creative commons

C O M M O N S D E E D

저작자표시-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

**저작자표시.** 귀하는 원저작자를 표시하여야 합니다.

**변경금지.** 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

**저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.**

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

**Master Thesis in Science in Business Administration**

# The Effect of Network Position on the Performance of Open Innovation in Service Systems

## - Econometric Analysis Based on Social Network Theory -

연결망의 위치가 소프트웨어 서비스의 개방형 혁신에 미치는 영향
: 사회연결망 이론에 기반한 계량경제분석

**August 2014**


**Graduate School of Seoul National University**

**Technology Management, Economics, and Policy Program**

**Jiwon Bang**

# The Effect of Network Position on the Performance of Open Innovation in Service Systems

**- Econometric Analysis Based on Social Network Theory -**

지도교수 **Jörn Altmann**

이 논문을 경영학석사학위 논문으로 제출함

**2014** 년 **8** 월

서울대학교 대학원

협동과정 기술경영경제정책 전공

방지원

방지원의 경영학석사학위 논문을 인준함

**2014** 년 **8** 월

위 원 장 **Junseok Hwang** (인)

부위원장 **Jörn Altmann** (인)

위 원 **Kibae Kim** (인)

# Abstract

# The Effect of Network Position on the Performance of Open Innovation in Service Systems

Service systems today provide a platform for open innovation enabling the sharing of service resources with any innovation agents to create new services. While the prior research on service systems for open innovation pioneered the network structure of innovation collaboration, it has also raised the important issue of whether the network position of the service system enhances creativity with static data. In this thesis, it is suggested that the network position of a service network at individual service affects innovation performance, and a time lag exists between those relations. One research question is if the network position affects innovation performance in a service system, what position is good for innovation performance? Furthermore, this thesis explores whether a time lag exists when a network position affects the innovation performance of service systems. To do so, a service system for open innovation through an open Applications Programming Interface (API) is investigated from a network perspective, in which a node and a link represent a software service and the co-development of two software services for a composite service, respectively. And then several software services are selected with high performance and a central position. Cross sectional regression is conducted to test the hypothesis that the network position of services, measured with degree and betweenness centralities, affects their performance, measured by the number of visitors. The data on the network position were gathered from www.programmableweb.com between October 2010 and April 2014, and the performance data from www.alexa.com between November 2013 and May 2014. The results show that network position affects the innovation performance and there exists a time lag in the relation between network position and innovation performance. The results suggest to make the strategy of network position from academic and pragmatic views. Furthermore, findings expected to encourage innovation studies from a network perspective to consider the time lag in the relation between network position and innovation.

**Keywords: Open Innovation, Service System, Software-as-a-Service, Software Services, Composite Services, Network Position**

**Student Number: 2011-23442**

# Contents

# List of Tables

# List of Figures

# Chapter 1. Introduction

## 1.1 Motivation

The market demand on new business models attracting users to innovation and the advance of related IT technologies motivates service systems to provide a platform where to share the knowledge for innovation freely. A service system is defined as a dynamic value-cocreation configuration of resources, including people, organizations, shared information (language, laws, measures, methods), and technology, all connected internally and externally to other service systems by value propositions (Maglio et al., 2009). Service systems viewed as a system of systems (SoS) are incorporated by several businesses for developing and co-creating applications, comprised of tangibles and intangibles (activities), to fulfill specific customer requirements (Lopes and Pineda, 2013).

Some products, and several ideas for the products, new business model emerge by reusing the ideas both internal and external to a company through open innovation. Innovation in this way is called “open innovation” (Chesbrough, 2005). In this sense, service systems are the software industry version of platforms for “open innovation”. There are a variety of examples of service systems for open innovation: e.g., the database of academic journals where researchers submit and download academic articles (Wagner and Leydesdorff, 2005), the open source development community where developers exchange the information on the software development projects (Valverde and Soleê, 2007), and Software-as-a-Service (SaaS) with open APIs for allowing the access of service users to its intrinsic technology (Kim et al., 2011). In case of APIs, mashups, an empowering technology, doesn't limit users to reuse an existing APIs (Michael, 2009).

