Demand and Supply-side Dynamics of Piracy Diffusion in P2P Networks

JooHee Oh, Marshall School of Business, University of Southern California

Il-Horn Hann, R.H. Smith School of Business, University of Maryland

I. Introduction

Digital piracy, the sharing files of music on peer-to-peer (P2P) networks has been cited as the key threat to the music industry by distorting promoting plans of albums (Billboard.biz, 2009). To combat widespread digital piracy on file-sharing networks, the content industries have employed diverse anti-piracy strategies including pollution technologies and stronger copyright legislation for music CDs, DVDs, games, etc. For instance, companies such as Overpeer, deliberately deteriorate song file dispersion by spoofing or corrupting files. Organizations such as the RIAA aim to reduce the potential impact of so-called pre-release piracy, where content on P2P networks becomes available before the official launch date. Much effort by the music and film industry is directed to achieve a supply and demand reduction through technology and legal means. Legal and monitoring activities against digital piracy have focused on the deterioration of the supply side. Technologies such as spoofing or corrupting files decay the supply level of desired content. Legislation also mostly aims at penalizing file uploaders instead of file downloaders. However, there is little understanding in how supply and demand are formed in P2P networks. Specifically, there is little understanding whether monitoring and legal efforts to curb piracy in P2P networks is effective.

We began our work by asking how the supply-side of file sharing behavior, the provision of files for downloads, affects the demand-side of file sharing, the number of file requests and vice versa. The purpose of this study is to describe the demand- and supply- dynamics of piracy diffusion processes of an individual *title* (an album or a song) on P2P networks. Many studies have developed various approaches to explain the diffusion of new products and technology adoption which incorporate various internal and external factors, e.g., advertising, promotion and WOM (Bass 1969; Givon et al. 1995; Mahajan et al. 2000). Certain limitations in the existing models call our attention to explain digital piracy diffusion. For instance, studies found that the initial phase of the product introduction processes hardly fit into the Bass model (Goldenberg et al. 2009). What previous models lack in their representation of piracy diffusion are the unique properties of file sharing behavior and the topological system characteristics. Unlike other processes, users' willingness to share their song files significantly influences the piracy diffusion process regarding the intensive legal and monitoring attention on the part of file-suppliers; the uploading behavior and the fraction of users who is willing to share downloaded songs critically shape the diffusion processes.

In this study, an individual title level diffusion model has been developed to represent unique properties of piracy on P2P systems. Our unit of analysis is *aggregated* file-sharing behavior of peers for a single title, instead of a user or the market in general. In particular, we are interested in how the shape of aggregated data depicts patterns in both demand- and supply- side file-sharing behavior per title. Our contention is that the shape characteristics of the title-level diffusion can be represented as interlinked processes of demand- and supply-side behavior with parameters representing properties of piracy. We utilize the dynamics between the demand- and supply- side file sharing behavior of albums over time. We empirically illustrate the shape of aggregated piracy behavior using functional data analysis, pioneered by Ramsay and Silverman (2005). This functional approach enables us estimate the parameters representing the size of supply-side impact on the demand-side behavior and vice versa as described in the differential equations system. By applying functional approach in the diffusion model, we empirically illustrate the dynamics of piracy diffusion processes and generate out-of-sample forecasts of digital piracy for the new music titles.

II. Model Development: Piracy Diffusion Processes in P2P Networks

This study examines the dynamics of file sharing behaviors. Given our focus on aggregate effects of file sharing behavior per title, we develop an empirical model that represents the characteristics of a network structure. Modern P2P networks display a "hybrid" architecture. File sharing behavior can be measured by two different ways in the hybrid architecture; one is by counting number of hash-requests by leaf-nodes. Another is by counting total number of nodes, representing individual users, who supply the song file to the requesting node. In a hybrid structured P2P network, how many file requests will be expanded from an incremental increase in supplying node or how fast total supplying nodes will be expanded from an incremental increase in file-request for a certain song? To answer the questions, we examine two processes of diffusion; first, a demand process of file-downloading requests and secondly, a supplying process of file provision nodes. As we shall see, there is an intricate relationship between the proliferation of the downloading requests and the proliferation of the available supplying nodes.

