December 2002

DIFFUSION OF THE INTERNET HOSTS

Yi-Cheng Ku
Fortune Institute of Technology

Her-Sen Doong
National Sun Yat-sen University

DIFFUSION OF THE INTERNET HOSTS

Yi-Cheng Ku
Information Management Department
Fortune Institute of Technology
ycku@center.fjtc.edu.tw

Her-Sen Doong
Information Management Department
National Sun Yat-sen University
hsdoong@mail.hamastar.com.tw

Abstract

The purpose of this study is to explore the extent to which exploitation of the global diffusion of the Internet hosts can be adequately described by traditional mathematical models of the diffusion of an innovation. Internal-influence model, mixed-influence model, and two exponential growth models are applied to examine the adoption pattern. The research results demonstrate that the diffusion pattern of Internet global hosts is best described by the internal-influence model which illustrates the "social system" as a competitive rather than a cooperative context. Though imitation is a very popular phenomenon in the business community, the growth of Internet hosts has the effect on increasing the value of participation in the process. Within the diffusion process of Internet hosts, adoption is a very powerful and effective agent for disseminating information about itself to the community of users.

Keywords: Innovation diffusion; diffusion model; Internet

Introduction

Internet is one of the most important innovations over the last decade of the 20th century: with which data become available anytime and any place. People and organizations can communicate with each other, trade and learn online. Internet changes not only the life style of people but the enterprise's business model. Since people give Internet an important position, the host number of Internet grows by times every year. According to the Network Wizards (http://www.nw.com), there were 376,000 Internet hosts in January 1991. However, as Internet becomes part of our daily life, the number of Internet hosts has shockingly increased to 109,574,429, growing nearly 300%, in January 2001. And the pick of increase numbers counted per six months showed in July 2000, it increased 20,649,693 units in six months. Therefore, there are more than 125 million hosts in July 2001 over the world.

Who, then, pushes the growth of the Internet? Network externalities could offer a good explanation (Gurbaxani 1990): The value of the network increases with an increase in the number of users it connects. Moreover, the commercial application may be another critical reason. Now, the most proportion of Internet traffic is accounted for commercial use, it does not fit to the network's original purposes in research and education (Goodman et al. 1994).

Internet's global diffusion is a practical issue that also has a significant impact on information system development. Therefore, it has drawn a great deal of attention from both practitioners and researchers. The prior researchers have done many investigations related to Internet diffusion (Gurbaxani 1990; Goodman et al.1994; Rai 1998). Nevertheless, Internet is growing so rapidly that it is difficult to predict the development trend under the early stage. We argued that it is necessary to re-evaluate the Internet diffusion phenomenon after many EC companies went out of business in late half of 2000.

Rai et al. (1998) adopted Logistic, Gompertz and Exponential Models to predict the development of Internet hosts with the data regarding the number of hosts from August 1981—July 1997 at quarterly intervals. As the result shown, Exponential model conveys the situation better than the others because the Internet is still in the early stages of its diffusion (Rai et al. 1998). Rai's study has made an important contribution to researches on the Internet diffusion. However, as Table 1, the growth rate of Internet hosts has flatted down from July 1999, and Internet changes a lot during the last 4 years. In order to understand the new trend of Internet, this study tries to re-examine the diffusion pattern of Internet hosts with new data and evaluation techniques.

Table 1. Internet Hosts Growth Data by Half a Year, January 1988-July 2001

Time Period	Number of Hosts	% Growth	Time Period	Number of Hosts	% Growth
1988/1/1	28,863	N/A	1995/1/1	5,846,000	82.00
1988/7/1	33,000	14.33	1995/7/1	8,200,000	40.27
1989/1/1	80,000	142.42	1996/1/1	14,352,000	75.02
1989/7/1	130,000	62.50	1996/7/1	16,729,000	16.56
1990/1/1	197,500	51.92	1997/1/1	21,819,000	30.43
1990/7/1	274,500	38.99	1997/7/1	26,053,000	19.41
1991/1/1	376,000	36.98	1998/1/1	29,670,000	13.88
1991/7/1	535,000	42.29	1998/7/1	36,739,000	23.83
1992/1/1	727,000	35.89	1999/1/1	43,230,000	17.67
1992/7/1	992,000	36.45	1999/7/1	56,218,000	30.04
1993/1/1	1,313,000	32.36	2000/1/1	72,398,092	28.78
1993/7/1	1,776,000	35.26	2000/7/1	93,047,785	28.52
1994/1/1	2,217,000	24.83	2001/1/1	109,574,429	17.76
1994/7/1	3,212,000	44.88	2001/7/1	125,888,197	14.89