## 1.2 Problem Description

As the open innovation in service systems promotes reusing of shared resources, the innovation studies is interested in the structure of an innovation system and its evolutionary pattern. One of the main efforts of innovation studies is to investigate the structure of large service systems and their evolutionary patterns (Albert et al., 1999; Newman, 2001; Wagner and Leydesdorff, 2005). Those studies tried to apply network analysis to the cooperation of agents and resources in service systems in the same line with the network science including social networks (Freeman, 1979) and statistical physics (Albert et al., 1999). Prior research on the structure of service systems revealed the structure of a service system (Kim et al., 2009; Kim et al., 2011; Kim and Altmann, 2013) from those perspectives.

On the basis of the understandings of network structure and position, some of previous studies statistically tested whether the network position of software services in a service system affects their innovation performance in a service system (Grewal et al., 2006; Sasidhgaran et al., 2011). However, the prior research on service systems performed a simple cross sectional analysis of the relation between network position and innovation performance with a hidden assumption that the network is static, so that it messes the fact that the network position of nodes varies as time goes on (Kim et al., 2013).

## 1.3 Research Goal

The main objectives of this thesis are to analyze the effect of the network position of a software service on the innovation performance of the network, and to examine whether there exists time lag when the position of nodes in a service network affects their innovation performance.

## 1.4 Research Question

Research questions are: Does the position of a SaaS in a service network affect the innovation performance of the SaaS? If the network position is considerable, what position is related with the high innovation performance? Is there time lag exist when the network position affects the innovation performance of service systems?

## 1.5 Research Outline

To answer these research questions, it is considered that the service network formed on the basis of the empirical data surveyed from Programmableweb (www.programmableweb.com). And then cross sectional regression models are conducted to test the hypotheses on the relationship between the network position of each software service, and its innovation performance, and on the existence of time lag effect in the relation above.

In the regression models innovation performance is measured with the reach, or the number of users visiting the site a day, gathered from Alexa (www.alexa.com). Furthermore, the degree centrality and betweenness centrality of each software service are measured to identify the network position (Freeman, 1979).

The analysis results show that the network position affects the innovation performance of software services. The statistical test suggests that the innovation performance of a software service improves according to the increase in the degree centrality of the service. However, the effect of the betweenness centrality on innovation performance is statistically insignificant. The joint effect of degree centrality and betweenness centrality seems to exist, but it is statistically insignificant. The results propose that structural embeddedness in a service network gives the best innovation performance compare to the other positions.

The results of the time lag effect show that time lag exists when the network position affects the innovation performance of service systems. By separating models at six-month intervals, it is better when the network position before about one year positively affects performance. The effect of the betweenness centrality seems to be weaker than degree centrality for significant probability. In conclusion, the structural embeddedness before one year in a service network would affect mostly the performance of service systems.

The remainder of this paper is organized as follows. The next chapter introduces the theoretical background of SaaS platform as a service network and innovation in the networks. Chapter 3 describes how to form the SaaS network with the empirical data and measure the centralities, and suggest a research model which is tested. Chapter 4 shows the results of statistical test on the effect of network position on the innovation performance in a service network. And also shows the results of a time lag effect when the network position affect the performance in a service network. Finally, chapter 5 concludes with a discussion on the academic and entrepreneurial implications of this research and its limitations.

# Chapter 2. Literature Review

## 2.1 Open Innovation of Software Services

One of the recent trends in software industry is that software is provided as a service instead of a product. The software delivered by this model is called "Software-as-a-Service" (SaaS) (Campbell-Kelly, 2009). In this model customers utilize software running in the programs and computer infrastructure of its provider through the Internet. For example, Salesforce.com and Google offer through the Internet customer relationship management (CRM) services, and office services such as email, calendar and docs, respectively.

As software is provided as a service, the pattern of software innovation has changed, too. While the software was developed in a company to be provisioned customers as a product, the customers also participate in the software innovation on the basis of the software services. In the new paradigm of innovation, that is, a customer can combine existing software services and add his/her own special functions to create a "composite service" (Haines and Rothenberger, 2010), or a "mashup" (Ogrinz, 2009). For example, a mashup Weather Bonk, the weather forecasting service on a map, was developed with the functions of NOAA Weather Service, Weather Bug Service and Google Maps, etc. (Kim et al., 2013).