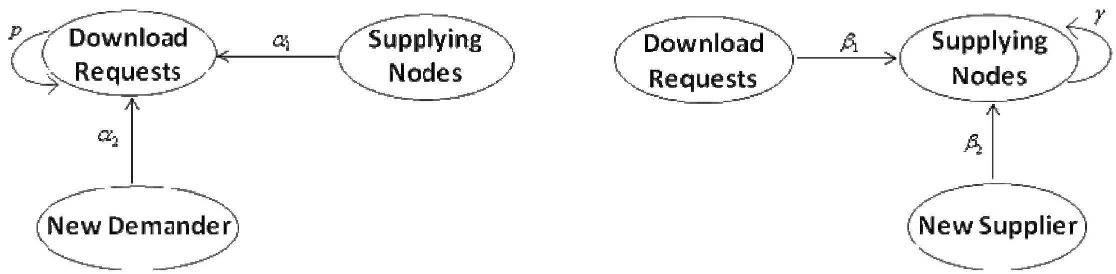
The Process of Demand- Side of File sharing Behavior

The rate of demand-side process of adoption depends on the external effect of the mass communication media and the internal effect of word-of-mouth of adopters on non-adopters (Bass 1969; Givon et al. 1995; Mahajan et al. 2000). In the case of piracy, the downloading request rate with respect to the potential demand directly affects the demand process. A popular title will be requested for downloads with higher rate than a nonpopular title. The parameter α_2 in the Figure 1 characterizes this effect of demand potential. Word-of-mouth also influences the new potential demanders' download request rate. A prevalent title will create more imitation effect on new demanders, thereby highly requested than a nonpopular title. The parameter p denotes WOM effect of adopters. One particular aspect of piracy is that the demand-side downloading request can only be realized if the supplying side file exists. Not only the existence, but also the number of supplying nodes, that represents the availability of files, influences the speed of file-downloads. If the available files are spoofed or corrupted, or hardly connected to the requesting node, the downloading request would be broken down. The parameter α_1 stands for the supply side impact on file demand. Figure 1 illustrates these three main features that govern the demand-side downloading request rate.

The Process of Supply- Side of File sharing Behavior

The file-supplying side on P2P networks has been critical in piracy dispersion; legislators and anti-piracy pollution companies have been mainly targeted supplying behavior. In response to such strategies, unless a peer moves downloaded files to a personal folder, most recent P2P networks allow users to share their downloaded song files with others by default. The rate of supply-side process, the change in total number of supplying nodes, determined by two processes; a influx of new suppliers who did not initially belong to supplying nodes and by conversion of downloaders who intentionally or nonintentionally share their files. The potential ultimate number of file suppliers affects the file-supplying rate by the influx process, denoted as the parameter β_2. Also the fraction of downloaders who likely be converted to the supplying nodes alters the rate of file-supplying process, indicated as the parameter β_1. Lastly, the retention rate of supplying peers to be remained as supplying nodes, determines the sustainability of supplying process considering heavy legal and monitoring attentions. If the cost of file-supplying cancels out the benefit of supplying, the outflow rate of supplying nodes can be significantly increased. The supply side process of piracy diffusion can be described as shown in Figure 1.

[Figure 1: Diffusion Processes of P2P Piracy: Demand- side vs. Supply- side]



α_1: Coefficient of file availability representing the impact of supplying-nodes on downloading demand

α_2: Coefficient of demand potential representing the impact of piracy potential on downloading demand

p : Coefficient of demand imitation representing the WOM influence of peers who already adopted files

β_1: Coefficient of sharing propensity representing the portion of supplying node among downloaders

β_2: Coefficient of supplying potential representing the impact of potential piracy supply on current nodes

γ: Coefficient of retention representing the remaining propensity of supplying nodes without outflux

D_H (Download Requests): Total number of downloads for a title at time t

D_S (Supplying Nodes): Total number of supplying nodes for a title at time t

$\overline{H} - D_H$ (New Demander): Potential market demand-Cumulative number of downloaders

$\overline{S} - D_S$ (New Supplier): Potential market Supply-Cumulative number of supplier

A Model of Diffusion Processes

The number of download-requests at time t, $D_H(t)$ increases when a peer, who do not have copy of a title, requests a file downloading of the title, which happens with rate $\alpha_2[\overline{H} - D_H(t)]$. The potential effect of WOM from existing peers who already downloaded the title is represented by the coefficient p.The WOM influence on the downloading request rate can be simply represented by $pD_H(t)$ or can be formulated as $p\frac{D_H(t)}{\overline{H}}[\overline{H} - D_H(t)]$ similar to the Bass model. The downloading response rate from available supplying nodes, increase the number of downloads, with rate $\alpha_1 D_S(t)$. This leads to the equations (1)-(1)' presented below. Note that without this supply-side effect on demand-side equation, denoted as $\alpha_1 D_S(t)$, the demand process in the equation (1)' parallels the Bass model. Similarly, the number of supplying nodes for a song, $D_S(t)$, increases when a peer who is not supplying files for a certain title, turned into a supplying node, which happens with rate $\beta_2[\overline{S} - D_S(t)]$; in addition $D_S(t)$ increases when a peer who downloaded the song, decided to become also a supplying node for the song, which occurs with rate $\beta_1 D_H(t)$. The fraction of supplying nodes sustaining as file suppliers takes place with rate $\gamma D_S(t)$. While the other two terms, depicts dynamics between the supply level and the rate, the term $\beta_1 D_H(t)$, measures the demand-side impact on the supply-side. This leads to the differential equation (2). The differential equation system (1)-(2), a simplified version, has a closed form solution. In turn, the system (1)'-(2) can be empirically illustrated by using a functional approach.

