The Role of Hubs in the Adoption Processes

Jacob Goldenberg*, Sangman Han**, Donald R. Lehmann*** and Jae Weon Hong**

March 26, 2008

* The Hebrew University of Jerusalem.
** SKK University
*** Columbia University.

The authors thank Gita Johar and Christophe Van den Bulte for their helpful comments.

Reprinted with permission from the *Journal of Marketing*, published by the American Marketing Association, Jacob Goldenberg, Sangman Han, Donald Lehmann, and Jae Weon Hong, March 2009, 73, no. 2.

The Role of Hubs in Adoption Processes

Abstract

The role of hubs (people with an exceptionally large number of social ties) in the diffusion and adoption processes is beginning to be explored in marketing. Using data on a large network with multiple adoptions, we identify two types of hubs - innovative and follower hubs. Contrary to recent arguments, we find that hubs tend to adopt earlier in the diffusion process even if they are not innovative. While innovative hubs have more impact on the speed of the adoption process, follower hubs have more impact on market size (total number of adoptions). Importantly, a small sample of hubs offers accurate success vs. failure predictions early in the diffusion process.

Introduction

Growth processes are important to marketing in general and new product adoption in particular, where the diffusion of an innovation is governed, among other things, by word of mouth. In social systems, growth processes are thought to be strongly influenced by individuals who have large number of ties to other people. One example is the famous ride of Paul Revere who initiated (almost single handed) the American Revolution by riding through the night convincing mainly people he knew to start an armed resistance to the English army (Gladwel, 2000). In the social network literature, such people are called influentials, opinion leaders, mavens or sometimes hubs (Van Den Bulte and Wuyts, 2007) depending on the aspect of influence in question. Somewhat surprisingly, until recently there has been relatively little attention paid to these individuals in the marketing literature on innovation adoption. Further, when the marketing literature does address such individuals, the focus is typically on not how they influence the overall market but rather on either assessing their influence on people they are in direct contact with or identifying their characteristics.

Broadly speaking, influential people are thought to have three important traits: 1) they are convincing (maybe even charismatic), 2) they know a lot (i.e., are experts), and 3) they have large number of social ties, i.e. they know a lot of people. In this paper we focus on the third trait and identify social hubs – individuals who maintain a large number of ties to other people – and their influence on the overall process of innovation adoption. We argue, somewhat contrary to recent suggestions (Watts and Dodds, 2007), that social hubs adopt sooner than other people not because they are innovative but rather because they are exposed earlier to an innovation due to their multiple social links. We further distinguish between innovator and follower hubs and show that the first influence mainly the speed of the adoption in a network while the latter influence mainly the number of people that eventually adopt the innovation. We also show that a small sample of hubs can be used to make an early forecast of the entire diffusion process. Our analysis utilizes a relatively unique data set with large number of diffusion processes on the Cyworld social networking site in Korea.

We begin by reviewing the literature and developing hypotheses. Next we examine how hubs influence the speed of the adoption process as well as the size of the total market. We also demonstrate how adoption by hubs provides an early prediction of the likely success of a new product.

Background

There is growing agreement among practitioners and academics on the fundamental role social networks play in the way information reaches consumers, channel members, and suppliers, (Achrol and Kotler 1999; Iacobucci 1996; Rosen 2000), Van Den Bulte and Wuyts 2007). Recent research has tied social network properties to the success of marketing actions such as pricing or promotion strategies (Mayzlin 2002; Shi 2003). Much of the empirical research in this area has focused on relatively small networks (see Houston et al 2004 for a review), tie strength (Brown and Reingen 1987; Rindfleisch and Moorman 2001) or social capital (Ronchetto, Hutt and Reingen 1989).

Most diffusion models essentially treat the market as homogenous, with the exception of some general adopter categories (e.g., innovators, main market, laggards). This is done for a very practical reason: networks and other more complex structures make modeling and estimation much more complicated. One attempt to examine the impact of a network structures is in the area of "international diffusion" where Putsis, Balasubramanian, Kaplan and Sen (1997) examined how adoption in one country impacts adoption in others and demonstrated the importance of communication within and across countries. While this is an important direction for diffusion models, the size of the network is fairly small and the analysis is at an aggregate level, i.e. the nodes are countries and not consumers. Similar work can be found in the B2B field. For example, Jones and Ritz (1991) suggest a two stage diffusion where first an organization and then individuals within the organization adopt a product (for a different model see Kim and Srivastava, 1998).

Perhaps the best known segmentation scheme in diffusion literature is the classification of consumers into types of adopters (Rogers, 1995). Recently Van Den Bulte and Yogesh (2006) introduced a model of growth using two adopter segments: influentials who affect another segment of imitators whose own adoptions do not affect influentials. This two-segment structure with asymmetric influence is consistent with several papers in sociology and diffusion research (See also Lehmann and Esteban-Bravo, 2006; Goldenberg, Libai and Muller, 2002).

Network Research in Marketing

A social network is defined by a set of actors and the relationships (ties) among them. Often, the importance of an individual can be inferred from their location in the network (Iacobucci, 1998). Although it has been clear for a long time that network analysis is important for understanding growth processes (Rogers, 1976) and that network analysis of field experiments is a promising direction for diffusion studies, relatively little has been done in this direction (Van Den Bulte and Wuyts, 2007).

One approach is to examine both the aggregate growth process as well as what happens at the individual level in a network. Reingen and Kernan (1986) demonstrate how such analysis can be carried out, emphasizing the roles of subgroups and referral flow in small social groups. Brown and Reingen (1987) examined an interpersonal network of a few hundred individuals and related various roles and effects to tie strength. Broadly speaking, most research in marketing has either focused on the interaction within a link (e.g., Narayan and Yang 2006; Reingen 1987) or on aggregate level measures (e.g.., Dholakia et al. 2003, Godes and Mayzlin 2004). Our research links individual measures to the aggregate diffusion process in a large population.

The Role of Key Individuals

Research which suggests that a relatively small number of people have a substantial influence on the opinions and decisions of the majority can be traced back at least 50 years (Katz and Lazarsfeld, 1955). “Opinion leaders" are thought to have expertise in an area. By contrast "hubs" (Valente, 1995, Barabassi, 2003) are individuals with a large number of social ties. Thus while related, these two constructs are different. The literature on opinion leaders is relatively broad and has been examined in a variety of areas including marketing, public opinion, health care, communication, education, agriculture, and epidemiology. There is wide agreement that opinion leaders can have a major impact on opinion formation and change. Summarizing this work, Roch (2005) argues that “a small group of influential opinion leaders may accelerate or block the adoption of a product”.

Weinmann (1991) suggested that influence is as a combination of three personal and social factors. (1) the personification of certain values (or “who one is”); (2) competence (“what one knows”) and (3) strategic social location (“whom one knows”). The first factor is associated mainly with personality traits such as how persuasive is a person.

The second factor is associated with knowledge. Understanding product advantages and/or technical details is often important to people seeking advice. In general, opinion leaders and people who offer advice are more knowledgeable about and enduringly involved with the relevant product class (e.g., Richins and Root-Shaffar 1988; Venkatraman 1988). Myers and Robertson (1972) examined the “knowledgeability” of opinion leaders in twelve categories using 400 households in the Los Angeles area. The correlations between opinion leadership and various measures of knowledge and interest were moderate to high, ranging from a low of .37 (interest in household furnishing) to a high of .87 (knowledge about cosmetics and personal care).

A related concept is market-mavenism (e.g Coulter et al., 2002; Engelland., Hopkins & Larson, 2001; Feick & Price, 1987; Goldsmith, Flynn & Goldsmith, 2003; Steenkamp & Gielens, 2003). Feick and Price (1987) emphasize the knowledge market mavens have about multiple products and places to shop (typically over a variety of categories) as well as their tendency to initiate discussions with consumers and offer information, partially due to their high involvement with the product class. As suggested by Coulter et al. (2002), because opinion leaders are involved in the product category and spend time shopping, they may also acquire more general marketplace expertise.