This new style of innovation is called "open innovation", suggesting that a leading company provides a platform of innovation to invite third party developers in its innovation process (Chesbrough, 2003). That is, the third party developers can utilize the innovation resources so that the cost of innovation is reduced owing to the free access to the pre-developed resources. When the innovation by third party developers is vigorous, the platform attracts more customers who give benefit both to the platform provider and third party developers. This is the mechanism running the ecosystem of open innovation in service systems (Baek et al., 2014; Kim and Altmann, 2013).

The key to this process is that a leading company does not earn benefit directly through selling services but indirectly through its innovation partners. Therefore, a leading company should promote third party developers to achieve more innovation, while the strategy in old style of innovation was making itself more attractive through innovation. This new strategy for innovation ecosystem is named as "platform leadership" (Gawer and Cusumano, 2002). And this is the reason that the structure of the society and the pattern of third party developers' behavior has been investigated by a variety of innovation studies on service systems (Hwang et al., 2009; Kim et al., 2011; Kim and Altmann, 2013; Kim et al., 2013).

## 2.2 Network Position and Innovation Performance

Social network theory is a part of social science that has long been providing implications on the life of people in a society (Festinger, 1963; Milgram, 1967; Barabási, 1999; Girvan and Newman, 2002). People prefer making a small group with friends who meet by chance near their home, and they share information, benefit and norm in the small groups (Festinger, 1963). And a whole society is also really so small that a man in Texas can reach somebody in Massachusetts through the connection of few friends (Milgram, 1967). Recently, statistical physicists found the actual structure of a society with the help of high computation power that can analyze a big size of data. For example, Barabási (1999) found that several hubs connect almost all people in a society who have few connections.

Two of the most popular categories of position in an entire network are structural embeddedness and junctional embeddedness (Grewal et al., 2006). Structural embeddedness captures the extent to which an entity is entrenched in a network thorough direct connections. On the other hand, junctional embeddeness assess the extent to which an entity connects the other. Junctional embeddedness implies better position in a network is mediating the other agents to gain benefit through their connection (e.g. to catch information between them), while structural embeddedness implies securing resources through a node's neighbors (Grewal et al., 2006).

Considering clusters within which connectivity is dense comparing to those out of it, the network position of the previous studies is categorized into four types: hub, core, bridge and periphery (Baek et al., 2014). "Hub" is defined as the nodes connected with a majority of nodes in a whole network (Barabási, 1999; Albert and Barabási, 2002). "Bridge" is the node connecting other nodes within different clusters though it is not connected with a lot of neighbors (Burt, 1992). "Core" is defined as the node connected with other nodes within a cluster but does not connect pairs of nodes within different clusters, and classify the nodes not including any of these into “periphery”.

This classification is considerable in innovation studies whose interest is whether the network position affects its innovation performance, and what position gives benefit if it does so (Everard and Henry, 2002; Granovetter, 1973; Grewal et al., 2006; Hansen, 1999; Hargadon, 2002; Sasidharan et al., 2011; Tsai, 2001). It is well known that embeddedness affect the innovation performance (Grewal et al., 2006). It is also well known that hub position raises power and influence through its high connectivity (Scott, 1991). Moreover, bridge is beneficial for innovation because it has chance to combine information separated in different clusters (Burt, 1992; Hargadon, 2002). However, contrary results also come in some conditions (Grewal et al., 2006; Hansen, 1999; Tsai, 2001). For example, core is beneficial for innovation with complicate knowledge while the opposite works with simple information (Hansen, 1999).

Figure 1 shows examples of network position in a simple network with twenty five nodes and thirty five links. There are four groups of nodes within which they are densely connected, or four clusters (A, B, C, and D). According to the four categories of network position, node 1 locates at a hub position which connecting the whole network. Node 2 is a bridge that connects cluster A and cluster B through only two links. Node 3 is a core that is densely connected with the nodes in cluster A, but does not connected well out of the cluster. Node 4 has only a directly connected neighbor, or locates at periphery.