(1) $$\frac{dD_H(t)}{dt} = pD_H + \alpha_1 D_S + \alpha_2[\overline{H} - D_H(t)] = (p - \alpha_2)D_H + \alpha_1 D_S + \alpha_2 \overline{H}$$

(1)' $$\frac{dD_H(t)}{dt} = p\frac{D_H}{\overline{H}}[\overline{H} - D_H(t)] + \alpha_1 D_S + \alpha_2[\overline{H} - D_H(t)] = [\overline{H} - D_H(t)]\left[\frac{D_H}{\overline{H}} p + \alpha_2\right] + \alpha_1 D_S$$

(2) $$\frac{dD_S(t)}{dt} = \gamma D_S + \beta_1 D_H + \beta_2[\overline{S} - D_S(t)] = \beta_1 D_H + (\gamma - \beta_2)D_S(t) + \beta_2 \overline{S}$$

Hypotheses

Each node in P2P networks functions as both a client and a server. Supplying nodes, one of our units of analysis, consist of peers or nodes who share their downloaded song files. In the provision side, positive network externality is exhibited when an additional supplying node added to the system not only provides song files but also introduces extra resources in the system (Boever 2007). Despite huge differences among peers with respect to processing, connection speed, local network configuration or operating system, each member of the P2P network has the same functionality at the application layer (Karagiannis et al. 2004) and the network resources are typically provided to all peers. Supplying nodes in this study only include peers who share the file, thereby representing nodes that introduce and share extra resources. In the case, increase in supplying nodes for the title results in providing more resources, e.g., song files, bandwidth to the file requesters. Thereby, more number of supplying nodes will drive more benefits of peers who demand the song file. This implies that the availability of supplying nodes for a title positively affects the rate of file-demand through the parameter, α_1 in the above equation (1)'.

Hypotheses 1 (Availability of Supplying Nodes) The rate of file demand increases with respect to the file supplying level, the number of supplying nodes, i.e., $\alpha_1 > 0$.

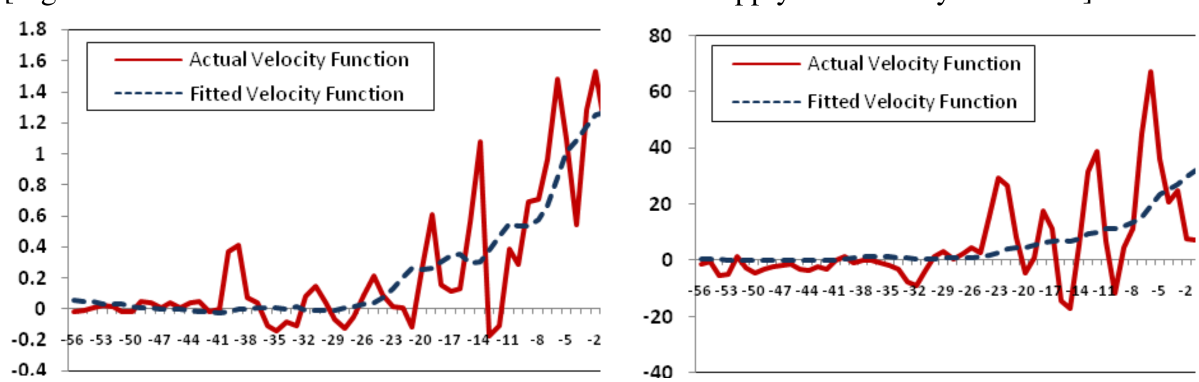
Studies discussed underlying motivation of file-sharing behavior on P2P networks. While Krishnan et al. (2004) commented on the free-rider problem and potential under-provision of sharing resources, they also mentioned the possibility that peers may share their content based entirely on self-interest; it is to draw traffic away from other peers in the network to the sharing peer, thereby increasing the chance that the sharing peer will be able to get her desired content from other peers on the network. What we seek here is not how or why individual decide to share files; our focus is on the aggregated impact of the number of download requests on the number of supplying nodes. While not all peers who downloaded song files will turn into supplying nodes, unless the conversion ratio of downloaders to become supplying nodes dramatically shifts, as the number of downloading request increases, the number of supplying nodes will increase. Whether individuals purposefully make decisions on sharing or eventually find themselves as supplying nodes by default and update their status, the conversion ratio of downloaders will not be changed; individuals' preference parameter over sharing decision or average number of status examination over P2P usage remain the same without drastic changes in preference systems or environments. This implies that the number of downloading requests for a title positively affects the rate of file-supply through the parameter, β_1 in the equation (2).