The third factor is associated with social capital (Burt, 1997) and the type of social connectivity opinion leaders possess. In his work, Burt demonstrates how the value of social capital to an individual is contingent on the number of people doing the same work. An individual's main advantage consists of bridging structural holes – or, disconnections between different "nodes" in a network. Burt argues that people with high social capital stand at the crossroads of a large social organization and therefore have the option of bringing together otherwise disconnected individuals in the network. Because their contacts are more diverse, they are more likely to be a candidate for inclusion in new opportunities. Schott (1987), in examining interpersonal influence in science, suggested that a national community's influence is enhanced by its expertise (indicated by its number of Nobel laureates) and that the influence of one community on another is promoted by collegial and educational ties between them (indicated by co-authorships and student exchanges, respectively). Similarly, Weinmann (1994) suggested that centrally positioned scholars, i.e., scientific opinion leaders, determine the direction of scientific progress because innovations adopted by central figures are more widely accepted by other members of the profession. Opinion leaders in a field tended to be inter-connected, thus creating a powerful "invisible college" that dominates the adoption or the rejection of new scientific models, ideas and methods. Keller and Berry (2003) discuss people who influence others and their relatively large numbers of social links. Similarly, Gladwell (2000) describes "connectors" as people with mega-influence on their surroundings, not because they are experts but rather because they are acquainted with an order of magnitude more people than other people.

Richmond (1977) describes two major explanations for the informational superiority of opinion leaders over followers. The most common explanation is that opinion leaders have more exposure to the mass media than their followers (Katz & Lazarsfeld 1964). An alternative explanation is that opinion leaders acquire more information than non-opinion leaders from the same sources, including personal communications with large numbers of individuals who already have the relevant information.

In this paper we focus on the third factor, connectivity. Specifically we examine individuals that are hubs (have an exceptionally large number of social ties).

The Influence of Social Hubs

A social hub is a person (node in a network) with a large number of ties. The influence of hubs on propagation in networks has been an object of study in the social network field since Rapaport (1953a, 1953b). The network literature refers to propagation or diffusion as the transport from node to node of some quantity, e.g., information, opinion, or disease. The spread of socially-transmitted diseases is one example (see Newman 2002 for a modeling approach based on theoretical physics and Eames and Keeling 2002 for an approach from the bio-mathematics perspective). Rapaport established the influence of network characteristics such as the transitivity of node linking on disease propagation. In Coleman, Katz and Menzel (1966), dependence was assessed via a measure of “embeddedness” akin to the various measures of centrality evolved since Bavelas (1948) and surveyed by Freeman (1978). Studies done in this area include Rogers, Ascroft and Roling (1970) and Rogers and Kincaid (1981). In Barabasi and Albert’s (1999) scale-free model, a few nodes dominate the connectivity of a network due to their extremely large number of ties. The number of ties (connections) is often termed the degree of a node. The distribution of individuals' degrees often follows a power law with a few individuals with the highest degree considered to be hubs. Barabasi and Albert (1999) also show that individuals remain connected even when a large number of links are broken (disconnected) due to the role of hubs. Those individuals with high degree (hubs) should be central in any network. For example, in the case of the spread of a computer virus (Goldenberg, Shavit, Shir and Solomon, 2005), hubs are central in the infection process. The higher the degree of a person, the more neighbors they can impact.

On the other hand, having many links does not necessarily make a person an innovator. Clearly a person who is both innovative and a hub is likely to adopt earlier. Even if a hub is not an innovator, however, there are likely to be a number of individuals that are connected to the hub who adopt the new product early. Such repeated exposure can lead to early adoption by the hub as well, not because of an innovative disposition but rather due to greater exposure (Goldenberg, Shavit, Shir, and Solomon, 2006). Since hubs vary in their innovativeness, there may be different roles for and effects of innovative and less innovative hubs.

Recent work by Watts and Dodds (2007) based on simulation offers a different argument. More specifically Watts and Dodds (2007) report that large cascades of influence can be driven not by hubs (they use the term influentials) but by a critical mass of easily influenced individuals. Nonetheless, in their simulations they found conditions in which hubs are disproportionately responsible for triggering large-scale “cascades” of influence. They emphasize that their results do not exclude the possibility that hubs can be important and suggest that examination of the role of hubs requires more careful specification and testing than it has received so far.

Recently Trusov, Bodapati and Bucklin (2008) examined an internet social network. They found that the average individual in a network is influenced by few other individuals and also influences only a few others. In addition, strong heterogeneity was observed with a small proportion of users participating in a substantial share of the influential dyads identified in the network. More precisely, they found that some users whose total network impact is greater by a factor of 8 more than most other members. However, they did not find that many having links (high degree) makes users influential per se. While their research focuses on network activity rather than adoption processes over the network, it suggests that hubs may indeed be important for diffusion.

Hypotheses

The main purpose of this paper is to examine the role of hubs in diffusion in a natural setting. In order to consider both individual level and aggregate behavior, data is required that includes a large number of individuals (and nodes) and their adoption behavior for multiple products. This paper follows the general perspective of examining an interpersonal network first used by Brown and Reingen (1987). However, we concentrate on the impact of individuals on the aggregate diffusion process rather than on individual adoption.

The previous discussion suggests that there is a small group of people who have a large number of social connections. Thus a basic issue is to identify such hubs. The literature suggests that to some extent opinion leaders are expert and innovative. There are no findings, however, that connect expertise or innovativeness to social connectivity and having a large number of acquaintances.

Our first hypothesis concerns the timing of adoption by hubs. The general two stage model that has been adopted widely is that information moves from the media to opinion leaders, and then from opinion leaders to their followers. We argue that hubs will adopt first because of their greater exposure to an innovation, even it they are not innovators.

Consider a hub with a large number of ties (e.g., 500) who is not innovative and an innovator who has a smaller number of social ties (e.g., 25). Innovators require little exposure in order to make a decision to adopt. Let us assume for this example that an innovator needs 2 product exposures and the hub needs 10. However, because hubs are well connected, their number of indirect exposures to the new product is large even at early stages of the diffusion process. This can create a situation in which the social hub reaches their adoption threshold of 10 exposures before the innovator reaches their threshold of 2. In other words, even hubs who are not innovative may be convinced at early stages of the process to adopt due to their large number of contacts, and mistakenly identified as innovators based on their time of adoption. Thus,

H1: Due to their large numbers of ties, social hubs are more likely to adopt at the early stages of a process.

There is no a-priori reason to assume that all hubs share other traits. More specifically, there is no evidence that personal innovativeness is correlated with social connectivity, i.e., hubs can be innovative and connected or can simply be socially connected. Therefore we distinguish between those hubs that are genuine innovators (innovator hubs) and those who adopt early due to exposure to other adopters (follower hubs). Unlike Watts and Dodds who base the adoption threshold on the proportion of adoptions in a neighborhood to trigger adoption, we use a fixed number as a threshold.

Given that hubs adopt early, and due to their having a large number of links that connect them to a large number of other individuals, their adoption should increase the speed of adoption in the period after they adopt (assuming the product adopted performs satisfactorily). Assume the probability of adoption by an individual (see Goldenberg, Libai and Muller, 2004) is :

$$P= 1-(1-p)(1-q)^{\alpha(t)}$$

Here P is the probability an individual adopts, p is the effect of exposure to external forces (e.g., marketing efforts), q is the impact of word of mouth, (network effects) and $\alpha(t)$ is the number of links to current adopters. The number of the adopters at a time t would be then E(P) x(M-N(t)). This means that an individual with a large number of links (e.g., 500) contributes much more to the adoption through interactions (word of mouth) than an individual with a moderate number of links (e.g., 25). Even if we take the conservative view that social hubs are not more persuasive than other individuals nor that they contact everyone they are linked to, more connections will be activated once they adopt, resulting in a significant increase in adoption rates. In addition, because innovator hubs adopt earlier than follower hubs, obviously they have more time to influence the network. Thus:

H2: When hubs adopt, the overall adoption process speeds up with innovator hubs having a larger effect on speed of adoption than follower hubs.

