), see Valente (2005) for a review), these could only be

³See e.g., Oren and Schwartz (1988) and Urban and Hauser (1993) for early papers and Goldenberg, Libai and Muller (2002) and Van den Bulte and Joshi (2007) for more recent applications.

tested on aggregate data.⁴

In recent years, various technological innovations have made it finally possible to access network data on the interpersonal relationships between consumers. Models to assess the impact of network characteristics on the diffusion process have finally been developed and used in empirical studies. For example, Godes and Mayzlin (2004) studied how WOM can be a driver of personal preferences in an environment where consumer communication via newsgroups is observed. Van den Bulte and Lilien (2001) analyzed the Medical Innovation data set which tracks the medical community's understanding of a new drug. To obtain the structure of WOM, they combined data from two overlaid relationship networks ("discussion" and "advice") to get a *social influence weight* matrix over the doctors whose behavior was recorded. They then estimated an individual-level diffusion model to demonstrate that earlier findings on social contagion over the same network were confounded with the marketing efforts (pricing and promotion) of the manufacturer. Nair, Manchanda and Bhatia (2010) studied physician prescription behavior and found that opinion-leaders in the physician's reference group may have a significant influence on the physician's behavior. In another domain, Hill, Provost and Volinsky (2006) used telecommunication data to provide evidence that those customers having communicated with a customer of a particular service have increased likelihood of adopting that service. Recently, Iyengar, Van den Bulte and Valente (2008) investigated the relationship between self-reported leadership and sociometric leadership, that is, when a person is nominated by others as someone they turn to for advice. They found that these two types of opinion leaderships are only weakly correlated and that there are significant differences between the adoption behavior of different types of opinion leaders. Studying a different type of network, Stephen and Toubia (2010) examined the role of the link structure in seller networks, where links facilitate customers' navigation between different stores. They found that the network had a positive overall effect on store performance and that the position of stores in the network had a significant effect on their profitability.

⁴Goldenberg, Libai and Muller (2001, 2002) generated adoption data by applying stochastic cellular automata models to simulate the diffusion process on a two-dimensional grid. In Goldenberg et al. (2002), they compared the results of their simulations with consumer electronics sales data to explain double-peaked diffusion curves, observed in many product categories. Shaikh, Rangaswamy and Balakrishnan (2005) developed a diffusion model taking into account local network characteristics. They demonstrated how it may be used to infer structural properties of the consumer network from aggregate sales data.

Our work directly follows this stream of research. Our primary goal is to understand how network characteristics (particular patterns of connections between already adopted network members) may influence the adoption probabilities of their not-yet-adopted peers. To establish this relationship, we estimate a discrete-time proportional hazards model. This is the same approach that Bell and Song (2007) follow to estimate network effects on purchase behavior across regions in the US. However, as in our data set we can observe the individual network connections of participants, we are able to use the same approach to study the impact of personal network structure on adoption.

How network structure affects social influence has been extensively studied in sociology, although only in small social networks.⁵ An influential paper by Krackhardt (1998) suggests that when assessing the influence individuals' contacts have over them, we should not only count the number of related actors but look at how those relationships are embedded in the entire network of relationships. Coleman (1988) and later Burt (2005) establish two important social phenomena, which may be tied to structural properties of the network. When two related individuals are connected to the same third parties, the network becomes better at transmitting information, and ultimately the affected relationships become stronger. Burt (2005) labels this as *network closure*. He argues that the shared third parties create redundant paths for information flow, leading to increased trust between the two related actors. As a consequence, friends of individuals in social networks are typically very densely connected to each other compared to the average connectivity in the network (Watts and Strogatz 1998). This is consistent with Granovetter (1973) and Rogers (2003), who stated that social networks in general consist of clusters of densely connected individuals with strong ties among them, and sparse weak ties connecting such clusters to each other. Burt (2005) highlights another phenomenon related to social networks' structural properties, namely the fact that individuals interconnecting these clusters may have higher influence on their peers because they have control over information originating from other groups. This phenomenon is termed *brokerage* and the argument is often cited as *structural hole theory* (Burt 2005). These arguments will provide critical input for our network measures.

Our model is partly based on the approaches proposed by Holtz (2004) and Shaikh,

⁵See also relevant early studies in marketing: Reingen and Kernan (1986) and Reingen, Foster, Brown et al. (1984).

Rangaswamy and Balakrishnan (2005). However, these papers do not estimate the adoption process at the individual level. The former paper attempts to disaggregate the Bass-model to the network level and uses simulations to confirm the results, whereas the latter compares the fit of a small-world-based diffusion model to the traditional models using aggregate-level data. In this paper, on the other hand, we develop and estimate an individual-level diffusion model, where each individual is described both by her local network characteristics and demographic information. At any point in time during the adoption process and for each potential adopter, we compute the network spanned by their already adopted peers. We then estimate how the structure of the network of their adopted peers affects the adoption probability of the potential adopters. Further, we also show empirically how network characteristics of already adopted individuals may predict their influential power on others. The estimates also provide insights on some of the common sociological theories in network analysis. Our model is presented next.

3 Model

We start by describing the individual level adoption model. This is followed by the description of network measures.

3.1 The Adoption Process

To examine the process of diffusion over a social network, we use a hazard rate model. We assume that all members of the network are potential adopters of a certain product or service. The relationships among network members serve as paths for WOM communication, through which network members may influence each others' adoption. We denote such a social network by $G(V, E)$, where V is the set of potential adopters and E is the set of symmetric binary relationships among members of V .

By *neighbors* of a potential adopter v , denoted $N(v)$, we refer to those network members (actors) who are connected to v . That is, $N(v) = \{w | w \in V \text{ and } \{v, w\} \in E\}$. The degree of a node, $d(v)$, is then the number of its neighbors: $d(v) = |N(v)|$. For notational convenience, we further

define an indicator function of the relationships in the network. For every $v_1, v_2 \in V$, we let

$$e(v_1, v_2) = \begin{cases} 1 & \text{if } (v_1, v_2) \in E \\ 0 & \text{otherwise.} \end{cases}$$

We assume that at the start of the diffusion process, a small set of actors, $A_0 \subseteq V$, have already adopted the innovation in question. We then model time in a discrete way. In every time step, some actors adopt the innovation, resulting in the series $A_0 \subseteq A_1 \subseteq \dots \subseteq A_T$, where A_t denotes the set of actors in the network who have adopted the innovation during the first t steps. For the individuals in the network, let T_v denote the adoption time of actor v , yielding

$$A_t = \{v \in V \mid T_v \leq t\}.$$

It is important to make the distinction between the overall social network and the subnetworks induced by A_t . The network $G(V, E)$ defined on the entire set of individuals V (the potential adopters) remains unchanged throughout the process. On the other hand, as the number of adopters increases over time, A_t covers a larger and larger subset of V . The network *induced* by A_t consists of the nodes of A_t links of the original network $G(V, E)$ leading between two already adopted members (of A_t). If we were to plot the network induced by A_t for every $t < T$, with the overall network in the background, we could see how the innovation diffused across the network over time.

In our model, the adoption likelihood of potential adopters depends on their individual propensity to adopt and on the influence of their neighbors who have already adopted. As explained earlier, conceptually, we distinguish between three types of influences. First, (local) *network effects* refer to the impact of the *structure* of connection patterns among the potential adopter's already adopted neighbors. Thus, in this case we only consider the networks that are induced by A_t for a given time t . Figure 1 illustrates this phenomenon: the network among the black nodes 1, 2 and 3 may influence the adoption likelihood of later adopter 8. It is important to emphasize that these network effects do not distinguish individuals in the personal networks of potential adopters, it is only the local network structure that drives such effects. Over time, this local network changes as more and more neighbors of the potential adopter adopt themselves. As such, network variables are time varying covariates.

[– Insert Figure 1 around here –]

Second, by *influencer effects* we refer to the impact of average individual characteristics of already adopted network members on their not yet adopted friends. These average characteristics may be demographics (e.g., age), but they may also include measures that describe the global network position of the influencers, i.e. their position in the final friendship network, G . While these measures do not change over time for a single influencer, as we measure the average effect of influencers, these variables also become time varying covariates as additional influencers adopt the innovation. Influencer effects may impact multiple potential adopters. Figure 2 illustrates this: the already adopted actor 8 may impact the adoption likelihood of later adopters 9, 10, 12 and 13, indicated by black nodes. We assume that such personal influences are the same on all of the influencer’s later adopting neighbors.

[– Insert Figure 2 around here –]

Finally, we also consider *adopter effects*, which refer to the impact of the adopters’ individual characteristics on their adoption propensity. These are constant variables over time. Again, some of these measures are demographic variables, but we also include the overall network characteristics of a certain adopter, for example, the total number of connections an individual has in the final network, G . (By considering the number of real-life friends an individual has, we control for differences in individuals’ propensity to make friends.)

To capture the above described effects, let f denote the influence function $f : 2^V \times V \rightarrow [0, 1]$, with the following interpretation. If an actor $v \in V$ has not yet adopted by time t , then the probability that it adopts at time $(t + 1)$ is $f(N(v) \cap A_t, v)$, i.e., a function of the set consisting of the neighbors of the actor who have already adopted and the actor itself. Formally,

$$\Pr[T_v = t + 1 | T_v > t] = f(N(v) \cap A_t, v). \quad (1)$$

The function $f(S, v)$ thus captures both the propensity of actor v to adopt the innovation in question and the strength of influence the set of adopted neighbors $S \subseteq V$ have on actor v . (Note that our model formulation also allows for individual characteristics of members of S to

influence the adoption likelihood of v .) In order to estimate the effect of certain structural properties of the network on adoption probabilities, we interpret the daily adoption decisions as binary choices of actors between adoption and non-adoption. There are a number of different link functions that can be used to model such binary decisions as dependent variables, including the most frequently used logit and probit. Herein, we employ the complementary log-log link function formulating the equation

$$f(S, v) \equiv 1 - \exp[-\exp(\alpha + \beta \times X(S) + \gamma \times W(S) + \phi \times Z(v))], \quad (2)$$

where $X(S)$ is a set of variables describing the local network structure, $W(S)$ is a set of variables describing the average characteristics of the members of S (the influencers), while $Z(v)$ is a set of node-specific (i.e. adopter-specific) covariates. The choice of the complementary log-log transformation has two important advantages over the logit and probit functions. First, the so-derived discrete time parameter estimates are also the estimates of an underlying continuous-time proportional hazards model (Prentice and Gloeckler 1978). Further, the complementary log-log link also allows us to directly relate our model to hazard rate models with utility-maximizing consumers (Bell and Song 2007).