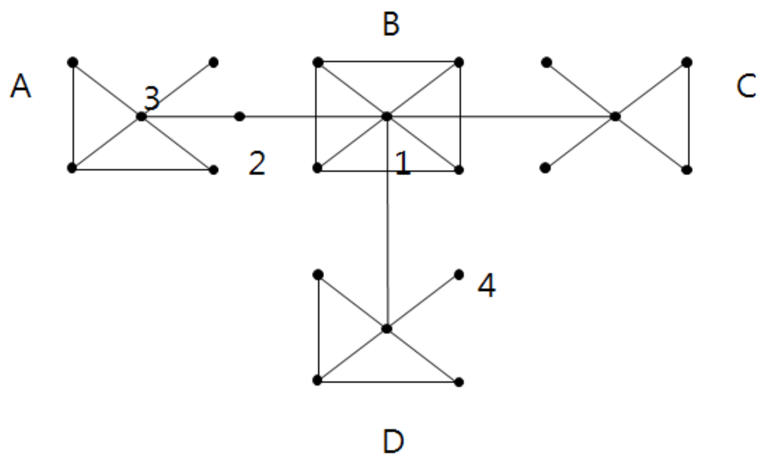



**Figure 1.** Example of network position

According to the previous studies on network and innovation, an interesting question is whether the network position affects the innovation performance. In detail, this question is investigated with the following five sub-hypotheses on the basis of the classification of network position defined above. With these sub-hypotheses, it is inferred what network position is beneficial for innovation.

H1. The position of a software service in a service network affects the performance of the software service.

H1.1. The structural embeddedness gives better performance.

H1.2. The junctional embeddedness gives better performance.

H1.3. The hub position gives better performance.

H1.4. The bridge position gives better performance.

H1.5. The core position gives better performance.

If the network position affects the innovation performance, following question is whether a time lag exists among the relation of network position and innovation performance of service systems.

H2. There exists a time lag when the network position affects the innovation performance.

H2.1. An improvement of performance can be noticed with a time lag for structural embeddedness.

H2.2. An improvement of performance can be noticed with a time lag for junctional embeddedness.

H2.3. An improvement of performance can be noticed with a time lag for hub positions.

H2.4. An improvement of performance can be noticed with a time lag for bridge positions.

H2.5. An improvement of performance can be noticed with a time lag for core positions.

# Chapter 3. Methodology

## 3.1 Data

Two data sets are used, one for measure the innovation performance of the services, and the other for identifying the network position of software services.

The innovation performance of the services was collected from Alexa (www.alexa.com), which provides commercial web traffic data. Alexa estimates the traffic of each domain according to the global traffic panel, i.e., sampled from millions of Internet users over 25,000 different browser extensions. It shows that the ranking of a site according to reach, pageviews and bounce rate. Reach is the number of users who visit a site on a day. Pageviews of a site is the number that users clicked on the page on a day. And bounce rate of a site represents the rate of the number of visitors who leave the site as soon as enter it to the number of visitors who remain the site. These indicators are not measured only for domains (e.g. www.google.com), but also its subdomains (e.g. www.google.com/maps). Alexa provides other demographic information as well.

The network data was collected from Programmableweb (www.programm-ableweb.com), which is an open platform publishing the information of software services and composite services. The gathered data includes the name of mashup developed between 1 October 2010 and 27 April 2014, and Application Programming Interfaces (APIs) used to develop the mashup. With this data, a service network is defined as a set of nodes representing software services and links between these nodes. A link between a pair of software services indicates that the two software services are used concurrently in a mashup. When more than two software services are used to develop a mashup, a complete graph is generated by definition. For example, a mashup developed with five software services generates ten links among the software services. We distinguish the links between a pair of nodes representing the development of different mashups. That is, a service network is weighted, and the value of a link between two software services means the number of mashups developed with the software services.

## 3.2 Research Method

To test the hypotheses that the network position affects the innovation performance, cross sectional regression is implemented. In the regression model the dependent variable is innovation performance, and the independent variables are centralities identifying how central a node is in a network. In this first hypotheses, only current variables are used.