Hubs are characterized by both in and out degree, i.e., the number of people who talk to them vs. the number of people they talk to. While in degree should be primarily related to when a hub adopts (and is what leads hubs to adopt early), out degree should primarily determine their influence on subsequent adoption. Therefore we hypothesize that, ceteris paribus:

H3: The higher the relative out degree of a hub, the greater impact it has on adoption.

We also hypothesize that social hubs influence market size. Social connections provide indirect information, and the larger the number of connections, the greater the information possessed. From a network point of view, access to diverse resources typically requires that one be connected to diverse actors and sub-networks. Such status can be obtained by having high both *Degree* (number of ties) and *Betweeness* (links to different groups) centralities. These centralities are typically correlated, partially because people with an extremely high degree have a higher probability to be connected to people in different social circles. In general the extent to which someone has an information advantage depends on crossing structural holes, which means linking separate parts of the network (Burt 1992). Put differently, being connected to many interconnected people creates an information advantage from collecting different bits of information sooner than the average network member (Van Den Bulte and Wuyts, 2007).

Even with the conservative assumption that social hubs are not more persuasive than the rest of the members in the network, hubs still have a large number of ties, and therefore potentially more influence than other individuals on people who are not necessarily connected to adopters. If a sufficiently large number of hubs adopt a product, it is more likely that the new product will be exposed to people who otherwise may have not been exposed to it. For example, if we remove hubs from a network, some individuals may not be exposed to the product sufficiently to trigger their adoption. Hence, we argue that adoption by hubs not only increases the speed of adoption but also market size.

Which type of hub has more influence on market size? Follower hubs are more similar to the rest of the population in terms of innovativeness. Therefore their tastes and risk aversion are likely to be more similar to the main market than are those of innovative hubs. Although not too much empirical research exists on this, it seems logical that individuals trust information from similar peers more. A related concept is homophily, defined as the degree to which pairs of individuals are similar in terms of certain attributes, (Rogers 1995). Brown and Reingen (1987) showed that while homophily is related to tie strength, it is a different construct. Homophily fosters trust and reciprocity: it is easier to trust someone who is similar. Although homophily can become a barrier to innovation when different groups are involved, when it exists in a coherent market it can enhance diffusion. Consistent with recent literature that examines the issue of main market vs. early market adoption (see Van Den Bulte and Yogesh, 2006; Goldenberg, Libai and Muller, 2002; and Lehmann and Esteban Bravo, 2006), we argue that innovative hubs will have more influence on the early market while follower hubs will have more influence on the main market. Since the main market is typically at least four times larger, we thus propose that follower hubs have a larger effect on the overall market size:

H4: Hubs adoption increases the eventual size of a market with follower hubs having a larger effect than the innovative hubs.

H2 and H4 would be in contradiction if adoption speed and market size are highly correlated. However, logically they are different. A process can be fast in either a large or small market. In fact, many fads have a rapid adoption rate among a small population which many “really new” products require decades to reach their potential.

If hubs do not adopt a product soon after its introduction, this may impede adoption for those who are connected to the hubs. As a result, adoption rate of hubs at the early stage of the diffusion process increases the probability of success of the product. Thus,

H5: Hub adoption at an early stage can be used to predict product success

Data and Method

Data

In order to examine the role of hubs, data is needed in which a large network is mapped and there is information about the timing of individual (node) adoption for multiple diffusion processes. Fortunately one such data set was available for this research. Specifically we use data from a social network website in Korea—Cyworld.com.

Cyworld was founded in 1999. As of October 2006, there were about 22 million registered members (compared to about 100 million for Myspace). Many people consider Cyworld as a part of every day life with regard to building relationship and sharing information about their life on their home pages. The number of monthly unique visitors is about 20 million in Cyworld (vs. 24.2 million a day in Myspace according to wikipedia.com, 2006 and Businessweek, 2006).

A key aspect of the service, for our purpose, allows people to customize their homepages by including documents, photos, and other "goodies" for free as well as to decorate their minihompy (personal homepage) with paid items such as virtual household items—furniture, electronics, wallpaper, etc. (Cyworld generates money from these paid items and from advertising. People can also adopt items such as pictures or video clips directly from the minihompies they visit (called "scraping" in Cyworld.). In this study we focus on this type of adoption of items.

Measures

A common definition for connectedness in networks is the degree of each node (person in our case), i.e. the number of links (connections) to other nodes in the network. We use degree to identify hubs. We also separately measure out-degree as the number of other nodes ever visited by the hub and in-degree as the number of other nodes that have visited the hub. The data contains information on "scrap" items - item number, time of scrap, and creator ID of each item. By combining the network information with the information on "scrap" items, we track the diffusion process of items over the network. We adopt the Trusov, Bodapati and Bucklin (2008) approach and define links by activity (e.g., visits) and not by pointers such as membership in address books. A link between two individuals in a social networking site (such as LinkedIn or Face Book) does not necessarily imply influence. We define hubs as people with both in and out degree larger than three standard deviations above the mean. (This definition is conservative and may weaken some of the effects as some hubs have a much higher degree). The number of members in the database increased over 5 four-month periods of data from 2,492,036 in December 2003 to 12,685,214 in July 2005. Across those five periods, the number of hubs ranged from 1.28% to 3.30% (averaging 2.63%).

We examined other definitions of hubs. For example, we assumed total connections followed a geometric distribution and computed its mean and variance. We then selected as hubs those individuals whose total connections were three standard deviations above the geometric mean. Because the results were essentially unchanged, we report the simple three standard deviation above the arithmetic mean hub results in the paper.

General descriptive statistics are reported in table 1. We divided the data into 5 time periods and report the means, standard deviations and medians both for in degree and out degree. In and out degree are highly related; the correlation between in and out degree in each period is in the range of .90 - .95.

Insert Table 1 about here

Hubs can be innovative or not. We measured innovativeness based on the adoption timing of each hub across multiple products. Specifically for each product a hub adopted, we measured how many of their neighbors had adopted before them. The lower the number, the more "independent" they are, i.e. they adopt with less social influence.[1] We defined an innovative hub as one who adopted before anyone else in their neighborhood. This resulted in 38.4% of the hubs being classified as innovative. All other hubs are classified as followers. In other words, innovativeness is defined relative to the others a person is connected with.[2]

This definition is imperfect. Consider two hubs, A and B. Assume A is connected mostly to people who adopt early and B is connected to people who adopt later. Even if the

[1] An alternative measure can be constructed based on the number of items adopted. This measure, however, is confounded with both ability to pay and acquisitiveness. For that reason, we assessed innovativeness based on adoption within a social network.

[2] If hubs adopt earlier due to mass exposure rather than innovativeness, a population based measure may classify people who are basically followers as innovators. Since we can examine whether a hub adopts before the 16% of their own directly connected acquaintances, this measure seems most appropriate.

two individuals adopt an identical item at the same time, using our definition B is likely to be identified as "innovator hub" simply because they are linked to "followers". We therefore tested an alternative definition of innovative hubs where innovativeness was determined relative to the entire population and any hub adopting before 16% of the eventual adopters was labeled as innovative. The results based on this population based measure were less strong (e.g., in making early predictions). Therefore we use the neighborhood based definition of innovativeness on the analyses.

We concentrate our analyses on items which produced at least 400 scraps. This means our analyses focused on the 1,067 items which were most popular (out of an unmanageable total of 7,500,488). We defined a highly successful item as one which was in the top 30 items in terms of total adoption. Similarly we defined a moderately (less) successful item as the 30 items whose total adoption was just below the average of the 1,067. This makes our tests conservative ones, vs. one which compares the very best vs. the very worst. The average time for the entire adoption process (time until the last adoption) of the most successful items was 265 days compared to 163 for the somewhat successful ones. This suggests that the diffusion process is quite rapid and the appropriate time interval to analyze this data may be less than a day.