We also note that the complementary log-log and logit formulations yield very similar results for small probabilities. However, besides allowing a direct interpretation of the results as hazard ratios, the prior method also provides a slightly better fit in such cases.

3.2 Measures

Below, we describe the set of variables that capture the three factors affecting the diffusion process.

3.2.1 Network effect measures

We start by discussing the effects that the local structure of already adopted friends has on the adoption likelihood of an individual. We keep the notation $S = N(v) \cap A_t$ for convenience.

Degree

The most natural question to ask regarding the influence of an actor's adopted neighbors is how their number affects the likelihood of the actor's adoption. Granovetter (1973), Valente (2005) suggest that one friend may have a greater impact on an actor's behavior when the actor has, in total, less friends. Hence, we choose our first local network variable to be the proportion of already adopted friends:⁶

$$X^1(S) = \frac{|S|}{|N(v)|}.$$

Intuitively, if a person has more friends already using a certain service or product, she will adopt with a higher probability. Hence, we expect degree to have a positive effect on adoption probability.

Clustering

Our second local network variable is the *clustering coefficient*, which measures the extent to which a set of members are interconnected. Clearly, this measure might be relevant in a context where these members exert an influence on another member. The definition of the clustering coefficient is:

$$X^2(S) = \rho(S) = \frac{\sum_{s,t \in S} e(s,t)}{\sum_{s,t \in S} 1},$$

where the numerator counts the number of links among the already adopted neighbors of v and the denominator is the maximum number of relationships possible among them. Network closure theory (Coleman (1988), Burt (2005)) proposes that if two actors related to the same individual are also related to each other, they have greater power over that individual than if they were unrelated. In our context, we could expect that if a potential adopter hears about the social network service from two friends then the attractiveness of the service is higher when these two friends also know each other. The density of relationships among adopted friends of potential adopters may thus impact their adoption likelihood. Based on this stream of research in sociology, we expect that clusteredness has a positive effect on adoption probability.

As discussed above, a higher clustering coefficient, on average, indicates stronger relationships. However, keeping the density of one's personal network constant, the number of

⁶In Section 4.4.2 we explore an alternative formulation where we use the number of adopted friends an individual has.

relationships in the network increase quadratically with the number of related actors. In other words, a larger personal network with the same clustering coefficient requires more ties per neighbor between its members. Therefore, we expect that larger personal networks with the same clustering coefficient indicate a stronger network closure. In other words, we expect a positive interaction effect between the degree and clustering variables. To this end, our third variable becomes the degree-clustering interaction, or

$$X^3(S) = X^1(S) \cdot X^2(S).$$

3.2.2 Influencer measures

Here, we focus on the variables that describe the characteristics of the individuals who may influence a potential adopter.

Influencer total degree

Opinion-leaders have always been in the focus of marketers. They are considered important targets for marketing communication. Nair, Manchanda and Bhatia (2010) empirically showed that in referral networks of physicians, opinion-leaders significantly altered the behavior of other individuals in the networks. Studying the aggregate impact of influencers on diffusion, Watts and Dodds (2007) ran a series of computer simulations. They showed how the structure of social influence may decrease the relative importance of highly connected individuals over a critical mass of easily influenced individuals. Goldenberg, Han, Lehmann et al. (2009) examined this issue in an empirical study, finding that members of a social network with large network degree (*hubs*) actually had a larger-than-average overall impact on adoption. In our model, we revisit this question, but we focus strictly on the micro-level effects of high network degree. We analyze how the average influence of adopted network members on their later adopting friends depends on the total number of friends these adopted actors have. That is, our first influencer variable is the average number of connections that potential adopters' friends have:

$$W^1(S) = \frac{\sum_{w \in S} d(w)}{|S|}.$$

Lin (1999) argues that in social networks, the larger the personal network of actors becomes, the easier these actors get access to more diverse social resources. Goldenberg et al. (2006) also find

support for this argument: surveying consumers' opinion-seeking habits, they identify conditions when social connectivity is more important for opinion leadership than product expertise. In the context of our analysis, however, social status alone is insignificant to influence the behavior of not-yet-adopted individuals: new adopters had to be actively *informed* about this novel type of service. Personal communication (at least one of the adopted friends sending an invitation, plus perhaps other, potentially offline conversations) had to precede every adoption and many of the instrumental actions mentioned in Lin (1999) were not supported by the medium. For this reason, we argue that the average intensity of friendships of an actor likely diminishes with the number of friends the actor has. This is also consistent with the findings of a recent study in marketing by Stephen et al. (2010) and with early sociological work by (French and Raven 1960) suggesting that people only have a limited amount of influential power. Put differently, whereas it is likely that highly connected individuals have a high influence on their closest ties, it is also likely that they could not spend much time communicating with all of their friends. In sum, we expect that the average influential power of actors (on average over all actors) decreases in the total number of friends the actors have.

Betweenness

The “betweenness centrality” of v is defined by Freeman (1977, 1979) so that for every pair s, t of the other nodes in the network, if v lies on the shortest path between s and t then that pair of nodes contributes to the betweenness centrality of v . The intuition behind such a definition is that if a message traveling from s to t has to pass through v then the structure of connections indeed increases the influence of v over t (and, in an undirected network, over s , with the roles of source and destination being interchangeable).

While such influence may very well be present in general (for instance: organizational) social networks, we argue that in our network of real-life friendships, the influence is irrelevant if s and t are not actually neighbors (friends) of v , i.e. we argue that influence depletes rapidly with additional intermediaries. Therefore, we propose a definition for *local betweenness* to focus on structural holes at v by examining only pairs of neighbors of v :

$$B(v) = \sum_{s \neq t \in V} \frac{e(s, v) \cdot e(v, t) \cdot (1 - e(s, t))}{\sum_{w \in V} e(s, w) \cdot e(w, t)}. \quad (3)$$

For every unrelated pair of actors s, t among the neighbors of v , the contribution of the pair s, t to the betweenness of v is inversely proportional to the number of length-2 paths between s and t . For simplicity, from here on, we refer to the local betweenness measure as *betweenness*.

[– Insert Figure 3 around here –]

To illustrate the concept, consider Figure 3. In the network on the left, actor 1 interconnects the pairs of not-related actors (2,4) and (3,4) but not (2,3) since 2 and 3 are related in that network. Since actor 1 is the middleman of the only 2-step path between actors (2,4) and (3,4), $B(v)$ in Equation (3) becomes 2. In the same figure, on the right-hand side, actor 1 interconnects the same two pairs of not-related actors. However, actors (2,4) are also connected through actor 5, while actors (3,4) are also connected through both actors 6 and 7. In this way, $B(v)$ becomes $1/2 + 1/3 = 5/6$.

To account for the effect of local betweenness on the adoption likelihood of v , we take the average betweenness of v 's neighbors at the time of the adoption decision. Thus, our second influencer variable becomes:

$$W^2(S) = \frac{\sum_{w \in S} B(w)}{|S|}. \quad (4)$$

How does this variable affect adoption probabilities? Besides network closure, Burt (2005) details another structural pattern of network connections that may alter the influential power of network members involved. When an actor is interconnecting two, otherwise not well communicating parts of a network, then we talk about a *structural hole* in the network. The interconnecting actor may be able to *broker* information between the two sides. Clearly, such brokers may have higher influence over related actors on both sides of the network. Thus, the literature on structural holes would suggest that structural holes increase the influential power of brokers. However, when the personal network of an actor has two parts that otherwise do not communicate, then the influence of the actor in the middle of the structural hole is governed by two opposing effects. It may increase with the sizes of the corresponding parts because of the higher social status that brokerage gives, but just like we point out above, as the personal network grows bigger, on average, these relationships become weaker, and the influence of the brokering actor on her potential adopter friends may decrease.

Control variables

Beyond the variables outlined above, we also use a number of control variables in our estimations. Due to the lack of theory, we do not have explicit predictions on the effect of these measures. However, from a practical perspective, including them in the estimation may be interesting as they can help predicting adoption probabilities and identifying influential consumers.

One of our control variables is the influencer’s clustering coefficient in the final friendship network. This variable - as discussed before - is a proxy for how dense the influencer’s network is. Furthermore, we also examine the effect of two demographic variables. In particular we look at how an influencer’s age and gender affects his influential power. As in the case of the network variables, we take an average over the neighborhood of the individual. For example, in the case of age the variable is:

$$W^3(S) = \frac{\sum_{w \in S} AGE_w}{|S|}.$$

For gender, we simply include the proportion of females within S .

3.2.3 Adopter measures

This group of variables measure how the characteristics of potential adopters affect their likelihood of adoption in each time period. Again, these variables are of two types. The first set describes the adopter’s final network characteristics (total degree, betweenness and clusteredness) as in the case of influencers. For the lack of theory, we do not have explicit predictions concerning these variables. Rather, they should be considered control variables. The second set of (control) variables are demographics. Beyond age and gender, here we can also include an additional variable: population density of the city of residence.

4 Empirical Analysis

4.1 Data Description

Our data originates from a major European social networking site. The goal of this web-based networking service is to build an online community of people who then may use the tools

provided by the web site to interact: to send messages to their friends, to share their pictures, to maintain a profile page with personal information, etc. People can represent their “friendship network” graphically and can search the membership base by various criteria. An important feature of the site is that proposed friendships need to be confirmed by the other party and can be severed as well. As a result, the network contains information on mutual relationships only.

Today, the portal we study has more than four million registered users and over 100 million friendship links between them. In this paper, we analyze adoption data from the first three and a half years (1247 days) of the service. The reasons to consider this early time frame are twofold. First, during this period, the service was not advertised and its media appearance was minimal. This means that membership growth was entirely due to WOM effects. Second, during the period studied, the service was the only of its kind in the country. At the end of the examined time period, the web site started experiencing technical difficulties providing the service and they limited the number of new members who could join. Soon after the alleviation of the technical problems, social networking got exposure in the national media, resulting in both the launch of competing portals and in a sudden growth of membership.