Data Source: Internet Software Consortium (http://www.isc.org)

Literature Review

Studies of the Internet Diffusion

Diffusion is defined as the process by which an innovation is communicated through certain channels over a period of time among the numbers of a social system (Rogers 1995). There are four critical factors -- innovation, communication channels, time and social system— that accelerate the diffusion process. The first one, Innovation, often means a new idea or object that have not been previously explored or adopted. The second one, communication channel, makes messages pass from one individual to another. Therefore, people can obtain the innovation messages from the mass media channels, interpersonal channels or both. And the third one is time that means an innovation's rate of adoption in a system. It's determined by five variables: perceived attributes of innovations, type of innovation-decision, communication channels, nature of the social system and extent of change agents' promotion efforts. The last factors is a social system that is defined as a set of interrelated units that are engaged in joint problem-solving to accomplish a common goal (Rogers 1995).

Traditionally speaking, innovation diffusion research centers around the above four factors and their interrelationships. Since the time that the members of a social system adopted the innovation differs, according to diffusion of innovation theory, the distribution of adopters over time is expected to be a bell-shaped curve. We can divide adopters as five categories by the different standard deviation from the mean, including innovators, early adopters, early majority, late majority and laggards.

Diffusion Model

Diffusion models are mathematical models that can be applied to depict the successive increase in the number of adopters or adopting units over time (Mahajan and Peterson 1985). Therefore, they can be used to explain the diffusion pattern, predict future distribution of an innovative technology, illustrate possible effect of a policy, or market a new product. The basic diffusion model can be expressed as equation (1). The equation (1) indicates that diffusion rate is a function of the difference between absolutely potential adopters and cumulative adopters. The diffusion rate decreases as the actual adopters increase. The coefficient of diffusion, g(t), is the probability that potential adopters will adopt the innovation at time t. It is often affected by the nature of innovation, communication channel and others elements. Therefore, $g(t)[m - N(t)]$ is the number of adopters at a given time

t, expressed as n(t). In previous research, g(t) has two kinds of definitions: one is the function of t; the other one is the function of cumulative adopters, i.e. g(t)=a+bN(t)+cN(t)2+…

$$\frac{dN(t)}{dt} = g(t)[m - N(t)]. \tag{1}$$

Where: N(t) is the cumulative number of adopters, $N(t) = \int_{t_0}^{t} n(t)dt$;

$N(t= t_0) = m_0$, the initial number of adopters in a social system;

$\frac{dN(t)}{dt}$ = diffusion rate at time t;

g(t) is the coefficient of diffusion;

m = total number of potential adopters in a social system;

Mahajan and Peterson (1985) indicated three general diffusion models: the External-Influence model, Internal-Influence model and the Mixed-Influence model. External-Influence model means the diffusion rate is affected by the mass media. The coefficient of diffusion g(t) is a constant, and the diffusion model becomes the equation (2). Internal-Influence model hypothesize that the interpersonal communication has a great influence on the diffusion rate. Earlier adopters influenced the later adopters. The coefficient of diffusion g(t) is a function of the cumulative number of adopters, and the diffusion model was expressed as the equation (3). Mix-influence model, equation (4), combined the effect of external and internal influence. All three general diffusion models are in the following:

- g(t)=p, $N(t) = m(1 - \exp(-pt))$ (2)

- g(t)=qN(t), $$N(t) = \frac{m}{1 + \frac{m - m_0}{m_0}\exp(-qmt)} \tag{3}$$

- g(t)=p+qN(t), $$N(t) = \frac{m - \frac{p(m - m_0)}{p + qm_0}\exp(-(p + qm)t)}{1 + \frac{q(m - m_0)}{p + qm_0}\exp(-(p + qm)t)} \tag{4}$$