Results

Network structure

Figure 1 presents a sample network with links among 77 individuals from a panel who were selected using the snowball approach which is widely used in mapping networks. The node in the center is connected to a relatively large number of other nodes. Almost any efficient information flow that is to cover the entire sub-net passes through this local hub.

The entire network looks similar to a large version of Figure 1 with many more hubs linked to a large number of other individuals. In order to demonstrate that the entire network has the standard properties of networks, the degree distribution (number of links for each node) was compared to a typical network distribution of links. In most cases this distribution follows a power law, and such networks are termed "scale free" (Barabassi, 2003). Figure 2 presents the degree distribution of connections. As is typical, the distribution is highly skewed; some nodes are hyper connected while the majority of the nodes have only a few connections.

Insert Figures 1 and 2 about here

Most successful processes exhibit a takeoff followed by intensive growth until a peak is reached and then a decay to zero new adoptions. Figure 3 shows the growth pattern of scrap items posted over time; it follows a typical S-shaped diffusion pattern with a fast increase and saturation at the 400,000 level.

Insert Figure 3 about here

Adoption timing

In order to test whether hubs adopt earlier than the rest of the population, we divided each diffusion process into five categories—$t_{5\%}$, $t_{16\%}$, $t_{30\%}$, $t_{50\%}$, and $t_{100\%}$ where $t_{5\%}$ is the time when first 5% of total adoption occurs, $t_{16\%}$ is the time it takes for the first 16% of total adoption to occur, etc. The adoption patterns of hubs and other individuals are shown in Figure 4. As can be seen, hub adoption has a tendency to slowly decline while non hub adoption increases over time. Thus the concentration of hubs among first adopters is larger than among later adopters; hubs on average adopt sooner than non hubs ($t = 16.49$, $p < .0001$).

Insert Figure 4 about here

In order to examine the importance of neighborhood exposure, we compared adoption of hub and non hubs. First we randomly selected 100 hubs and 100 non hubs. We then measured both the number of neighbors who adopted before them and the proportion who did so for the 60 items mentioned earlier. The results are presented in table 2.

Insert Table 2 about here

As expected, hubs adopt earlier in their neighborhood in terms of the proportion of their neighbors that have adopted. However, consistent with our argument, that the number of exposure drives adoption, hubs are not more innovative per se since they on average need 1.68 neighbors to adopt first vs. only 0.61 for non-hubs. Thus hubs appear to adopt early more because of their large number of connections (contacts) rather than innate innovativeness (which in one sense is even below average). This finding can explain the difference between our results and the simulation results in Watts and Dodds (2007). If hubs follow a proportion threshold rule they would indeed adopt later as suggested by Watts and Dodds (2007). However if they follow a number of exposures rule, they adopt earlier because of the large number of connections they are exposed sooner. A detailed comparison based on simulation is provided in the Appendix.

Do Hubs speed the adoption process?

To examine this we selected the 30 most successful items and performed a linear regression of the number of adopters at time t as a function of time, cumulative number of adopters at t-1, the squared cumulative number of adopters at t-1, and the number of new hubs adopting at t-1. The two cumulative adopter terms capture the Bass (1969) model. The scraping process is fairly fast, and the entire adoption process is often concluded in a few months during which millions of people may adopt an item. Separate regressions were performed using a time period of a day, six hours and two hours respectively.

Insert Table 3 about here

The results (Table 3) are notable for several reasons. First, the model fits fairly well (R^2 = .55 to .75), especially so for the two hour time interval, suggesting in this arena imitation occurs rapidly. Second, a hub that adopts in the previous time period has a much stronger influence than a typical adopter, suggesting hubs are indeed critical to growth. Most important, the number of hubs adopting adds predictive power to cumulative adopters and cumulative adopters squared, the terms in the discrete Bass (1969) model.

We also randomly selected 30 items from the same 1067 items and ran the regression analysis. The results are consistent with those of top 30 with one exception. Here, adopters are better predicted using the daily window, reflecting the relatively slower growth of some of these products.

An interesting question is whether the direction of the links is important. The correlation of in degree (number if in links) and out degree (number of our links) is large (over .9). We therefore measured the difference between in links and out links (in–out degree) and added it as another variable to the regression presented in Table 3. The adjusted R^2 increased to .60. The standardized coefficient of the in-out degree variable was - .60 ($p<0.01$), indicating that hubs that have higher out degree than in degree seem to be more effective in speeding the process, consistent with H3. The coefficient of the number of hubs was reduced to .17, close to the cumulative number of adopters coefficient (.21) which can be explained by the high correlation between the number of hubs and in-out degree.

As a further demonstration of the role of hubs in diffusion, we took a sample of 113 hubs. We then observed adoption in their neighborhoods on a daily basis. Using the data up until the day they adopted, we estimated the Bass model and then forecasted the number of adoptions the day after the hub had adopted. We then compared the forecast with actual adoptions. In 106 of the 113 cases, adoption the day after the hub adopted exceeded that predicted based on adoptions up to that point (Average = 1.79 vs. 0.77, $p< .01$). Thus at the neighborhood level, evidence suggests hub adoption speeds overall adoption.

In order to test the relative impact of innovative and non-innovative hubs on the speed of adoption, we examined 1) time to saturation, 2) the time to the inflection point of cumulative adoption $t^* = (p + q)^{-1} \ln (q/p)$, 3) the time to takeoff $t^{**} = - (p + q)^{-1} \ln [2 + \sqrt{3} p/q)]$ and 4) t^*-t^{**}. The last three measures are adopted from Van Den Bulte (2000) and Lim, Choi, and Park (2003) and p and q are the Bass model coefficients that are estimated from the data. We examined the same 60 items. The range of time to saturation was 23 to 546 (average = 214) days. Linear regressions were performed with the four dependent variables and the number of innovative and follower hubs which had adopted as independent variables. Specifically, we measured the number of hubs that adopted in the first 23 days for t16% and 55 days for t50%.

For all 4 measures, both innovative and follower hubs have a significant effect on the speed of the adoption process (Table 4). Moreover, an innovative hub's effect is more than twice that of a follower hub in most cases, supporting our hypothesis.

Insert Table 4 about here

In addition to aggregate effects, the data allows us to examine our hypotheses on an individual level. We isolated the hubs directly linked neighbors and identified their time of adoption. The mean time of adoption of an individual linked to innovator hubs was found to be when 55% had adopted while individuals who are linked to a follower hub adopt at a mean time of .69 ($t = 35.1$, $p<.000$). This re-enforces the argument that innovative hubs drive early adoption.

The relation of hub adoption to market size

We have shown that hubs adopt early and speed up the adoption process. We now examine whether there is a correlation between hub adoption and market potential/size. We again analyzed the 30 most successful scraps and the first 30 just below average of the top 1067 in terms of eventual adoption. The total number of adopters (market size) in each process is the dependent variable. The results are presented in Table 5 with different time frames used as data in the regressions.

Insert Table 5 about here

The R^2 of the regression using the entire process is extremely high (0.99) because of the large number of hubs who eventually adopt (if they are representative they will accurately reflect total demand). However, even after 10 days the number of hub adoptions is a good prediction ($R^2 = 0.88$) of eventual market size. Interestingly, follower hubs have around seven times the impact on market size that innovator hubs have. Thus, in contrast to the results on speed of adoption, follower hubs appear to be more responsible for or at least predictive of the mass adoption of less innovative users and therefore total market size.

Predicting product success

In order to see if we can discriminate between highly and modestly successful products, a logistic regression was performed. The results are presented in Table 6.

The classification table indicated 100% correct predictions. In this case, both the innovative and follower hubs were significant predictors and their coefficients were essentially equal. One explanation for these results is that innovator hubs initiate the process. Hence, if they choose not to adopt an item, they block the adoption process. Follower hubs, by contrast, join only if innovator hubs adopt, giving an item a second "push" in the main market which is necessary for widespread adoption. It is also possible that the follower hubs influence more people because they are more similar in tastes and preferences to the larger population.