To register to the site, potential users had to receive an invitation from a member. During the time period we study, a member had unlimited invitations to send, hence the availability of invitations did not limit the growth of the network. During the period analyzed, 138,964 users registered to the website. To minimize the potential bias caused by including only users who adopted during the studied time frame, we also include some of the next adopters in the dataset. Due to computational limitations, we set the size of the analyzed sample at 250,000 users (of whom the latter 111,036 members signed up during the first month after the last day we look at).

When analyzing the diffusion process, we investigate the adoption of the portal (the social network site) over time by members of the *real-life* friendship network of potential adopters. We consider all the relationships confirmed in the system 36 weeks after the end of the studied period as this real-life friendship network. By doing so, we implicitly assume two things: (i) that the network recorded on the site 36 weeks after our study period contains all the “real-life” relationships between potential adopters, and (ii) the relationships indicated in the network were not “formed on the network”, i.e., not even partly resulted from individuals’ adoption behavior.

Obviously, the validity of these assumptions may be critical to our findings. In Section 4.4, we conduct several tests to conclude that neither of them weaken our empirical results.

The friendship data (sampled as described above) contains 13,152,323 links among the 250,000 registered users, corresponding to a friendship density of $4.21 \cdot 10^{-4}$. Table 1 summarizes the distributions of the demographic variables and network characteristics defined in Section 3.

[– Insert Table 1 around here –]

In Table 1, we also report the correlations of demographics and network characteristics between related actors in the network. We can see that, whereas the network characteristics of neighbors are indeed positively correlated, these weak correlations are unlikely to be the drivers of the results of our estimations. Finally, Table 2 reports correlations of the independent variables used in our estimations. These, together with the network autocorrelation information in Table 1 suggest that Age, Gender and Population Density are good control variables. We note that the network autocorrelation is higher for the age and the population density variables, suggesting to consider group effects within age clusters and within cities. We deal with these estimation issues in Section 4.4.1.

[– Insert Table 2 around here –]

In the outline of the stochastic model, we defined a distinguished group of people who are already using the service at the start of the analyzed period. In our empirical analysis, we include the first 2121 members in this initial group because these members received their membership directly from the creators of the social network. Finally, as the database contains registration dates for every member, we choose days to be the unit of time (thus estimating daily adoption probabilities in our model). Next, we present the details of the estimation.

4.2 Estimation

The estimation is based on the network-based influence model described in (1) and (2). We apply the complementary log-log regression using the ML method to obtain parameter estimates. The

observations we use are based on daily adoption decisions. Every day we record the decision of each individual who has not adopted yet along with the variables described in the previous section. Some of these variables are constant over time, some of them change as the adoption process moves forward. Since we have 250,000 individuals in the dataset that we analyze throughout a period of 1247 days, we obtain close to two hundred million such observations. Just storing the values of our variables for all of these observations would exceed our computational capacities. To overcome this problem, we discretize some of our variables so that we can store the observations in a grouped table. For the variables we discretize, we split the range of the variable into 100 equal intervals. For each interval, we only store the average value of the variable over observations that lie in the interval. In this way, we are able to reduce the space required to store our observations (which can be represented by integer coordinates of points in a multi-dimensional grid) and run the maximum likelihood estimation. To examine in detail how the number of intervals chosen affects the results, we build the multi-dimensional grid several times, each time with randomly selected dimensions containing only 50 equal intervals (keeping 99 ticks at the others). We find that the 95% confidence intervals overlap for the vast majority of the coefficients, confirming the validity of this approach. Due to spatial limitations, we omit the exact results from the paper.

Since our primary focus is on investigating how observed network characteristics can predict adoption probabilities, we first estimate a model (Model 1) with *network effects* only. Beyond these effects, Model 2 also includes *influencer effects* and *adopter effects* that are still network related, i.e. have been calculated from the ultimate friendship network of the population. In Model 3 (the full model), in addition to the above, we also include all the demographic variables.

4.3 Results

Table 3 summarizes the results of the complementary log-log regressions. Across Models 1,2, and 3, we find consistent support for the positive effect of degree and clustering. Specifically, following the general intuition expressed in Section 3.2, we observe a clear *degree effect*: more adopted neighbors increase the likelihood of a potential adopter to adopt, and while holding the

number of adopted friends constant, a higher total number of friends decreases adoption likelihood. More interesting is the strong support for the *clustering effect*. Consistently with network closure theory, a set of highly connected individuals have a stronger influence on a potential adopter than an identical number of sparsely connected ones. It is interesting to see how strong the clustering effect is. When an individual has 100 friends and 6 of them have already adopted, then one extra adopted friend has the same effect on his/her adoption probability as one extra friendship between the 6 adopted friends. Finally, the *interaction* of degree and clustering is also significant and positive in the richer models providing partial support for the general intuition that the marginal effect of an additional influencer is larger for highly connected networks.

[– Insert Table 3 around here –]

For influencer variables, we do not find significant effects for *betweenness*, indicating that structural hole theory may not apply to very large social networks in a straightforward fashion. An interesting further finding in this flavor is that *total degree* has a relatively weak but significant negative effect suggesting that individuals with many connections have less influential power on a particular neighbor.

The remaining influencer variables are all significant. The significance of the demographic variables shows that, in a practical setting, these are useful predictors for influential power. For this social network site, we find that younger people and female network members have a higher influence. This result is somewhat surprising since many researchers in sociology have acknowledged that men have a greater social power (Dépret, Fiske and Taylor 1993). However, influencing adoption likelihood to a social network portal may be substantially different from the general notion of social power. To further examine the effects of age and gender we estimate Model 3 separately for different genders and age groups. Besides confirming the direction and strength of the effects listed herein, we generally find higher influencer effects for similar individuals and that females have a significantly higher influence among younger individuals. We do not report the results of these estimations here, but in Section 4.4.1, we test if such unobserved heterogeneity in the relationships affects the validity of our findings.

Regarding adopter effects, we also find significant results. These are generally consistent across models except for adopter betweenness, which changes sign between models 2 and 3. The total degree of an individual, for example, has a positive effect on his or her adoption probability. This is not surprising considering that we also included the proportion of already adopted friends in the equation, through which the total number of friends relate to the dependent variable negatively. However, an alternative cause to this may be endogeneity whereby the “more enthusiastic” network users gather more friends online. This could invalidate our assumption on the exogeneity of the “final” network. To verify that this is not the case, we conduct several tests, which we present in Section 4.4.3.

With respect to demographics, we find that age and population density of the town of residence both have small but significant effects while gender does not. It is interesting that age has a positive effect, which generally contradicts empirical findings in the diffusion literature, since younger people have been shown to be generally more likely to adopt early. However, we already control for network related variables that may cause younger users to sign up sooner if the network density is higher among them. In addition to age, we find that population density of adopters’ home city has a marginal positive effect. Note that we find this latter effect in addition to the already discussed network effects, which already capture the potentially higher density of social connections in highly populated areas.

We also conclude that adopters’ propensity to join the network increased with the overall network size. This may indicate the presence of some offline buzz about the site. Yet, since network size and time are almost perfectly correlated, it is also entirely possible that this positive coefficient indicates that later network members got more accustomed to Internet technologies (in particular, the concept of a web-based online social network) over time, hence adopting with larger probability.

Comparing the fit of models, the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and McFadden’s pseudo R^2 measure all rank the models in the same order. The AIC/BIC measures require the same number of observations across estimations - therefore, we compute them over the 136,829,184 full observations that are included in Model 3. The actual values of these metrics are reported in Table 4. However, as we believe that

considering broader samples for the simpler models is more useful for interpreting the results, in most tables, including Table 3, we only report the pseudo R^2 values. We can see that although Model 3 gives the best fit based on all of the three measures, the corresponding pseudo R^2 value is still only 0.0496. The explanation for such low values is that, as the mean of our dependent variable is less than 0.001, the null model makes many good predictions (by always predicting non-adoption). Thus, low fit measures do not necessarily weaken our findings. However, as the independent variables are weakly positively correlated for network neighbors (see Column 4 in Table 1), it is possible that even characteristics of the real-life friendship network are driven by unobserved similarities among network neighbors (Manski 1993). The fact that the coefficients of Model 2 and Model 3 are similar while the explanatory power of Model 3 is higher indicates that this is not the case, providing support for the results. As we are not interested in the formation of the final friendship network per se, but rather, in the diffusion on an existing network of friendships, we leave all further discussion of this issue to Section 4.4.

[– Insert Table 4 around here –]

4.3.1 Predictive Power

The results of the previous section demonstrate how local network characteristics may influence adoption likelihoods. Here, we demonstrate how our models can help in predicting which individuals will adopt in the near future.⁷ Existing models such as the Bass-model (1969) perform well in forecasting the overall number of adopters, but do not provide any prediction on which individuals are more or less likely to adopt. Our methodology allows us to predict the adoption probabilities of individuals who have not adopted up to a certain point in time. However, it is not straightforward to measure the predictive power of our model if the outputs are probabilities, especially if these probabilities are as low as our daily adoption rates (less than 0.001). In real applications, it is often required to predict who the next adopters will be and not only provide probability predictions. In order to test the predictive power of our models, we therefore rank order the individuals in decreasing adoption likelihood, and compare the top m individuals of this list with the set of the next m individuals who adopted in reality. We do this for different values of

⁷We would like to thank the AE for suggesting this.

m - to determine the number m to be used in the context of a real application. One could, for example, use an aggregate-level model to predict the number of individuals adopting in the next period under consideration.

As our adoption data gets distorted after the studied period for previously mentioned reasons, here we make predictions for the last 10 days⁸ before the cut-off point, day 1247. Therefore, we first estimate our models on the first 1237 days of the adoption data. Then, using the resulting coefficients and the total network data, we calculate the predicted probability of adoption for each individual on day 1238. Then we rank these predicted probabilities and list the nodes to which the m highest probabilities belong as a prediction for the next m adopters. Let M denote the set of these individuals with $|M| = m$.

We test the predictive power of four models. The first three are the same three main models we estimated in Section 4.2. Model 4 is a benchmark where we ignore the network variables, and only include demographics and network size. We drop all observations where any of the variables used in any of the models are missing. To test predictive power, we calculate what percentage of individuals in M really adopted in the 10-day period. Table 5 shows the results for different values of m .