Since three general diffusion models are built under several hypotheses, it would appear to be some conflict between assumptions underlying the models and the context of the Internet diffusion phenomenon (Rai 1998). For example, the population of the social system is constant and homogeneous; the Internal-Influence model assumes that external factors do not affect the diffusion process, and so on (Mahajan and Peterson 1985). Those assumptions may restrict the validity of this study. In addition to three general diffusion models, we adopt two kind of Exponential model to fit the Internet diffusion pattern depending on Rai (1998) (referred to hereafter as Rai98) study.

Parameter Estimation Consideration

Application of any diffusion model involves estimating its parameters. Parameters can be estimated in nonlinear estimation procedures (Srinivasan and Mason 1986) or maximum likelihood estimation procedures (Schmittlein and Mahajan 1982). We adopted nonlinear procedures in SAS statistic software to estimate all parameters by historical data. Because diffusion models are built under some kind of patterns rather than random process, it is necessary to verify the distribution of diffusion patterns after we found the parameters. Mahajan et al. (1988) provided the White-Noise model, equation (5), to be the null hypothesis in order to test the validity of diffusion model. It means that diffusion model is useful when it can explain better than the White-Noise model on the innovation-adopting process.

$$x(t)=x(t-1)+e(t) \tag{5}$$

Where: x(t) is the number of adopters at time t;
e(t) is assumed to be $N(0,£m^2)$

We adopt J-test (Davidson and Mackinnon 1981) to test of influence models against White-Noise model, because White-Noise model is linear and J-test was designed to test the specification of an econometric model in the presence of an alternative model that purport to explain the same phenomenon.

Data and Methods

Data

We collected Internet hosts growth data of Internet Software Consortium by half a year, January 1988-July 2001, to estimate the parameters of diffusion model. After comparing with the Rai98 data set, it shows different data set between this study adopted and Rai98's in following Table 2.

Table 2. The Different Data between This Study Adopted and Rai98's

Date	Hosts	Rai98 Hosts
1995/1/1	5,846,000	4,852,000
1995/7/1	8,200,000	6,642,000
1996/1/1	14,352,000	9,472,000
1996/7/1	16,729,000	12,881,000
1997/1/1	21,819,000	16,146,000
1997/7/1	26,053,000	19,540,000

As the Internet Software Consortium website describes, there are two kinds of Internet Domain Surveys (http://www.isc.org/ds/new-survey.html):

> *The Internet Domain Survey has been taken twice a year since 1987. The original survey methodology counted hosts by walking the domain name tree and doing zone transfers of domain data in order to discover hosts and further subdomains. It is described more completely in RFC1296. The old survey counted the number of domain names that had IP addresses assigned to them.*
>
> *However, by July 1997 the Domain Survey was not able to count a significant portion of the hosts in the domain system, due to some organizations restricting download access to their domain data. The blocking of downloads (or zone transfers as they are called) had increased to the point where in the July 1997 survey we could only download 75% of the domains we discovered. We decided to try a new survey technique before the old one became useless.*
>
> *In January 1998, we ran the first "new" Internet Domain Survey. The new domain survey is the reverse of the old survey. It counts the number of IP addresses that have been assigned a name. This distinction is subtle but it does mean the new survey is counting a different "thing" than the old survey.*

Due to the differences between the past and present survey, they adjust the past domain survey hosts counts in order to make a comparison. Therefore, our research based upon "adjusted hosts count" from January 1995 to July 1997, and the hosts count of new domain survey from January 1998 to July 2001.

Methods

The purpose of this study is to explore the extent to which exploitation of the Internet hosts can be adequately described by traditional mathematical models of the diffusion pattern of an innovation. Internal-Influence model, External-Influence model,

Mixed-Influence mode, and two Exponential growth models are applied to examine the adoption pattern. Since the External-Influence model failed to converge, we adopted the other four models to fit the diffusion pattern of Internet.
Moreover, we used two types of examination that are conducted to determine which model best fits the observations of Internet hosts diffusion. The first one estimates the model parameters of all four diffusion models with the nonlinear least squares regression method. The second determines the true model among the four alternatives using the J-test and the minimum difference between the predictive value and actual growth data.