Insert Table 6 about here

Overall, prediction is quite good, even when using only hubs that had adopted at 5% adoption level, suggesting early forecasting success based on hub adoption is potentially useful. (Again when the cumulative number of adopters was added as a covariate, the results remained the same, and the coefficient of this variable was not significant). In this analysis, adoption by follower hubs has a stronger link to whether a product eventually succeeds.

In the previous analysis we used all the hubs in the analysis. This is not realistic in terms of managerial applications because firms may not have access to such (complete) data. A more practical alternative would be to focus on a small sample of hubs, (i.e. a hub panel), identifying them through market research (e.g., by examining their previous adoption history). To see if this approach is effective, we randomly selected a small sample of 280 hubs (139 innovative and 141 non-innovative). We then repeated the logistic regression to predict successes vs. failures using this sample. As a stringent test, we examined whether this sample could give successful predictions when only 5% of the market has adopted and compared it to the results based on an equal sized sample of non-hubs. Table 6 presents the result of these both regressions.

Even though the sample we chose is very small compared to the overall population (only 0.006% of the hub population), 70% of the predictions were correct, close to the rate for the entire population of hubs. Further, for the random sample of non hubs, correct predictions occur in only 56.7% of the cases, close to chance and significantly worse than the results for the hubs.

Hubs correlates

Given the importance of hubs, it would be useful to be able to identify them without collecting network data. To try to do this, we used two widely available demographic variables, age and gender. We also used two general characteristics of respondents, namely how long they had been members of Cyworld and how many items in total they had scraped (in essence a measure of acquisitiveness). We compared a sample of 30,723 hubs with 289,001 non-hubs who adopted at least one of the 60 items we studied using logistic regression. Hub status is positively correlated with membership period and total number of acquisition. The results (Table 7) also show that hubs are more likely to be customers who have been active a long time as well as male and younger.

Insert Table 7 about here

Discussion

This paper examined diffusion in a well documented social network. We find that hubs can be identified and classified into two types: innovator and follower hubs. Overall social hubs appear to adopt earlier due to their larger number of connections rather than innate innovativeness.

Adoption by hubs speeds up the growth process and directly influences eventual market size. Innovator hubs influence mainly the speed of adoption while follower hubs mainly influence market size. Moreover, hub adoption serves as a useful predictor of eventual product success; even a small sample of hubs can give reasonable predictions in very early stages. The implication of the last finding is that a firm can collect a sample of hubs and use this panel as a "test tube" for early predictions. This is important because there is generally little indication whether and when a takeoff should be expected. (For exceptions, see Golder and Tellis 1997 and Garber, Goldenberg, Libai and Muller, 2004). It is important to note that this approach should be performed with "minimal invasion". For example, effects like "mere questions" (Levav and Fitzsimons 2006) can contaminate the results by influencing hub behavior.

One application of the findings, other than as an aid to forecasting is to buzz marketing. If social hubs can be identified (and privacy concerns overcome), they could be an efficient target for word of mouth campaigns, leading to both faster growth and increased market size.

One limitation of this study is that the nature of the items is relatively unique and may not generalize to the entire span of innovations and new products. While the network itself and adoption process are very close in structure and dynamics to typical networks and adoption processes, clearly replication is called for in different contexts. Specifically it is important to test the influence of hubs on the dynamics of adoption with more "classic" products which defuse more slowly, although mapping the relevant network may be a serious problem.

Since data sets on large networks with multiple processes are hard to construct, other approaches like tracing hubs and examining their behavior in a longitudinal study may be a useful avenue for future research. Another future direction is to develop methods to identify hubs in a network based on penetration and other data, i.e. general descriptors. In addition, accounting for characteristics of the items themselves (either as fixed or random effects) is worth pursuing at some point.

The value of a customer to the firm is more than the sum of their purchases, it also includes the effect that some individuals, i.e. hubs, have on others. Such “influentials” have substantially higher value than previously realized. Hopefully the results here will encourage work to better understand the motivations, behaviors, and impacts of well connected individuals in their social structures.

Tables and Figures

Table 1: Out and In Degree

Period	Membership	Arithmetic Mean	Arithmetic Std	Geometric Mean	Geometric Std	Median
1	2492036	13.76	19.53	7.07	3.60	5
		13.77	18.79	7.14	3.56	6
2	5941393	19.92	25.32	10.92	3.64	11
		20.04	24.73	10.40	3.79	11
3	8675741	21.43	25.12	12.53	3.48	13
		21.59	24.54	11.74	3.69	14
4	10264533	20.66	25.82	12.59	3.34	12
		20.90	24.92	10.98	3.77	13
5	12685214	17.34	24.64	12.14	3.05	8
		17.71	23.18	7.80	4.23	8

Table 2: Comparison between hubs and non hubs adoption timing (mean and std) within Neighborhood.

	Number of Neighbors who Adopt Earlier	Proportion
Hubs	1.68 (2.10)	0.01 (0.01)
Non-Hubs	0.61 (0.96)	0.25 (0.36)

Table 3: Determinants of adoption: a regression analysis with number of adoptions at time t as a dependent variable using 3 different time unites. All coefficients are standardized

Time period	Time	Cumulative number of adopters at t-1	Square of Cumulative number of adopters at t-1	# of hubs adopting at t-1	Adjusted R^2
Daily	-0.03**	0.31”**	-0.30**	0.75**	0.59
Six hours	-0.03**	0.31**	-0.27**	0.72**	0.55
Two hours	-0.01**	0.18**	-0.15**	0.85**	0.75

** denotes significant at the .01 level

Table 4: Hub adoption predictions of the diffusion speed. Regression analyses with dependent variables: time to saturation, time to first inflection point (peak) time to takeoff and the difference between takeoff and peak. All coefficients are standardized. For each dependent variable two regressions were performed: 1) with data until 16% of the process and 2) data until 50% have adopted.

Dependent Variable	Time Until Adoption	Innovative Hubs	Follower Hubs	Adjusted R^2
Time to saturation	$t_{16\%}$	-0,54**	-0.25**	0.43
	$t_{50\%}$	-0.63**	-0.30**	0.56
t*	$t_{16\%}$	-0.54**	-0.24*	0.42
Time to peak	$t_{50\%}$	-0.63**	-0.30**	0.56
t**	$t_{16\%}$	-0.31**	-0.14**	0.10
Time to Takeoff	$t_{50\%}$	-0.37**	-0.24**	0.2
t^{*} - t^{**}	$t_{16\%}$	-0.64**	-0.38*	0.60
	$t_{50\%}$	-0.70**	-0.27*	0.63

** significant at .01 level, * significant at .05 level

Table 5: Predicting market size based on early hub adoptions, all coefficients are standardized. The four regressions are using data accumulated up to 5%, 10%, 15% and 100%.

Time frame	Innovative Hubs coefficient	Followers Hubs coefficient	Adjusted R^2
$t_{5\%}$ (10 days)	0.11 (.n.s)	0.83 (.00)	0.88
$t_{10\%}$ (17)	0.15 (.00)	0.85 (.00)	0.99
$t_{15\%}$ (23)	0.12 (.00)	0.89 (.00)	0.99
$t_{100\%}$ (214)	0.14(.00)	0.86(.00)	0.99

(the numbers in parentheses are the significance levels)

Table 6: Logistic regressions to predict large vs. moderate success with all hubs and a small sample of hubs, using data accumulated within 5%, 10% and 15% adoption. The fifth regression is based on non hubs, for comparison.