[–Insert Table 5 around here –]

The highest value we use, $m = 9944$, is the actual number of adopters during this period. This constitutes 11.66% of the potential adopters in our sample. Thus, a random set of size 9944 would have a hit rate of 11.66%. We use this as a benchmark, in addition to the prediction based on only the available demographic variables (Model 4), which successfully predicts 13.25% of the adopters. Using any of our three main models, we can almost double the hit rate from the random benchmark to around 21%, substantially improving the predictive power over that of Model 4. Displaying the success rates for Models 3 and 4 as a function of m , Figure 4 shows this pattern. As Table 5 shows, Models 1 and 2 yield results very similar to those of Model 3. If the application only requires the identification of a lower number of adopters than the total expected

⁸The prediction process does not depend on the number of days and the predicted sets would be the same for different periods. However, the actual probabilities would be very small for one day and it is more realistic to consider a longer period.

number of adopters, then the hit rate becomes even higher. For values around $m = 500$, Models 1, 2, and 3 predict adoption successfully with around a 30% probability. That is, for applications where not only successful predictions are of interest, but also the cost of “false positive” predictions is somewhat larger, it may be best to focus on only about 5-10% of future adopters. In such scenarios, our model performs much better than the random selection (11.66% success rate) and the model based solely on demographic variables (13.60% success).

[–Insert Figure 4 around here –]

4.4 Robustness and Validity Tests

In this section, we attempt to address several limitations of our model, data and estimation procedure. We start by allowing for latent individual heterogeneity in the propensity to adopt. Due to computational limitations discussed in Section 4.2, we are only able to estimate such models on a smaller (random) sample of individuals. In a different approach, we implode the friendship network into network layers containing only relationships within the same city of residence to examine whether our results also hold within the so-arising networks to further support our substantive findings. We continue by considering alternative model formulations. We change the link function, and we test an alternative definition of our main independent variable. Third, we discuss a number of issues related to network dynamics. By eliminating certain individuals, links and decisions from our dataset, we challenge our assumption on the exogeneity of the “real-life” friendship network, and explore the potential effects of heterogeneous link strength on our results.

4.4.1 Individual and group effects

In this section, we carry out tests to verify the validity of the assumption on the independence of the error terms in our main models. First, we estimate our model including random effects,

$$f(S, v) = 1 - \exp[-\exp(\alpha + \beta \times X_v(S) + \gamma \times W(S) + \phi \times Z(v) + u_v)], \quad (5)$$

where we assume that $u_v \sim N(0, \sigma^2)$. As the random effects vary by individual, we have to break with our estimation method of discretizing the independent variables, which assumes

homogeneity within each cell in the grid (that may contain observations from different individuals). Thus, to run the random effects model on the whole adoption dataset, we would have to deal with well over a hundred million observations. As this is clearly infeasible, we randomly select 5,000 individuals from the 250,000 network members and run the random effects estimation on the data comprised of observations concerning only these individuals. (However, we compute the independent variables from the network spanning over all the individuals in the data.) The results of this process are presented in Table 6. The large value of ρ confirms that the low model fit measures reported in Table 3 may indicate unobserved heterogeneity in the relationships. Nevertheless, whereas the coefficients cannot be directly compared to those presented in Table 3, we find that most of the effects identified by our main estimation are supported by the results herein as well. The fact that some of the demographic variables become insignificant suggests that demographics may be correlated to the unobserved propensity of adoption which is here captured by individual random coefficients. Finally, we note that our choice of restricting the structure of individual effects to the normal distribution is due to computational limitations: the data sample that we consider contains several million observations per regression, which makes estimating 5,000 fixed effect dummies infeasible.

[– Insert Table 6 around here –]

Next, we return to our original model formulation (without individual effects), but we allow the error terms to be correlated for homogeneous age or gender groups, arriving to more robust standard error estimates. We generally find that the network effects we look at, remain significant. A specific example leading to the biggest drop in the strength of the effects that we obtain this way, is presented in Table 7. For this estimation, we divide the population into five age clusters, grouping ages 21 or less, 22-23, 24-26, 27-29, 30 or more together. The parameter significances corresponding to the so obtained robust standard error estimates provide further support for the degree and clustering effects, and also for the negative effect of influencer total degree. We note, however, that the degree-clustering interaction and betweenness do not have significant effects under these circumstances.

[– Insert Table 7 around here –]

Finally, to explore the validity of some of our substantive findings, we refine the personal networks by dedicating special attention to friendships within the same city of residence. We split all network variables except clustering. We jointly estimate the impact of the number of friends from the same city and the impact of the number of friends from other towns. Similarly, the average betweenness of network neighbors from the same city and that of those from other towns become two independent variables. (In the case of the clustering variable, we do not have a straightforward way to do the same because across-town friendships in one's personal network may be present. Hence, clustering may not arise as a sum of within-city and across-cities components.) For the degree-clustering interaction, we have two choices: in Model 5, we include the degree variable as before (referring to the the proportion of adopted friends), and in Model 6, we include the within-city degree (proportion of adopted friends considering only within-city friendships).

[– Insert Table 8 around here –]

Table 8 shows the coefficient estimates of the so-formulated Models 5 and 6. We find support for the degree and clustering effects as well as the negative effect of influencer total degree. The degree-clustering interaction is only supported in Model 6, while we cannot detect the positive effect of betweenness (that structural hole theory would predict). An interesting phenomenon is the effect of network size: despite all the within-city network variables being more significant than their across-cities pairs, the opposite is true for network size, and the coefficient of within-city network size is negative. This suggests that the positive coefficient of network size in Models 1-3 rather indicates the evolution of users' affinity to technology than the presence of offline buzz.

4.4.2 Alternative model formulations

In this section, we discuss the most straightforward alternatives to some of our modeling choices. We begin by including two quadratic terms into our models to see if the negative effect of Influencer Total Degree is driven by actors with very large personal networks. As the results (presented in columns 5 and 6 of Table 3) indicate, the square of influencer total degree is not or

only marginally significant and it has a positive coefficient. This suggests that our main results are not driven by very highly connected individuals and that they hold for actors with smaller personal networks as well.

Next, we change our main independent variable from the *proportion* of already adopted friends within all friends of adopters, to their number. The estimates of the so-defined model are listed in Table 9. Whereas we find that all major results are consistent with those presented in Table 3, we also note that the model fit values are lower for this alternative specification. In sum, the proportion and not the sheer number of already adopted friends explains more about the adoption decisions studied.⁹

[– Insert Table 9 around here –]

Finally, we replace the complementary log-log link function (2) with the logit function and re-estimate our original models. As all coefficients and z -values are very similar to the results of the complementary log-log regression, we do not report the results herein.

4.4.3 Network Dynamics

Throughout our work, we have relied on the assumption that the observed network at the time of the study is “final” and static in the sense that we can observe all the real-life friendships among individuals in the network. This is clearly inaccurate for two related reasons. On the one hand, numerous friendships appearing in the data could have been made during the 3 and 1/2 years we analyze. On the other hand, there were probably some existing real-life friendships not confirmed in the network before the friendship data was collected. If the observed final network data is not a good proxy for the structure existing in real-life friendships, our results may be biased. In that case, one could argue that those individuals who had spent less time as users (i.e. people joining the network later) could not have enough time to map out all their friends present in the network. Then, the independent variables for these people (i.e. their network

⁹An interesting difference between the estimates of the two models is the coefficient of Adopter Total Degree. Goldenberg et al. (2009) proposed that social hubs tend to adopt earlier because of more exposure to the product in question. In Table 9, the number of (possible) exposures is directly captured in the Network Degree variable. As the coefficient of Adopter Total Degree is negative in this estimation, we indeed find empirical support for the above explanation to hubs’ early adoption.

characteristics) might be erroneous. Similarly, being a member of the service in order to make more friends could be a typical user behavior. In such a case, members with many friends would have a higher propensity to make new friends. In this case, the rate of the network's growth could vary based on how many friends a user already has, leading to sampling bias.

To verify that these concerns do not lead to qualitatively different results, we conduct a series of tests. We first look at whether the network spanned by the analyzed set of adopters grows uniformly. As we do not have access to the evolution of the complete network, we selected a random sample of users in our data and recorded the number of their friends among the analyzed pool of network users both at the time when the observed snapshot of the friendship network was taken and at a much later date (1002 days after the end of the studied period). We then analyzed the overall growth rate of network density, to find that the network of the observed 138,964 individuals continued to grow after the completion of the study at an average yearly growth rate of about 10%. Here, we test whether factors such as the number of friends already present or the length of time having been a member affect this rate. The results of this estimation are presented in Table 10. They show that neither the size of individuals' personal networks nor the date when they joined had a significant effect on the growth rate of their networks. In other words, the friendship network grows "uniformly", which in turn provides support against the presence of a sampling bias in the data we analyzed.

[– Insert Table 10 around here]

Whereas the above test eliminates concerns linked to particular types of user behavior, it also reveals that the "real-life" friendship network does slowly evolve over time. Below, we examine the potential impact of this trend by estimating three variants of our model. We begin with changing the sample of adopters to include only the 138,964 network members who adopt by the end of the analyzed period. The results of this estimation are presented in Table 11. The sign and significance of the coefficients of the degree and clustering variables do not change, while the effect of betweenness remains insignificant. Moreover, as the number of observations drops to less than half, we find that the model fit increases across Models 1-3.

[– Insert Table 11 around here –]

Next, we condense the friendship network, focusing on presumably stronger relationships. In two tests, we keep any relationship if and only if the endpoints of the corresponding edge in the network adopted the service within 30 and 60 days, respectively. We then recompute the network variables on the so-arising friendship networks, and re-estimate our models. The results of this procedure are presented in Table 12. Clearly, by tailoring the network the above way, the personal networks of users shrink. Such a condensation of the friendship network, therefore, works against our argument pointing toward a positive degree-clustering interaction, a negative effect of the influencer total degree, and it also may affect the chance of detecting the impact of structural holes (in the network defined by the stronger ties) on adoption. In Table 12 we show that after this transformation, we indeed do not find support for a positive degree-clustering interaction, and that the other two above effects also vanish when we apply the 30-day cutoff threshold. Interestingly, however, for the 60-days threshold we still find evidence for both the impact of structural holes and for that a larger total degree leads to a smaller average influence. In sum, we find that the negative effect of influencer total degree is robust, nevertheless it should be treated with care when one is only looking at the network spanned by the strongest ties between members.