To test the hypotheses that a time lag exists when the network position affects the innovation performance, regression model is implemented as first hypotheses. But there is difference in variables between two hypotheses. In the second hypotheses, independent variables are set for seven models with time variables. Each model has six independent variables, and those variables start from current time of variables to past time of variables at one month intervals. So each model can represent the effect of network position per six months. One model would get best explanation among those models, and the period of six months when the model is indicating would be time effect of network position on performance of service systems. By doing this, investigate whether a time lag exists in relation of network position and innovation performance of service systems.

### 3.2.1 Dependent Variable

Prior research proposed a variety of ways to measure innovation performance (Everard and Henry, 2002; Grewal et al., 2006; Sasidharan et al., 2011). For example, there are financial performance, degree of affiliation, technical or commercial success on work, task performance either individuals or team, and so on.

In this paper, innovation performance is measured through reach of a site which is expressed as the ratio of the number of Internet users who visit the site to the number of the sample of Internet users surfing on the Internet a day. For example, if a site alexarankingboost.com has 8% of reach, this means that if there is taken random samples of one million Internet users, and therefore 80,000 users visit that site. Reach is set as a dependent variable in analysis because it describes the actual performance of software services; the number of people using the site means that the site is valuable for the number of users. It is collected that reach data of subdomains of software services appearing in the service network. The reach for a domain is measured daily, and the one for its subdomains per a month, and a software service in this research is correspond to a subdomain. Therefore, the average of reach per domain is used to calculate the reach of a software service.

### 3.2.2 Independent Variable

The independent variables for this thesis are measured with two indicators for network position: degree centrality and betweenness centrality (Freeman, 1979).

Degree centrality is the number of ties or neighbors of a node (Opsahl et al., 2010), so that it measures the number of links that a node has, to indicate how many neighbors it has directly connected with. This measurement assumes that an agent has more benefit, power or influence as it has more neighbors because it can communicate with more agents on the connection. The definition of degree centrality is slightly modified when the links involve some weights. That is, degree centrality can be used to make an estimate of the strength of collaborative ties or neighbors (Newman, 2001), so that it is calculated in a weighted graph in this thesis. Accordingly, the degree centrality in a weighted graph is the sum of weights of each link that a node has, and is also called “strength” (Opsahl et al., 2010). In mathematical terms the degree centrality $C_D^w(i)$ of node $i$ is the sum of weight $w_{ij}$ of all the links to node $j$ in the neighbors set, $N(i)$, of node $i$:

$$C_D^w(i) = \sum_{i \in N(i)} w_{ij} \quad \text{Eq. (1)}$$

Betweenness centrality measures the extent to which a node connects other nodes in the network (Everard, Henry, 2002), so that it can be defined the number of shortest paths among nodes in the network (Opshal, 2010). Betweenness centrality indicates how many neighbors a node mediates while degree centrality shows how many neighbors a node has. With this indicator, it is diagnosed how effective the information a node can catch is in a network because information flows through an optimal path over agents. To define it in mathematical formula, the shortest path should be calculated first. It can be also measured for collaborative neighbors, so that betweenness centrality is calculated in weighted graph in this thesis. The path length between a pair of nodes *i* and *j* is defined as the minimum sum of inversed weights: $d_w(i, j) = \min(1/w_{ih} +\cdots+ 1/w_{kj})$ (Opsahl et al., 2010). With the definition of the shortest path length, it can be counted that the number $\sigma_{ij}{}^{w}(k)$ of shortest paths between the two nodes passing through node *k*, and the number $\sigma_{ij}{}^{w}$ of shortest paths between the pair of nodes *i* and *j*. Then betweenness centrality is calculated as:

$$C_B(i) = \sum_{i,j \neq i} \frac{\sigma_{ij}^{w}(v)}{\sigma_{ii}^{w}} \quad \text{Eq. (2)}$$

With these two centralities the network position is classified into hub, bridge, core and periphery (Baek et al., 2014). By definition, the degree centrality and betweenness centrality are both high for a node locating at hub. When the node is at bridge, its betweenness centrality is high and its degree centrality low comparing to another nodes. On the other hand the node at core has high degree centrality but low betweenness centrality. The node with both low degree centrality and betweenness centrality means that it is at periphery.