Data	Time of Hub Adopting	Innovators hubs	Followers hubs	Tenure	Correct predictions
All sample hubs	$t_{5\%}$ (10 days)	0.01 (.34)	0.06 (.07)	0.01 (0.3)	81.7%
All full sample hubs	$t_{10\%}$ (17)	0.01 (.34)	0.06 (.03)	0.01 (0.03)	80.0%
All full sample hubs	$t_{15\%}$ (23)	0.01 (.21)	0.05 (.05)	0.01 (0.03)	80.0%
Small sample hubs	$t_{5\%}$ (10)	2.53 (.02)	11.97 (.97)		70.0%
Small sample non hubs	$t_{5\%}$ (10)	1.25 (.29)	11.64 (.97)		56.7%

(The numbers in parentheses are the significance levels).

Table 7 : Predictors of Being a Hub (Dependent variable : hub=1, non-bub=0)

Parameter	Estimate	Error	P value
Intercept	-1.011	0.026	0.000
Age	-0.087	0.001	0.000
Gender (Male)	0.415	0.013	0.000
Membership Period	0.001	0.000	0.000
Total of Items Acquired	0.016	0.001	0.000

Figure 1: A sample of 77 individuals and how they are linked.

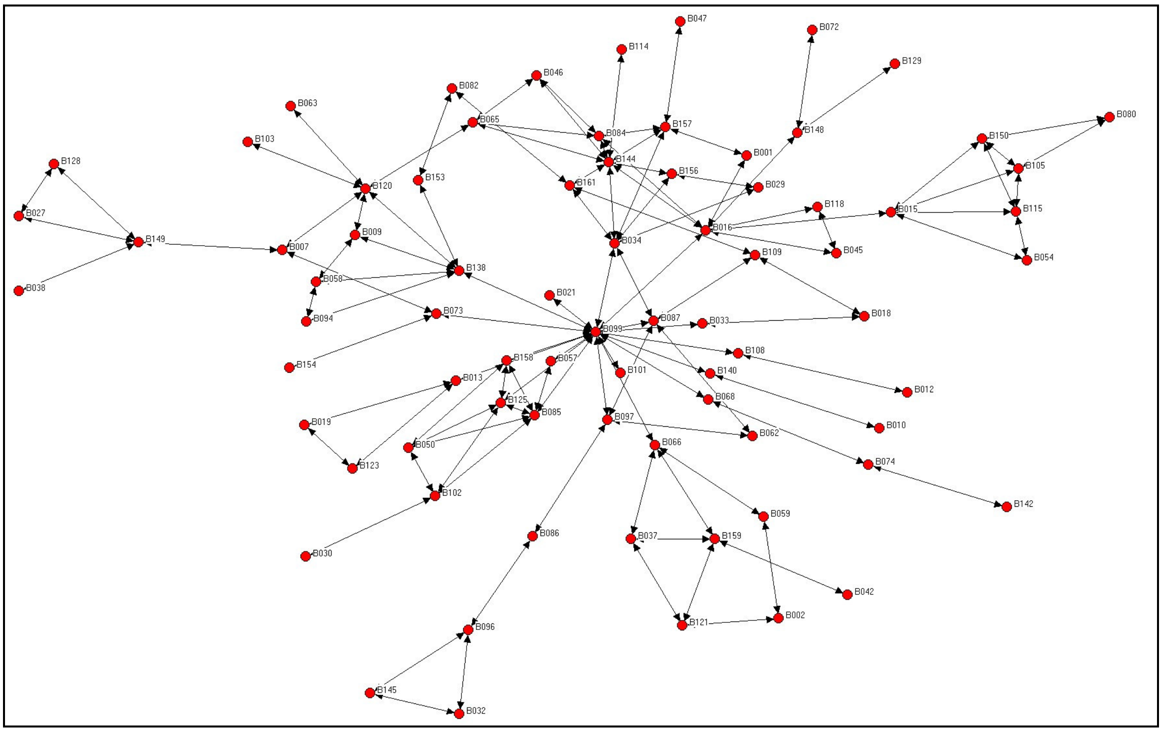


Figure 2: Degree distribution of the network.

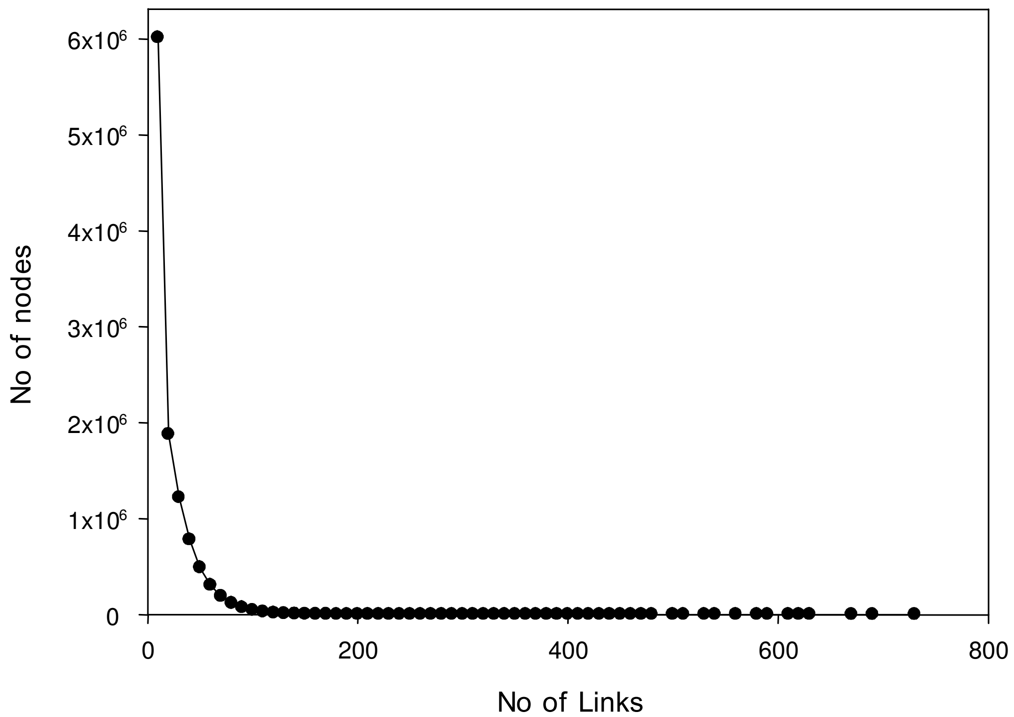


Figure 3: Number of items posted over time.