[– Insert Table 12 around here –]

Finally, in a similar vein to the above test, we re-estimate our models considering up to 30 and 60 adoptions per user, respectively.¹⁰ This results in censoring observations for those individuals who adopt the service much later than their first adopting friend. In this estimation, we keep the network variables from the original model. The results are presented in Table 13. We find that our results concerning the degree and clustering effects and the negative effect of influencer total degree are confirmed by this estimation. For the positive effect of local betweenness, we only find partial support. An important finding is that the significance of the degree-clustering interaction in the richer models drops. This indicates that the interaction effect is (at least) in part driven by “slow adopters” who only sign up for the service once large dense subnetworks of their friends have already joined.

[– Insert Table 13 around here –]

¹⁰We thank an anonymous reviewer for suggesting this test.

5 Concluding remarks

With the emergence of new communication technologies, the communication patterns of individuals are increasingly recorded, providing an opportunity to study word-of-mouth at the individual level. Network data allows us to discover how each individual's social connections influence her adoption decisions. Furthermore, it also allows for measuring each person's influential power on others. The insights gained from such analyses are invaluable for marketers who search for new viral marketing tactics to foster new product diffusion.

In this paper, we have analyzed data from a social networking site to identify word-of-mouth effects at the individual level. Our central goal was to discover how the local communication network structure affects the diffusion process. In this respect, we confirmed the general intuition that the number of already adopted friends of an individual has a positive effect on the probability of her adoption. More interestingly, we have also found that the interconnectedness of already adopted friends of an individual also has a positive effect on her adoption probability. This finding is important as it shows that beyond sheer network size, strong communities are more relevant for word-of-mouth influence.

In a second step, we have also measured the average influential power of already adopted friends. At this end, the interesting finding is that people with many friends have a lower average influence than those with fewer friends. Similarly, influencers who occupy structural holes in the network have, on average, higher influential power. Besides these global network variables, we also found that demographics are useful to identify strong influencers. Finally, demographics and global network variables are not only useful to identify influencers but adopters as well.

These findings have important implications for firms that want to better understand how customers adopt their product or service. If data is available on the interpersonal connection patterns among their (potential) customers, our model may serve as a methodology to identify customers who play an important role during the diffusion of the new product or service. For example, at the initial phase of the adoption process, marketers may be interested in targeting potential customers with a high probability of adoption. Our research shows that they can do so by using demographics as well as global network variables. Similarly, the promotion effort can

also be targeted to influencers, who - like potential adopters - may be identified by their demographics and global network positions. At a later stage of the diffusion process, marketers may have information about the connections of already adopted customers. The information on this sub-network can be used again both to target potential adopters (e.g. people surrounded by a large *and* dense network of adopted friends) as well as influencers (members of such sub-networks). Our findings show that when the goal is to make prediction on the next set of adopters, our model can provide predictions that are successful in up to 30% of the cases, compared to around 13% when using only demographic information.

Clearly, the use of similar statistical models to improve viral marketing effort need to be adapted to the actual application context. Nevertheless, our findings confirm the general insight that a viral marketing campaign may be more effective if it is initiated on a sample of highly connected customers rather than on a sample of customers with simply the largest number of connections. Specifically, our results indicate that, in the case of high-involvement products where high *average* influential power is required to convince others to adopt, the firm may be better off focusing on individuals who have less connections. On the other hand, in the case of low-involvement products, highly connected customers may play the most important role because they have the highest overall influence.

Our study has a number of limitations. We have addressed some of these in Section 4.4, including the potential endogeneity of connections. Our validity tests indicate that this does not influence the results significantly and that the “final network” we have is a good proxy for people’s real-life networks. Yet, we only consider individuals who eventually signed up for the service. Self-selection thus might bias our estimated adoption probabilities, but not the direction of the effects we find. Another limitation of our study is that we do not possess data on the actual communication between individuals, we are only aware of links based on revealed friendships. Although this is a very simplified mapping of social connections, we know that these links are mutual, since both parties had to agree when establishing such a connection. Further, because of the computational issues arising from the size of our data set, we had to keep relationships within the network homogeneous (the related actors and their personal networks being the only sources of heterogeneity). The results of Manchanda, Xie and Youn (2008) suggest that this may be an

important limitation. Finally, we only have access to limited demographic information, such as age, gender, and city of residence. It would be interesting to examine how other variables that are directly influencing social connections, such as the schools an individual went to, affect the diffusion process. These limitations undoubtedly weaken our study. However, despite the limitations, the results clearly show the potential impact of network analysis for efficient word-of-mouth marketing. Furthermore, the models presented provide a first step towards developing statistical tools to analyze the impact of network structure on new product diffusion.

References

- Bass, Frank M. (1969), "A new product growth model for consumer durables", *Management Science*, 15, 215–227.
- Bell, David R. and Sangyoung Song (2007), "Neighborhood effects and trial on the Internet: Evidence from online grocery retailing", *Quantitative Marketing and Economics*, 5, 361–400.
- Burt, Ron S. (2005), *Brokerage and Closure*, chap. Chapter 3: Closure, trust and reputation. Oxford University Press, 93–166.
- Coleman, James S. (1988), "Social Capital in the Creation of Human Capital", *American Journal of Sociology*, 94, 95–120.
- Dépret, Eric, Susan T. Fiske and Shelley E. Taylor (1993), *Control Motivation and Social Cognition*, G. Weary, F. Gleicher, and K Marsh (eds.), chap. Social cognition and power: Some cognitive consequences of social structure as a source of control deprivation. New York: Springer-Verlag, 176–202.
- Freeman, L. C. (1977), "A set of measures of centrality based on betweenness", *Sociometry*, 40, 35–41.
- (1979), "Centrality in social networks: Conceptual clarification", *Social Networks*, 1, 215–239.
- French, John R. P. Jr. and Bertram H. Raven (1960), *Group dynamics*, D. Cartwright and A. Zander (eds.), chap. The bases of social power. New York: Harper and Row, 607–623.
- Godes, David and Dina Mayzlin (2004), "Using Online Conversation to Study Word-of-Mouth Communication", *Marketing Science*, 23 (4), 545–560.
- Goldenberg, Jacob, Sangman Han, Donald R. Lehmann et al. (2009), "The Role of Hubs in the Adoption Process", *Journal of Marketing*, 73 (2), 1–13.
- Goldenberg, Jacob, Donald R. Lehmann, Daniella Shidlovski et al. (2006). "The role of expert versus social opinion leaders in new product adoption". Tech. Rep. Report No. 06-124, Marketing Science Institute, Cambridge, MA.
- Goldenberg, Jacob, Barak Libai and Eitan Muller (2001), "Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth", *Marketing Letters*, 12 (3), 211–223.
- (2002), "Riding the Saddle: How Cross-Market Communications Can Create a Major Slump in Sales", *Journal of Marketing*, 66, 1–16.
- Granovetter, Mark (1973), "The Strength of Weak Ties", *American Journal of Sociology*, 78 (6), 1360–1380.

- Hill, Shawndra, Foster Provost and Chris Volinsky (2006), “Network-based Marketing: Identifying Likely Adopters via Consumer Networks”, *Statistical Science*, 21 (2), 256–276.
- Holtz, Georg (2004), “An individual level diffusion model, carefully derived from the Bass-Model.”, working Paper, University of Osnabrueck.
- Iyengar, Raghuram, Christophe Van den Bulte and Thomas W. Valente (2008), “Opinion Leadership and Social Contagion in New Product Diffusion”, working Paper, Wharton.
- Krackhardt, David (1998), *Networks and organizations: Structure, form, and action*, R. Kramer and M. Neale, chap. Simmelian Ties: Super Strong and Sticky. Sage, Thousand Oaks, CA, 21–38.
- Lin, Nan (1999), “Social Networks and Status Attainment”, *Annual Review of Sociology*, 25, 467–487.
- Mahajan, Vijay, Eitan Muller and Frank Bass (1990), “New Product Diffusion Models - a Review and Directions for Research”, *Journal of Marketing*, 54 (1), 1–26.
- (1993), *Handbooks in Operations Research and Management Science*, Jehoshua Eliashberg and Gary L. Lilien Eds., vol. Marketing, chap. New Product Diffusion Models. Amsterdam, The Netherlands: Elsevier Science Publishers, 349–408.
- Manchanda, Puneet, Ying Xie and Nara Youn (2008), “The Role of Targeted Communication and Contagion in New Product Adoption”, *Marketing Science*, 27 (6), 950–961.
- Manski, Charles F. (1993), “The Reflection Problem”, *Review of Economic Studies*, 60, 531–542.
- Nair, Harikesh, Puneet Manchanda and Tulikaa Bhatia (2010), “Asymmetric Peer Effects in Physician Prescription Behavior: The Role of Opinion Leaders”, *Journal of Marketing Research*, forthcoming.
- Oren, Shmuel S. and Rick G. Schwartz (1988), “Diffusion of New Products in Risk-Sensitive Markets”, *Journal of Forecasting*, 7, 273–287.
- Prentice, R. L. and L. A. Gloeckler (1978), “Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data”, *Biometrics*, 57–67.
- Reingen, Peter H., Brian L. Foster, Jacqueline Johnson Brown et al. (1984), “Brand Congruence in Interpersonal Relations: A Social Network Analysis”, *Journal of Consumer Research*, 11, 771–783.
- Reingen, Peter H. and Jerome B. Kernan (1986), “Analysis of Referral Networks in Marketing: Methods and Illustration”, *Journal of Marketing Research*, 23, 370–378.
- Rogers, Everett M. (2003), *Diffusion of Innovations*. 5th edn. New York: The Free Press.

- Shaikh, Nazrul I., Arvind Rangaswamy and Anant Balakrishnan (2005), "Modeling the Diffusion of Innovations Through Small-World Networks", working Paper, Penn State University, PA.
- Stephen, Andrew T., Yaniv Dover and Jacob Goldenberg (2010), "You Snooze You Lose: Comparing the Roles of High Activity and Connectivity in Information Dissemination Over Online Social Networks", working paper, INSEAD, France.
- Stephen, Andrew T. and Olivier Toubia (2010), "Deriving Value from Social Commerce Networks", *Journal of Marketing Research*, 47 (2), 215–228.
- Urban, Glen L. and John R. Hauser (1993), *Design and Marketing of New Products*. Second edn. Englewood Cliffs: Prentice Hall.
- Valente, Thomas W. (2005). "Models and methods for innovation diffusion". In P. J. Carrington, J. Scott and S. Wasserman, eds., "Models and Methods in Social Network Analysis," Cambridge, UK: Cambridge University Press.
- Van den Bulte, Cristophe and Yogesh V. Joshi (2007), "New Product Diffusion with Influentials and Imitators", *Marketing Science*, 26 (3), 400–421.
- Van den Bulte, Cristophe and Gary L. Lilien (2001), "Medical Innovation Revisited: Social Contagion Versus Marketing Effort", *American Journal of Sociology*, 106 (5), 1409–1435.
- Watts, Duncan J. and Peter S. Dodds (2007), "Influentials, Networks, and Public Opinion Formation", *Journal of Consumer Research*, 34 (4), 441–458.
- Watts, Duncan J. and Steven H. Strogatz (1998), "Collective dynamics of 'small-worlds' networks", *Nature*, 393, 440–442.