NLIN, a nonlinear regression procedure of the SAS software package, is used to estimate the parameters in the four diffusion models. And it summarizes the parameter estimation and the model fit for the four models in Table 3. As Table 3 shows, the Internal-Influence model has an R^2 of 0.9973 which is the proportion of variance explained by this model. The Mixed-Influence has an R^2 of 0.9980; the Exponential-1 model has an R^2 of 0.9978; the Exponential-2 model has and R^2 of 0.9870. To draw a conclusion from Table 3, the R^2 of Mixed-Influence model is higher than the other models. After the model parameters are estimated, the J-tests is conducted to use each model in turn as the null hypothesis and the rest as the alternative hypotheses.

Table 3. Diffusion Results for Internet – Parameter Estimation and Model Fit

Parameter	Internal	Mixed	Exponential-1	Exponential-2
m	200,090	242,950		
m_0	0.213	0		
p		1.8128E-06	0.21089852	0.276266964
q	1.656E-06	1.1459E-06		
A			15.20553261	
R^2	0.9973	0.9980	0.9978	0.9870

Unit: 1,000 hosts
NOTE: 1. Exponential-1=A*exp(pt)
2. Exponential-2=exp(pt)

Table 4 presents the results of comparisons of the alternative specifications with the White-Noise model. The J-test result rejects the null hypothesis that the adoption pattern of Internet hosts is random process. All four alternative models can reject the White-Noise model at the significance level of $p<0.001$ or better. Thus we come to a conclusion that the diffusion of Internet hosts is not a random process.

Table 4. J-test Model Comparisons: t Values with Probabilities

J-test Model Comparisons: t Values with Probabilities				
Null Model	Alternative Models			
White Noise	Internal	Mixed	Exponential-1	Exponential-2
t statistic	12.2909	12.7701	11.7195	4.5648
p-value	0.0000****	0.0000****	0.0000****	0.0001****

Note: **** $p<0.001$

Best Model and Discussions

Rai98 adopted actual growth data to examine the predictive validity of the three diffusion models, not parameter estimation. As their results shown, the Exponential model performs better than the Logistic and Gompertz models. However, we use the data set from January 1998 – July 2001 to examine the predictive validity of the models. As the Table 5 shows, Internal-Influence model performs better than the others.

Mahajan and Peterson (1985, p.18) noted that the Internal-Influence perspective is "most appropriate when an innovation is … socially visible, and not adopting it places social system members at a disadvantage". With the rapid development of electronic

commerce, many companies adopted Internet hosts for the purpose of imitation. They imitated competitor's websites, Intranet and Inter-organization systems to avoid losing competition advantages. Even customers will ask suppliers to provide the Internet services. Many government institutions push business to adopt network technologies to be e-business, so that to raise the competition advantages of the organization. Many enterprises built the multimedia websites to promote or sell product. As the MIT surveyed (Fox 1995), there were 62.5% websites built for commerce use in America in 1995. Moreover, researcher pointed out that 83% advertiser among the top 100 in Britain had Internet addresses.

Furthermore, virtual communities and email attract many people to use Internet. Commercial service provider, such as America Online has enhanced the accessibility of the network to a broad spectrum of users. In 1997, there were approximately 4,000 Internet service providers in North America (Press 1997). Some of these ISPs provide services that make the network be attractive to groups ranging from school children to retired people. Those users may require their friends of social network to use Internet; especially email has become a popular communication tool for most people. This is a strongly relationship for computer networks where the value of the network increases with an increase in the number of users it connects. The Internet adoption is communicated among the members within the social system via interpersonal channels.