Hypotheses 2 (Sharing Propensity of Download-Requesters) The rate of file supplying increases with respect to the file demand level, i.e., $\beta_1 > 0$.

III. Empirical Results

We empirically illustrate the nearly continuous file-sharing data using penalized smoothing methods. Here, we focus on characterizing the shape of prerelease file-sharing data for the newly released albums. We used file-sharing data obtained from the Ares P2P network for the time period of April 2007 to September 2007 for the albums on the Billboards' Top 200s list. We first recovered the underlying functional objects of demand- and supply-side file sharing data for an album from two months prior to the release date from the discrete daily observations. To test our hypotheses in the Part II, we obtained the mean function of demand- and supply-side aggregated file-sharing data for per title. The Figure 2 depicts the first-derivative functions, the velocity functions of mean level file-sharing functions. The velocity function depicts volatility involved in the daily rate change of demand- and supply-side file sharing behavior. The dotted line in the Figure2 depicts the fitted value from the estimation based on equations (1)'-(2).

[Figure 2: First-Derivative Functions of Demand- and Supply- side Piracy Diffusion]



From the functional objects of file-sharing behavior, we tested our hypotheses by estimating the parameters in the differential equation system (1)'-(2). We found that the parameters α_1 and β_1, the coefficients representing the cross impacts between demand- and supply-side file sharing behavior, shows positive value, as summarized in the Table1.

[Table 1: Demand-side Equation (1)'and Supply-side Equation (2)]

(Demand) Parameters	p	α_1	α_2
Coefficients (Pr (>|t|))	0.0393 (0.391)	0.0026 (1.04e-05***)	-0.0182 (6.18e-08***)
Adj. R-sq.(p-value)	0.7457 (< 2.2e-16)		

(Supply) Parameters	γ	β_1	β_2
Coefficients (Pr (>ltl))	0.0228 (0.7165)	1.4087 (0.4978)	-0.0133 (0.0329 *)
Adj. R-sq.(p-value)	0.3319 (1.955e-05)		

IV. Conclusion

We develop a diffusion model to explain and predict the diffusion of digital piracy for a title. Our model represents known characteristics of file-sharing behavior and topological properties of P2P systems. While the demand process of file-sharing process in the model resembles the Bass model, we take into account the supply-side impact of availability that has been spotlighted in current anti-piracy efforts. Not only the demand-process influenced by the supply-side, the supply-side also critically affected from the demand-process. The differential equation system well illustrates the effect of supply-side constraint on the demand rate and vice versa. From a simplified version of differential equation system (1)-(2), we will derive analytical properties of the closed form solution. For the applied version of system (1)'-(2), we empirically tested our hypotheses based on functional data analysis. Our results show the limitation of Bass model in characterizing piracy diffusion because of the supply-side effect. Using the functional approach, we could directly estimate the parameters in the system from the derivative curves of file-sharing level functions. In the future, we will further describe empirical strategy to predict the level of piracy diffusion in a song-level based on the relationship between prerelease and post-release piracy behavior.

V. References

Asvanund,A., Clay,K., Krishnan,R., Smith, M.D., (2004). An Empirical Analysis of Network Externalities in Peer-To-Peer Music-Sharing Networks, *Inform. Sys.. Res* (15:2), pp. 155-174.

Bass, F.M., (1969). A New Product Growth Model for Consumer Durables, *Management Science* (15:5), pp.215-227.

Boever, J.D. (2007). Peer-to-Peer Networks as a Distribution and Publishing Model, *Proceedings ELPUB2007 Conference on Electronic Publishing, Vienna, Austria*

Givon, M., V. Mahajan, E. Muller. (1995). Software piracy: Estimation of lost sales and the impact on software diffusion. *Journal of Marketing* 59(1) 29–37.

Goldenberg, J., Lowengart, O., and Shapira,D., (2009). Zooming in: Self-Emergence of Movement in New Product Growth, *Marketing Science* (28:1), pp.274-292

Krishnan, R., Smith, M.D.,Tang, Z., Telang, R. (2004). The Impact of Free-Riding on Peer-to-Peer Networks, Proceedings of the 37th HICSS, IEEE

Karagiannis,T., Broido,A., Brownlee,N., Claffy,K.C. and Faloutsos,M. (2004). File-sharing in the Internet: A characterization of P2P traffic in the backbone. Technical report.

Mahajan, V.E., Muller, F.M.,Bass, J. Wind, (2000). New Product Diffusion Models, Sage, Thousand Oaks, CA

Ramsay, J.O. and Silverman, B.W., (2005). *Functional Data Analysis* (2nd ed.), Springer-Verlag, New York

The International Federation of the Phonographic Industry, (2009). Digital Music Market Report.