Tables and Figures

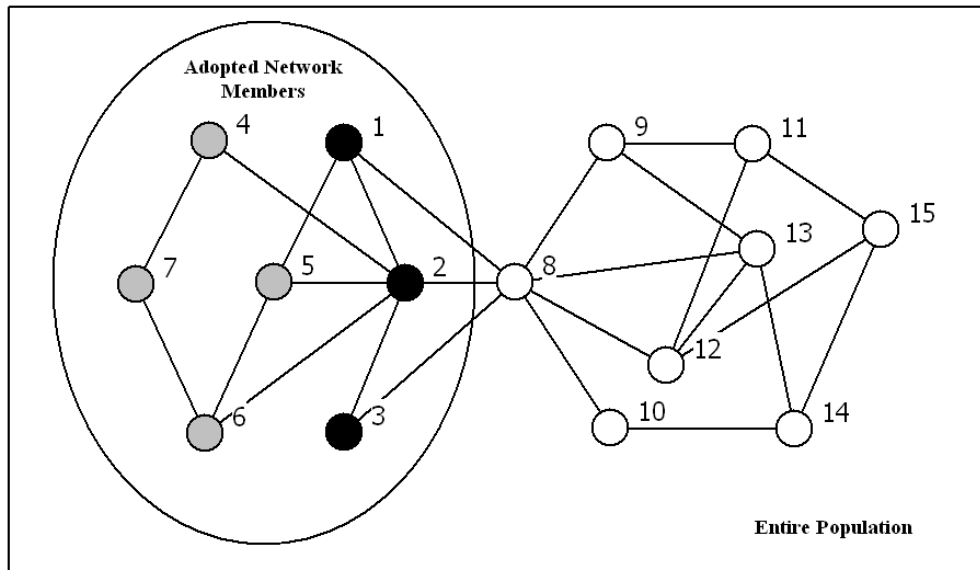


Figure 1: Illustrating “Network Effects”: the set of adopted neighbors $S = \{1, 2, 3\}$ influence the adoption likelihood of later adopter $v = 8$.

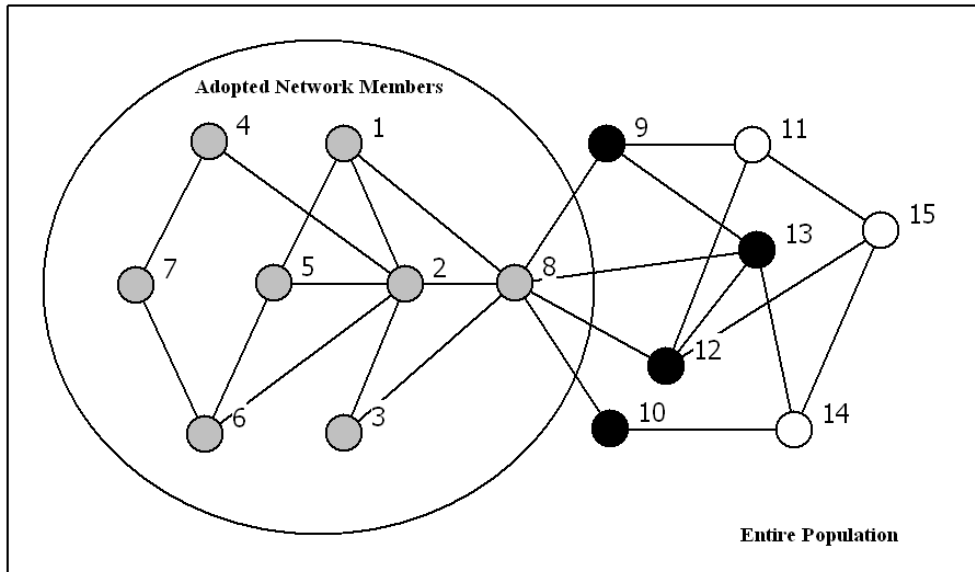


Figure 2: Illustrating “Influencer Effects”: characteristics of actor $v = 8$ impact the adoption likelihood of all later adopters $w_1 = 9, w_2 = 10, w_3 = 12, w_4 = 13$.

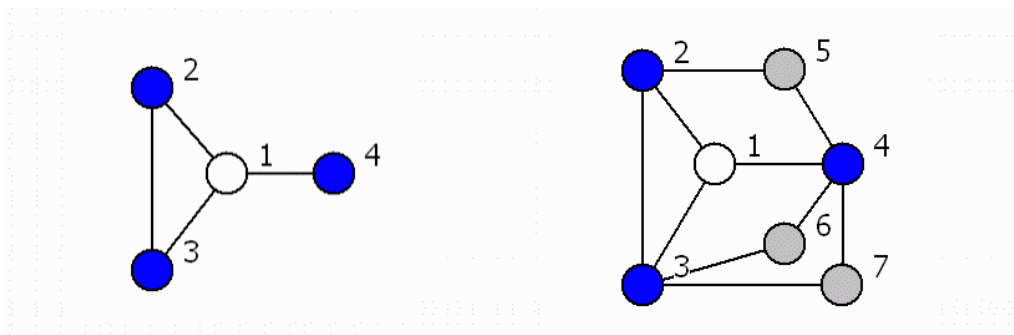


Figure 3: Sample networks to illustrate local betweenness: in the left network, the local betweenness of actor 1 is 2 while in the right network, actor 1’s local betweenness is 5/6.

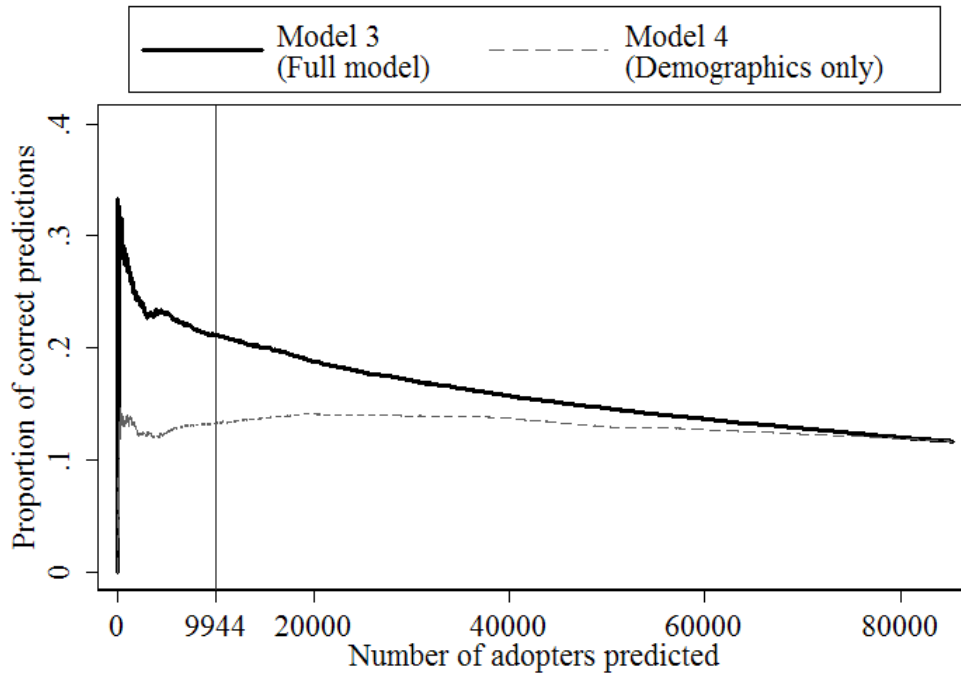


Figure 4: Proportion of adopters among predicted out-of-calibration sample adopters as a function of the size of the prediction list.

Table 1: Summary Statistics for the Studied Population of Network Members

Demographics				
Variable	Mean	St. Dev.	Netw. Autocorr.	Records
Gender (Female=1, Male=0)	0.52	0.50	0.07	250000
Age	26.45	7.89	0.38	203030
Population Density (in <i>pop / ha</i> , of city of residence)	23.59	12.93	0.37	204861
Network Characteristics				
Variable	Mean	St. Dev.	Netw. Autocorr.	Records
Degree	206.41	152.25	0.11	250000
Clustering Coefficient	0.17	0.13	0.22	235584
Local Betweenness ($\times 0.001$)	36.63	93.78	0.01	250000

Table 2: Correlation Coefficients of Adoption Determinants

Variable	Adoption Determinant Variables														
	X_1	X_2	X_3	W_1	W_2	W_3	W_4	W_5	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	
Degree (X_1)	-0.03														
Clustering (X_2)	0.67	0.40													
Degree \times Clustering (X_3)	-0.08	0.10	-0.01												
Influencer Total Degree (W_1)	-0.08	-0.00	-0.04	0.73											
Influencer Betweenness (W_2)	0.02	0.16	0.15	-0.19	-0.10										
Influencer Clustering (W_3)	0.03	-0.04	-0.02	-0.13	-0.09	-0.29									
Influencer Age (W_4)	0.08	-0.03	0.04	-0.10	-0.06	0.11	-0.11								
Influencer Gender (W_5)	0.08	-0.03	0.04	-0.10	-0.06	0.11	-0.11								
Adopter Total Degree (Z_1)	-0.06	-0.15	-0.17	0.25	0.03	0.03	-0.29	-0.05							
Adopter Betweenness (Z_2)	0.00	-0.03	-0.03	0.04	0.01	-0.02	-0.01	-0.01	0.38						
Adopter Clustering (Z_3)	0.03	0.39	0.44	0.01	0.00	0.35	-0.09	0.06	-0.37	-0.08					
Age (Z_4)	0.20	-0.06	0.09	-0.15	-0.04	-0.28	0.32	-0.05	-0.19	-0.00	-0.13				
Gender (Z_5)	-0.02	-0.02	-0.03	0.01	0.01	0.00	-0.02	0.19	-0.01	-0.02	-0.01	-0.08			
Population Density (Z_6)	0.24	-0.06	0.12	0.01	-0.02	-0.16	0.05	0.03	-0.06	0.00	-0.09	0.11	0.03		
Network Size (T_1)	0.50	-0.07	0.26	-0.19	-0.05	0.30	-0.11	0.18	-0.11	-0.02	0.12	-0.07	0.00	-0.11	