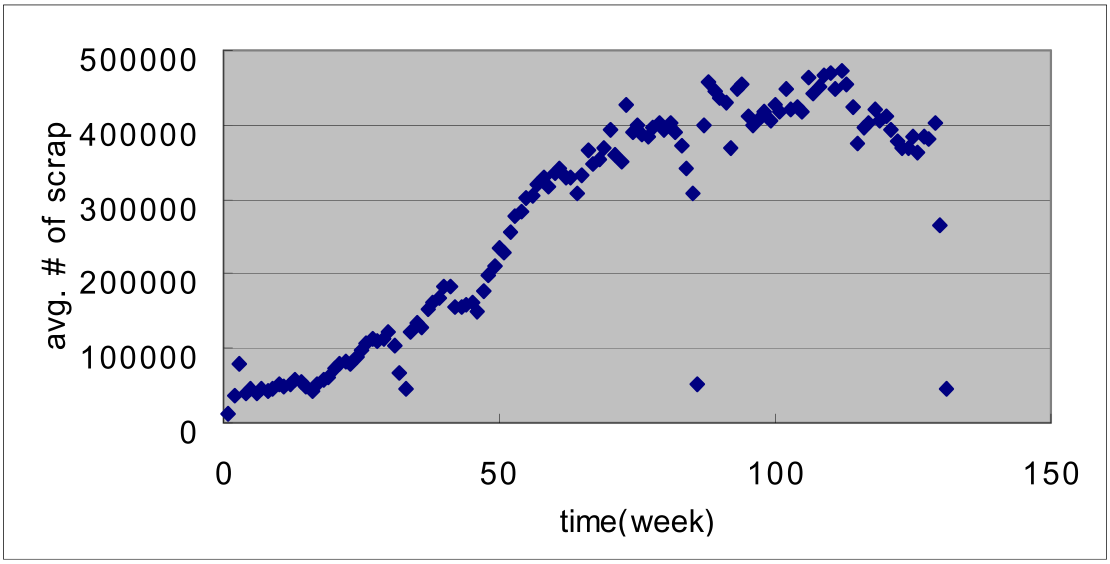
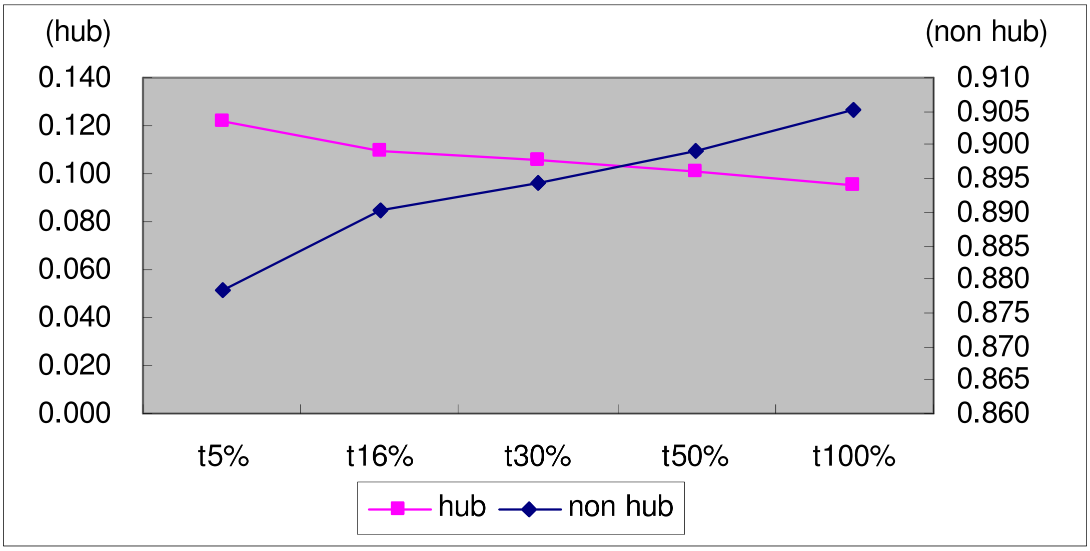


Figure 4: Adoption rate of Hubs and non-Hubs (in percentages) as a function of time (non dimensional).



References

Achrol, Ravi S. and Philip Kotler (1999), Marketing in the Network Economy, *Journal of Marketing*, Vol. 63; 146-163.

Achrol Ravi S.(1991) Evolution of the Marketing Organization: New Forms for Turbulent Environments, *Journal of Marketing*, Vol.55, 4, 77–94.

Barabassi, Alber-Laszio, (2003), Linked, New York: Plume Penguin Books.

Barabasi, Albert-Laszlo and Reka Albert (1999), Emergence of scaling in random *Networks, Science*, 286, 509-512.

Bass, Frank M. (1969), A New Product Growth Model for Consumer Durables, *Management Science*, 15, 215-227.

Batagelj. V. and A. Mrvar (1998). . Pajek: A program for large network analysis, Connections, 21(2), 47-*57*.

Bavelas, A. (1948) A mathematical model for group structure", Human Organizations, 7, 16-30.

Borgatti and Foster (2003), The Network Paradigm in Organizational Research: A Review and Typology, *Journal of Management*, 29, 991-1013.

Brown, Jacqueline Johnson; Reingen, Peter H. (1987) Social Ties and Word-of-Mouth Referral Behavior, *Journal of Consumer Research*; Dec 1987; 14, 3, 350-62.

Burt Ronald S, (1992), Structural holes: The Social Structure of Competition, Harvard University Press, NY.

Burt Ronald S, (1997), The Contingent Value of Social Capital, *Administrative Science Quarterly*, Vol. 42, Iss. 2; pg. 339-58

Coleman, James S., Elihu Katz, and Herbert Menzel. (1966), Medical Innovation: A DiffusionStudy, New York: Bobbs Merrill.

Coulter, Robin A., Lawrence F. Feick, and Linda L. Price (2002) Changing Faces: Cosmetics Opinion Leadership Among Women in the New Hungary, European *Journal of Marketing*, 36 (11/12), 127-1308.

De Sola Pool & Kochen (1978) Contacts and influence, Social Networks, 1, 5-51.

Eliashberg, Jehoshua, Jedid-Jah Jonker, Mohanbir S. Sawheny, and Berend Wierenga (2000) MOVIEMOD: An Implementable Decision-Support System for Prerelease MarketEvaluation of Motion Pictures, *Marketing Science*, 19 (3), 226-243.

Engelland, Brian T., Christopher D. Hopkins, and Dee Ann Larson (2001) Market Mavenship as an Influencer of Service Quality Evaluation, *Journal of Marketing Theory and Practice*, 9(4), 15-26.

Feick, Lawrence F. and Linda L. Price (1987) The Market Maven: A Diffuser of Marketplace Information, *Journal of Marketing*, 51(1), 83 –97.

Garber Tal, Goldenberg Jacob, Barak Libai and Muller Eitan, (2004), From Density to Destiny: Using Spatial Analysis for Early Prediction of New Product Success, *Marketing Science* Vol. 23, 3, p419-29.

Gladwell Malcolm (2000), The Tipping Point, NY: Little, Brown and Company.

Godes David and Dina Mayzlin (2004), Using Online Conversations to Study Word-of-Mouth Communication, *Marketing Science*, 23(4), 545-561.

Goldenberg Jacob, Barak Libai and Eitan Muller (2002), Riding the Saddle, How cross-Market Communications Creates a Major Slump in Sales, *Journal of Marketing*, Vol. 66 (April), p 1-16

Goldenberg Jacob, Barak Libai, and Eitan Muller (2004), Complex yet Simple: CellulaAutomata as an Enabling Technology in marketing Strategy Research, in Christine Moorman and Donald R Lehmann, eds, Assessing Marketing Strategy Performance, Marketing Science Institute, Cambridge, MA.

Goldenberg Jacob, Yuval Shavitt, Eran Shir,Sorin Solomon (2005), Distributive Immunization of Networks Against Viruses Using the 'Honey Pots' Architecture, Nature Physics 1, Dec, 184-188.

Goldsmith, Ronald E., Leisa R. Flynn, and Elizabeth B. Goldsmith (2003), Innovative Consumers and Market Mavens, *Journal of Marketing Theory and Practice*, 11(4), 54-65.

Herr Paul M., Frank R. Kardes and John, Kim (1991), Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: an Accessibility-Diagnosticity Perspective, *Journal of Consumer Research*, 17 (4), 454-462.

Houston, Mark, Mike Hutt, Christine Moorman, Peter Reingen, Aric Rindfleisch, Vanitha Swaminathan, and Beth Walker (2004), Marketing Networks and Firm Performance, in Assessing Marketing Strategy Performance, eds. Christine Moorman and Donald R. Lehmann, Cambridge, MA: Marketing Science Institute, 227-268.

Iacobucci, Dawn (1996), Networks in Marketing, Sage Publications.

Jones, J. Morgan, and Christopher J. Ritz (1991) "Incorporating Distribution into New Product Diffusion Models", *International Journal of Research in Marketing,* Vol.8, 2, 91-113.

Katz, Elihu and Paul F. Lazarsfeld,. (1955), Personal influence: The Part Played by People in the Flow of Mass Communications. Glencoe, Ill.: The Free Press

Keller, Ed and Jon Barry (2003), The Influentials: One American in Ten Tells the other Nine How to Vote, Where to Eat, and What to Buy, New York: The Free Press.

Kim Namwoon and Rajendra K. Srivastava (1998), Managing Intraorganizational Diffusion of Technological Innovations, *Industrial Marketing Management*, 27, 229–46.