Table 5. Model Forecasts vs. Actual Number of Hosts by Half of Year, January 1998 – July 2001

Date	Actual vs. Predicted Number of Hosts				
	Actual	**Internal-Influence**	**Mixed-Influence**	**Exponential –1**	**Exponential -2**
1998/1/1	29,670,000	27,862,421	31,073,620	30,151,986	20,860,458
1998/7/1	36,739,000	36,797,632	39,430,085	37,231,282	27,498,251
1999/1/1	43,230,000	47,802,956	49,508,595	45,972,704	36,248,187
1999/7/1	56,218,000	60,872,725	61,386,113	56,766,499	47,782,350
2000/1/1	72,398,092	75,738,691	75,008,148	70,094,538	62,986,680
2000/7/1	93,047,785	91,840,147	90,152,020	86,551,827	83,029,025
2001/1/1	109,574,429	108,381,215	106,415,960	106,873,074	109,448,838
2001/7/1	125,888,197	124,475,727	123,247,307	131,965,488	144,275,429
Mean sum of errors		2,280,977	3,355,824	2,730,454	8,926,344

In fact, mass medias also play a very important role in the earlier diffusion stage. These mass medias publicize the advantage about using Internet widely in several years ago, but the effect of mass medias are not outstanding in the adoption process. As discussed above, our research result clearly indicates that the Internal-Influence model best describes the diffusion process of Internet.

Conclusion

In 1969, the experimental ARPANET was developed by the U.S. Department of Defense consisted of four host computers located in the United States. 213 host computers in approximately a half dozen NATO countries were connected in 1980. By 1989, less than a half dozen years after the ARPANET migrated out of the Department of Defense and in reality became the Internet, connectivity jumped to approximately 20 countries and more than one hundred thousand hosts. Up till now the potential of Internet as a vehicle for the support of commerce is realized. Why has it taken so long? Can the traditional mathematical models of innovation diffusion explain this time lag? Is the Internet comparable to any other innovation in terms of the factors which have influenced its diffusion? Does it possess unique features that challenge some fundamental assumptions of traditional mathematical models of innovation of diffusion?

In this study, we aim to explore the extent to which exploitation of the Internet hosts can be adequately described by the features traditionally identified in the literature as critical to the diffusion of an innovation. Our analysis explores the generalized mathematical model of the diffusion of an innovation as developed by Mahajan and Peterson (1985). The research results demonstrate that the diffusion pattern of Internet global hosts is best described by the Internal-Influence model which illustrates the "social system" as a competitive rather than a cooperative context. Though imitation is a very popular phenomenon in the

business community, the growth of Internet hosts has the effect on increasing the value of participation in the process. Within the process of Internet hosts, adoption is very powerful and effective agent for disseminating information about itself to the community of users.

References

Davidson, R. and MacKinnon, J. G. "Several Tests for Model Specification in the Presence of Alternative Hypotheses," Econometrica (49:3), 1981, pp. 781-793.

Fox, B. "Retailing on the Internet Seeking Truth Beyond the Hype," Chain Stone Age (71:9), 1995, pp. 33-72.

Goodman, S., L. Press, S. Ruth, and Rutkowski, A. "The Global Diffusion of the Internet: Patterns and Problems," Communications of the ACM (37:8), 1994, pp. 27-31.

Gurbaxani, V. "Diffusion in Computing Networks: The Case of BITNET," Communications of the ACM (33:12), 1990, pp. 65-75.

Mahajan, V. and Peterson, R. A. Models for Innovation Diffusion, Beverly Hill, CA: Sage Publication Inc., 1985.

Mahajan, V., Sharma, S. and Bettis, R.A. "The Adoption of the M-Form Organizational Structure: A Test of Imitation Hypothesis," Management Science (34:10), 1988, pp. 1188-1201.

Press, L. "Tracking the global diffusion of the Internet," Communications of the ACM (40:11), 1997, pp. 11-18.

Rai, A., Ravichandran, T. and Samaddar, S. "How to Anticipated the Internet's Global Diffusion," Communications of the ACM (41:10), 1998, pp. 65-75.

Schmittlein, D. and Mahajan, V. "Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance," Marketing Science, 1982, pp. 57-78.

Srinivasan, V. and Mason, C. H. "Nonlinear Least Squares Estimation of New Product Diffusion Models," Marketing Science, 1986, pp. 169-178.