Table 3: Parameter Estimates of Different Model Specifications

Probability of Adoption		Model 1	Model 2	Model 3	Model 2b	Model 3b
		Mean (z-value)	Mean (z-value)	Mean (z-value)	Mean (z-value)	Mean (z-value)
Network Effects	Degree	4.476*** (215.95)	4.383*** (183.41)	4.476*** (114.27)	4.401*** (182.90)	4.483*** (114.21)
	Clustering	0.361*** (18.02)	0.689*** (32.48)	0.704*** (24.89)	0.703*** (33.06)	0.713*** (25.16)
	Degree × Clustering	-1.916*** (-43.31)	0.552*** (7.33)	1.305*** (10.00)	0.492*** (6.50)	1.274*** (9.76)
Influencer Effects	Avg. Total Degree		-0.00075*** (-17.50)	-0.00090*** (-15.64)	-0.00079*** (-18.18)	-0.00092*** (-15.91)
	Avg. (Total Degree) ² (×0.001)				0.000020 (1.86)	0.000027* (2.27)
	Avg. Betweenness (×0.001)		-0.0067 (-1.88)	0.00025 (0.59)	-0.000075 (-1.89)	-0.0000013 (-0.03)
	Avg. Clustering		3.867*** (28.48)	4.269*** (21.22)	3.777*** (27.70)	4.197*** (20.82)
	Avg. Age			-0.0082*** (-7.24)		-0.0082*** (-7.24)
	Gender (Fraction of females)			0.076*** (3.57)		0.077*** (3.62)
Adopter Effects	Total Degree		0.0010*** (39.02)	0.0011*** (34.78)	0.0012*** (33.57)	0.0013*** (27.63)
	(Total Degree) ² (×0.001)				-0.00027*** (-6.95)	-0.00020*** (-3.98)
	Betweenness (×0.001)		-0.00022*** (-7.02)	-0.00020*** (-5.44)	0.000066 (1.77)	-0.000016 (-0.33)
	Clustering		-2.614*** (-40.86)	-3.322*** (-34.38)	-2.548*** (-39.52)	-3.266*** (-33.59)
	Age			0.0029*** (5.89)		0.0030*** (6.11)
	Gender (Male=0, Female=1)			0.0068 (0.96)		0.0070 (0.98)
	Population Density (City of residence)			0.0012*** (3.97)		0.0012*** (4.06)
Network Size (×0.001)	0.013*** (134.39)	0.012*** (118.44)	0.012*** (92.28)	0.012*** (118.08)	0.012*** (92.18)	
Const.	-8.830*** (-1029.77)	-8.896*** (-455.45)	-8.903*** (-183.19)	-8.922*** (-450.13)	-8.925*** (-182.36)	
Observations	188,631,660	188,628,849	136,829,184	188,628,849	136,829,184	
Pseudo R ²	0.0455	0.0487	0.0496	0.0487	0.0496	

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 4: Comparing model fit over the same set of observations

Probability of Adoption		Model 1	Model 2	Model 3
		Mean	Mean	Mean
		(z-value)	(z-value)	(z-value)
Network Effects	Degree	4.929*** (154.49)	4.535*** (120.68)	4.476*** (114.27)
	Clustering	0.363*** (13.51)	0.690*** (24.47)	0.704*** (24.89)
	Degree × Clustering	-2.144*** (-22.88)	1.312*** (10.06)	1.305*** (10.00)
Influencer Effects	Avg. Total Degree		-0.00087*** (-15.57)	-0.00090*** (-15.64)
	Avg. Betweenness (×0.001)		0.000022 (0.52)	0.000025 (0.59)
	Avg. Clustering		4.516*** (24.20)	4.269*** (21.22)
	Avg. Age			-0.0082*** (-7.24)
	Gender (Fraction of females)			0.076*** (3.57)
Adopter Effects	Total Degree		0.0011*** (34.66)	0.0011*** (34.78)
	Betweenness (×0.001)		-0.00018*** (-5.22)	-0.00020*** (-5.44)
	Clustering		-3.354*** (-34.86)	-3.322*** (-34.38)
	Age			0.0029*** (5.89)
	Gender (Male=0, Female=1)			0.0068 (0.96)
	Population Density (City of residence)			0.0012*** (3.97)
	Network Size (×0.001)	0.013*** (106.69)	0.012*** (94.59)	0.012*** (92.28)
	Const.	-8.929*** (-845.80)	-8.981*** (-367.65)	-8.903*** (-183.19)
Observations		136,829,184	136,829,184	136,829,184
Pseudo R ²		0.0461	0.0495	0.0496
AIC		1,368,561.20	1,363,723.66	1,363,624.20
BIC		1,368,644.87	1,363,907.74	1,363,891.95

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 5: Adoption percentage of predicted adopters

Number of adopters predicted (m)	Successful prediction rate				
	Model 1	Model 2	Model 3	Model 4	Random
500	30.60%	29.60%	28.80%	13.60%	11.66%
1000	26.80%	27.60%	27.20%	13.40%	11.66%
2000	24.45%	24.45%	24.40%	12.35%	11.66%
4000	23.15%	23.32%	23.22%	12.10%	11.66%
9944	20.71%	21.19%	21.19%	13.25%	11.66%

Table 6: Parameter Estimates, Random Effects Complementary Log-log Regression

Probability of Adoption		Model 1	Model 2	Model 3
		Mean	Mean	Mean
		(z-value)	(z-value)	(z-value)
Network Effects	Degree	31.144*** (59.23)	36.963*** (36.60)	35.655*** (28.82)
	Clustering	1.275** (3.41)	3.013** (3.47)	1.400** (2.84)
	Degree × Clustering	-13.635*** (-10.90)	24.959*** (4.23)	25.448*** (5.46)
Influencer Effects	Avg. Total Degree		-0.012*** (-9.67)	-0.011*** (-7.38)
	Avg. Betweenness (×0.001)		0.0054*** (7.35)	0.0040*** (6.81)
	Avg. Clustering		31.374*** (6.33)	23.383*** (4.69)
	Avg. Age			0.012 (0.59)
	Gender (Fraction of females)			0.145 (0.34)
Adopter Effects	Total Degree		0.0088*** (4.48)	0.0081*** (4.91)
	Betweenness (×0.001)		0.0034 (0.52)	0.0064 (1.66)
	Clustering		-32.318*** (-13.60)	-33.609*** (-12.46)
	Age			0.027 (0.96)
	Gender (Male=0, Female=1)			0.067 (0.27)
	Population Density (City of residence)			0.0054 (0.52)
Network Size (×0.001)	0.079*** (45.93)	0.067*** (39.30)	0.088*** (28.52)	
Const.	-21.543*** (-99.44)	-22.411*** (-36.41)	-24.331*** (-17.02)	
Observations	3,780,234	3,772,389	2,730,623	
Individuals	4594	4552	3208	
ρ	0.9741	0.9653	0.9825	

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 7: Parameter Estimates, Correcting for Unobserved Heterogeneity

Probability of Adoption		Model 1	Model 2	Model 3
		Mean	Mean	Mean
		(z-value)	(z-value)	(z-value)
Network Effects	Degree	4.476*** (10.01)	4.383*** (11.54)	4.476*** (12.53)
	Clustering	0.361*** (7.43)	0.689*** (25.18)	0.704*** (70.19)
	Degree × Clustering	-1.916*** (-5.06)	0.552** (2.68)	1.305 (1.69)
Influencer Effects	Avg. Total Degree		-0.00075** (-3.40)	-0.00090*** (-5.53)
	Avg. Betweenness (×0.001)		-0.0067 (-1.86)	0.00025 (0.66)
	Avg. Clustering		3.867*** (11.31)	4.269*** (9.04)
	Avg. Age			-0.0082* (-2.39)
	Gender (Fraction of females)			0.076 (1.91)
Adopter Effects	Total Degree		0.0010*** (10.35)	0.0011*** (20.11)
	Betweenness (×0.001)		-0.00022*** (-8.89)	-0.00020*** (-3.72)
	Clustering		-2.614*** (-13.30)	-3.322*** (-6.28)
	Age			0.0029 (0.76)
	Gender (Male=0, Female=1)			0.0068 (0.26)
	Population Density (City of residence)			0.0012** (3.28)
	Network Size (×0.001)	0.013*** (4.19)	0.012*** (4.20)	0.012*** (4.91)
	Const.	-8.830*** (-27.25)	-8.896*** (-31.45)	-8.903*** (-25.46)
Observations		188,631,660	188,628,849	136,829,184
Pseudo R ²		0.0455	0.0487	0.0496

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 8: Parameter Estimates with Within-City Networks Considered (Models 5 and 6)

Probability of Adoption	Model 5		Model 6	
	Mean	z-value	Mean	z-value
Network Effects				
Degree, Within-City	2.731***	103.35	2.226***	97.05
Degree, Across-Cities	0.070***	4.65	-0.00084	-0.06
Clustering	0.588***	16.35	0.129***	4.10
Degree \times Clustering			6.027***	57.05
Degree, Within-City \times Clustering	-0.067	-0.77		
Influencer Effects				
Avg. Total Degree, Within-City	-0.00034***	-6.73	-0.00040***	-7.84
Avg. Total Degree, Across-Cities	0.000084	1.94	0.000048	1.17
Avg. Betweenness, Within-City (\times 0.001)	-0.00037***	-7.25	-0.00027***	-5.59
Avg. Betweenness, Across-Cities (\times 0.001)	-0.00023**	-2.64	-0.00013	-1.57
Avg. Clustering	3.067***	14.44	4.354***	21.94
Avg. Age	-0.017***	-14.87	-0.016***	-13.34
Gender (fraction of females)	0.010	0.47	0.063**	2.81
Adopter Effects				
Total Degree	0.0014***	40.64	0.0011***	32.47
Betweenness (\times 0.001)	-0.00042***	-8.33	-0.00025***	-5.82
Clustering	-1.520***	-17.52	-4.493***	-46.35
Age	0.0030***	6.02	0.0015**	2.90
Gender	0.0084	1.15	0.0072	0.99
Population Density (city of residence)	0.0073***	23.13	0.0035***	10.87
Network Size, Within-City (\times 0.001)	-0.0088***	-52.45	-0.0087***	-50.97
Network Size, Across-Cities (\times 0.001)	0.087***	253.59	0.087***	259.41
Const.	-8.733***	-182.87	-8.312***	-174.60
Observations	118,612,170		118,612,170	
Pseudo R ²	0.0582		0.0603	