Krider, Robert and Charles Weinberg,. (1998), Competitive Dynamics and the Introduction of New Products: The Motion Picture Timing Game, *Journal of Marketing Research*, 35 (1), 1-15.

Lazarsfeld, Paul F., Berelson, and Hazel Gaudet (1944), The People's Choice, New York: Duell, Sloan, and Pearce.

Lehmann, Donald R. and Mercedes Estelban-Bravo (2006), When Giving Some Away Makes Sense to Jump Start the Diffusion Process, *Marketing Letters*, 17, December243-254.

Levav, J. and G. J. Fitzsimons (2006), When Questions Change Behavior: The Role of Ease of Representation, *Psychological Science*, 17 (3), 207-13.

Lim, Bycong-Lak, Munkee Choi, and Mycong – Cheol Park (2008), The Late Take-off Phenomenon in the Diffusion of Telecommunication Service: Network Effect and Critical Mass, *Information Economics and Policy*, 15, 537-557.

Lusch, Robert F and Nicholas S. P Tay (2004), Agent Based Modeling: "Gaining Insights into Form and Industry Performance", in Christine Moorman and Donald R *L*ehmann, eds, Assessing Marketing Strategy Performance, *Marketing Science* Institute, Cambridge, MA.

Mahajan, Vijay, Eitan Muller and Yoram Wind (2000), New-Product Diffusion Models, Dordrecht: Kluwer Academic Publishers

Mahajan Vijay, Eitan Muller, and Roger A. Kerin (1984), Introduction Strategy for New Products with Positive and Negative Word of Mouth, *Management Science*, 30, 1389-1404.

Mayzlin, Dina (2002), The Influence of Social Networks on the Effectiveness of Promotional Strategies, Working Paper: Yale School of Management
Newman, M. E. J. The Spread of Epidemic Disease on Networks. Physics Review E 66, 16128.

Putsis P. Wiliam, Balasubramanian Seidhar, Kaplan H. Eduard and Sen K Subrata (1997), Mixing Behavior in Cross Country Diffusion, *Marketing Science*, Vol. 16, 4, 354-69.

Rapaport, A. (1953), *Spread of information through a population with sociostructural bias: assumption of transitivity*. Bull. Math. Biophys. 15, 523–533.

Reingen, Peter H., and Jerome B Kernan,, (1986), Analysis of Referral Networks in Marketing: Methods and Illustration, Journal of Marketing Research, Vol.23, 4, 370 – 9.

Reingen, Foster, Brown and Seidman (1984) Brand Congruence in Interpersonal Relations: A Social Network Analysis. Journal of Consumer Research 11, 3. D. J. Watts 1999. Small Worlds. Princeton University Press, Princeton.

Richmond, Virginia P. (1977) The Relationship Between Opinion Leadership and Information Acquisition, *Human Communication Research*, 4(1), 38-43.

Rindfleisch, Aric and Christine Moorman (2001) The Acquisition and Utilization of Information in New Product Alliances: A Strength-of-Ties Perspective, *Journal of Marketing*, 65(2), 1-18.

Rogers and Kincaid (1981), Communication Networks: A New Paradigm in Research. New York: Free Press.

Rogers M., Evertt (1976), New Product Adoption and Diffusion", *Journal of Consumer Research*, Mar .Vol.2, . 4, 290-202.

Rogers, Ascroft & Roling (1970), As reported in Valente, Network Models of the Diffusion of Innovations", New Jersey: Hampton Press

Rosen, Emmanuel, (2000), The Anatomy of Buzz: How to Create Word-of-Mouth Marketing, New York: Doubleday Currency

Shi, Mengze (2003), Social Network-Based Discriminatory Pricing Strategy, *Marketing* Letters, 14(4), 239 –

Steenkamp Jan-Benedict E. M. and Katrijn Gielens (2003), Consumer and Market Drivers of the Trial Probability of New Consumer Packaged Goods, *Journal of Consumer Research*, 30(3), 368-384.

Valente W.Thomas, (1995), Network Models of the Diffusion of Innovations, Hampton Press, New Jersey

Van den Bulte, Christophe (2000), New Product Diffusion Accelerations Measurement and Analysis, *Marketing Science*, 19 (4), 366-380.

Van den Bulte, Christophe and Joshi V.Yogesh, (2006), New Product Diffusion with Influentials and Imitators *Marketing Science*, forthcoming

Weimann, Gabriel (1994), The Influentials: People Who Influence People. Albany: State University of New York Press.

Weimann, Gabriel (1991), The Influentials: Back to the Concept of Opinion Leaders? *Public Opinion Quarterly*, 55(2), 267-279.

.

Appendix

One way to study dynamics in a system is to use Agent Based Modeling (ABM).Interactions between agents (adopters in our case) lead to distinct collective phenomena, whose so-called emergent properties can be described at the aggregate level (Lusch and Toy, 2004, Goldenberg Libai and Muller, 2004).

Here each agent is a potential adopter of the innovation. The probability that a given agent adopts an innovation at time t given that it has not yet adopted depends on two factors: external marketing (such as advertising), represented by parameter p, and internal social interaction (or word of mouth), represented by parameter q.

Given $k_i(t-1)$ is the number of agent i's neighbors that have adopted at time t-1, and $f_i(t)$ is the probability of agent i will adopt at time t. Garber, Goldenberg, Libai and Muller (2004) use the following specific form of f:

$$fi(t) = 1 - \left[(1-p)(1-q)^{k_i(t-1)}\right] \qquad (2)$$

We sampled one representative cluster (sub network) in the data and generated an agent based model. The model used as adoption threshold; when the number of neighbors exceeded a threshold, then the agent adopted. Specifically, the number of nodes (equivalent to adopters in the data) and their tie distribution was implemented in a net-logo software framework. The hubs are defined by the data itself. Thus contrary to most uses of ABM, here the simulation is tied to the sampled network.

Some limited empirical and experimental evidence supports the assumption that individuals follow threshold rules when making decisions in the presence of social influence. We sampled 2655 members from the network of adopters of the 60 items we studied using a snowball sampling approach – each identified adopter was a starting point from which to identify associates in the network. We based the distribution of degrees in the simulation on these numbers. In order to define the adoption threshold of adoption for each node (the number of exposures through linked adopters required to generate adoption), we used the threshold distribution in the sub-network and randomly assigned thresholds to the nodes. In addition, we kept the same proportion of hubs (2.63%) that there were among the 2655 members in the analysis. In order to start this process, seeding (assigning nodes that adopt at t_0) is required. We used seed ratios of 0.1% ~ 3%, similar to the range of the values of p found in the Bass model.

Results

Table 1 shows the average adoption time for hubs and regular nodes in the simulation. With the exception of the lowest seeding values, hubs were more likely to adopt, and when they did adopt, to do so earlier.

Table A1: Simulated adoption pattern

Seed ratio (%)	Percent Adopting	Average adoption Percent		t	P value	Average adoption time		t	P value
		Hub	Non Hub			Hub	Non-Hub		
0.1	13.6	14.4	10.2	1.4	0.167	3.8	4.1	0.8	0.414
0.3	6.8	9.8	6.7	2.5	0.017	1.8	2.4	3.6	0.001
0.5	15.6	17.3	15.6	2.6	0.013	3.0	3.6	2.5	0.015
0.7	14.2	22.8	14.0	10.5	0.000	2.3	2.9	4.6	0.000
1.0	22.3	26.3	22.2	4.8	0.000	2.1	2.8	5.0	0.000
1.5	20.9	30.0	20.7	21.2	0.000	1.9	2.4	5.6	0.000
2.0	14.4	23.7	14.2	9.6	0.000	1.8	2.1	4.6	0.000
2.5	21.9	27.7	21.8	8.6	0.000	1.9	2.5	5.1	0.000
3.0	22.3	30.2	22.0	12.1	0.000	1.8	2.2	6.0	0.000