*** $p < .001$, ** $p < .01$, * $p < .05$

Table 9: Parameter Estimates, Non-normalized Degree

Probability of Adoption		Model 1	Model 2	Model 3
		Mean	Mean	Mean
		(z-value)	(z-value)	(z-value)
Network Effects	Degree	0.0071*** (96.95)	0.011*** (96.84)	0.0088*** (61.27)
	Clustering	0.554*** (33.25)	0.333*** (17.17)	0.490*** (18.69)
	Degree × Clustering	0.019*** (40.20)	0.017*** (29.83)	0.020*** (28.34)
Influencer Effects	Avg. Total Degree		-0.00088*** (-19.82)	-0.00065*** (-11.13)
	Avg. Betweenness (×0.001)		-0.000070 (-1.96)	-0.00016** (-3.47)
	Avg. Clustering		-1.808*** (-11.29)	0.532* (2.32)
	Avg. Age			-0.0092*** (-8.35)
	Gender (Fraction of females)			0.036 (1.76)
Adopter Effects	Total Degree		-0.0015*** (-44.84)	-0.0010*** (-23.98)
	Betweenness (×0.001)		0.00014** (3.18)	0.00014*** (4.00)
	Clustering		0.664*** (16.00)	-1.803*** (-20.95)
	Age			0.0083*** (17.95)
	Gender (Male=0, Female=1)			0.0090 (1.27)
	Population Density (City of residence)			0.0059*** (19.99)
	Network Size (×0.001)	0.016*** (170.14)	0.013*** (131.62)	0.016*** (122.57)
	Const.	-8.768*** (-1069.21)	-7.964*** (-374.62)	-8.425*** (-169.13)
Observations		188,631,660	188,628,849	136,829,184
Pseudo R ²		0.0363	0.0392	0.0432

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 10: Uniformity of Network Growth

Percentage Growth of Personal Network During Two Years (748 Days) After Sampling	Coefficient	z-value
Day Joined	-0.0000025	-0.11
Number of Friends in Sample	-0.19	-0.99
Const.	55.53	1.95

* $p < .05$ Number of observations: 93. $F(2, 90) = 0.49$.

Table 11: Parameter Estimates of Different Model Specifications for the First 138,964 Adopters

Probability of Adoption		Model 1	Model 2	Model 3
		Mean	Mean	Mean
		(z-value)	(z-value)	(z-value)
Network Effects	Degree	2.554*** (102.58)	2.481*** (90.66)	2.151*** (49.72)
	Clustering	0.441*** (20.84)	0.694*** (30.91)	0.753*** (25.10)
	Degree × Clustering	-1.103*** (-22.26)	0.621*** (8.06)	1.514*** (11.06)
Influencer Effects	Avg. Total Degree		-0.0011*** (-25.55)	-0.0011*** (-18.70)
	Avg. Betweenness (×0.001)		0.00027*** (8.59)	0.00022*** (5.65)
	Avg. Clustering		2.797*** (20.00)	2.681*** (13.07)
	Avg. Age			-0.0062*** (-5.32)
	Gender (Fraction of females)			0.130*** (5.92)
Adopter Effects	Total Degree		0.00054*** (21.53)	0.00068*** (21.15)
	Betweenness (×0.001)		-0.000030 (-1.47)	-0.000086** (-2.91)
	Clustering		-1.898*** (-30.75)	-2.540*** (-26.88)
	Age			0.0054*** (11.25)
	Gender (Male=0, Female=1)			-0.033*** (-4.58)
	Population Density (City of residence)			-0.00056 (-1.87)
	Network Size (×0.001)	0.027*** (261.41)	0.026*** (237.42)	0.028*** (191.56)
	Const.	-8.553*** (-953.97)	-8.372*** (-425.88)	-8.489*** (-170.33)
Observations		84,784,888	84,782,177	61,276,734
Pseudo R ²		0.0692	0.0708	0.0743

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 12: Parameter Estimates for Condensed Friendship Networks

Probability of Adoption	Friends Adopting within 60 Days			Friends Adopting within 30 Days		
	Model 1 Mean (z-value)	Model 2 Mean (z-value)	Model 3 Mean (z-value)	Model 1 Mean (z-value)	Model 2 Mean (z-value)	Model 3 Mean (z-value)
Network Effects						
Degree	3.742*** (258.07)	3.682*** (214.03)	3.809*** (184.80)	3.066*** (195.11)	2.813*** (148.92)	2.857*** (127.96)
Clustering	0.287*** (15.68)	0.219*** (11.40)	0.163*** (6.99)	0.316*** (16.11)	0.128*** (6.26)	0.068*** (2.78)
Degree × Clustering	-0.758*** (-24.62)	-0.795*** (-17.37)	-0.722*** (-12.70)	-0.658*** (-20.79)	-0.739*** (-16.39)	-0.583*** (-10.55)
Influencer Effects						
Avg. Total Degree		-0.0022*** (-9.74)	-0.0021*** (-7.78)		0.0049*** (12.85)	0.0054*** (11.91)
Avg. Betweenness (×0.001)		0.037*** (19.24)	0.038*** (17.03)		-0.038*** (-8.30)	-0.041*** (-7.59)
Avg. Clustering		0.700*** (16.73)	0.649*** (12.39)		0.515*** (14.47)	0.483*** (11.10)
Avg. Age			0.00050 (0.71)			-0.00053 (-0.72)
Gender (Fraction of females)			-0.099*** (-7.10)			-0.100*** (-7.03)
Adopter Effects						
Total Degree		0.000024 (0.13)	0.00033 (1.66)		-0.0039*** (-10.54)	-0.0038*** (-8.86)
Betweenness (×0.001)		-0.022*** (-14.42)	-0.022*** (-13.07)		-0.080*** (-19.96)	-0.080*** (-17.62)
Clustering		0.0031 (0.09)	0.031 (0.77)		0.209*** (6.76)	0.187*** (4.95)
Age			0.00079 (1.45)			-0.0015** (-2.47)
Gender (Male=0, Female=1)			-0.0046 (-0.60)			-0.0013 (-0.16)
Population Density (City of residence)			0.0021*** (6.43)			0.0014*** (3.87)
Network Size						
Network Size (×0.001)	0.0026*** (29.19)	0.0026*** (26.55)	0.0030*** (24.28)	-0.00028** (-3.01)	0.0020*** (19.33)	0.0023*** (18.15)
Const.	-5.293*** (-474.07)	-5.317*** (-358.38)	-5.418*** (-175.02)	-4.120*** (-333.52)	-4.164*** (-258.03)	-4.125*** (-127.72)
Observations	5,078,035	5,053,555	3,718,551	1,817,996	1,795,121	1,322,861
Pseudo R²	0.0863	0.0875	0.0882	0.0840	0.0912	0.0917

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 13: Parameter Estimates for Models Considering only a Bounded Number of Adoption Decisions

Probability of Adoption	First 60 Adoption Decisions Considered			First 30 Adoption Decisions Considered		
	Model 1 Mean (z-value)	Model 2 Mean (z-value)	Model 3 Mean (z-value)	Model 1 Mean (z-value)	Model 2 Mean (z-value)	Model 3 Mean (z-value)
Network Effects						
Degree	4.686*** (35.30)	4.006*** (27.23)	4.573*** (14.71)	4.890*** (20.57)	3.400*** (12.48)	3.787*** (4.88)
Clustering	0.739*** (13.16)	0.666*** (11.68)	0.696*** (8.90)	1.251*** (13.83)	0.951*** (10.22)	0.974*** (7.65)
Degree × Clustering	-1.188*** (-7.67)	0.087 (0.31)	1.465* (2.46)	-1.502*** (-5.84)	-0.522 (-1.35)	-0.080 (-0.08)
Avg. Total Degree		-0.0016*** (-7.90)	-0.0012*** (-4.03)		-0.0028*** (-8.89)	-0.0026*** (-5.23)
Avg. Betweenness (×0.001)		0.00023 (0.93)	0.00014 (0.40)		0.00094** (2.66)	0.0011** (3.22)
Influencer Effects						
Avg. Clustering		1.763*** (4.11)	3.866*** (4.03)		1.460** (2.99)	2.661 (1.91)
Avg. Age			-0.0022 (-0.50)			-0.0014 (-0.19)
Gender (Fraction of females)			0.019 (0.19)			0.058 (0.38)
Adopter Effects						
Total Degree		-0.0021*** (-9.30)	-0.00027 (-0.92)		-0.0064*** (-13.05)	-0.0027*** (-4.11)
Betweenness (×0.001)		0.00058*** (5.76)	0.000089 (0.15)		0.0022** (2.78)	-0.0028 (-0.55)
Clustering		-1.262*** (-5.19)	-3.521*** (-6.73)		-0.686* (-2.39)	-2.221** (-3.15)
Age			0.029*** (11.76)			0.020*** (4.01)
Gender (Male=0, Female=1)			-0.398*** (-6.55)			-0.236* (-2.22)
Population Density (City of residence)			0.0096*** (3.67)			0.0015 (0.35)
Network Size						
Network Size (×0.001)	0.013*** (14.79)	0.0083*** (8.57)	0.017*** (9.56)	0.017*** (16.77)	0.0088*** (7.87)	0.020*** (9.80)
Const.	-8.311*** (-237.15)	-7.216*** (-76.60)	-8.775*** (-36.79)	-8.433*** (-119.31)	-6.267*** (-42.14)	-7.597*** (-17.73)
Observations	4,553,579	4,553,526	3,265,048	1,185,452	1,185,430	840,973
Pseudo R²	0.0718	0.0793	0.0406	0.1353	0.1637	0.0841

* $p < .05$, ** $p < .01$, *** $p < .001$