# Chapter 4.  Analysis Results

## 4.1 Descriptive Analysis

There were more than 2000 software services listed in Programmableweb (www.programmableweb.com), among which around 1000 software services were used to develop composite services. 179 software services were chosen, which were most frequently used for composite service development according to the data gathered from Programmableweb. And reach of the chosen software services from Alexa (www.alexa.com) was gathered. Among them the traffic data for 69 software services were not found, therefore 110 software services were available for this analysis.

Table 1 shows the descriptive statistics of reach, degree centrality and betweenness centrality, the variables used in this analysis. Though the table does not show, it is important that the innovation performance, the degree centrality and the betweenness centrality are all highly skewed (Appendix 1). That is, most of the cases have low value of reach, degree centrality and betweenness centrality, and the values for only few cases are very high. This is normal in network analysis. For example, the degree distribution decays by a power function in a lot of large complex networks (Albert et al., 1999; Albert and Barabási; 2000).

**Table 1.** Descriptive Statistics

| Variable | Reach | Degree Centrality | Betweenness Centrality |
|---|---|---|---|
| Mean<br>(SD) | 1.038<br>(4.681) | 0.154<br>(0.271) | 0.004<br>(0.0185) |

This highly skewed distribution of variables makes the statistical analysis hard because the analysis results are likely to depend on few outliers in any pattern in the low value area. In order to avoid this problem, the variables are transformed with a logarithm function. Figure 2 shows that the pairs of reach and degree centrality (left), and reach and betweenness centrality (right) have no outliers in log-log scales.

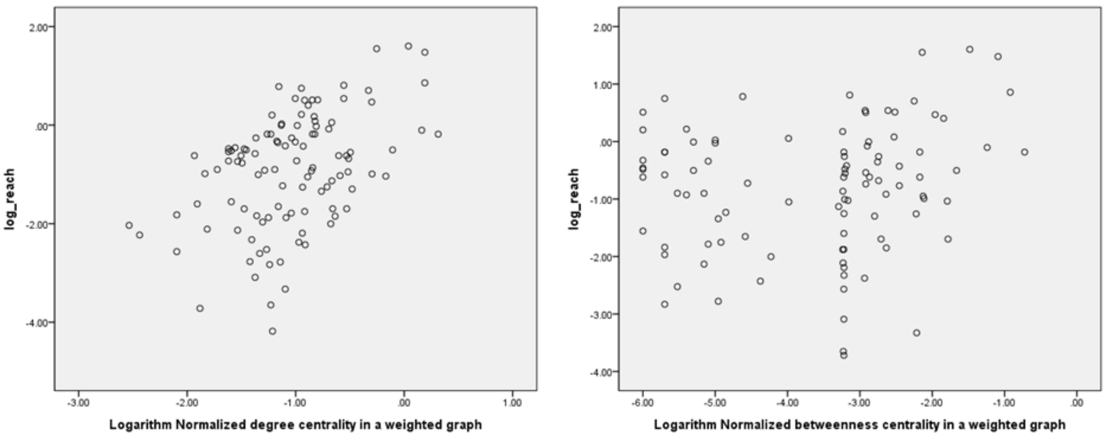



**Figure 2.** Scattograms of reach vs. degree centrality (left), and betweenness centrality (right) in log-log scales

## 4.2 Cross Sectional Regression for H1

In order to show the effect of network position on innovation performance, a cross sectional regression model is tested with backward method. In the model the dependent variable is the logarithm of reach $R(it)$, and independent variables are the logarithms of degree centrality $C_D(it)$ and betweenness centrality $C_B(it)$. For the first set of hypothesis, only current time of both dependent variables and independent variables are used. Those variables are indicated by $t = 0$. In mathematical form, for $Y_{it} = \log R(it)$, $X_{1it} = \log C_D(it)$ and $X_{2it} = \log C_B(it)$, model for this thesis is represented as:

$$Y_{it} = \alpha_{it} X_{1it} + \beta_{it} X_{2it} + \varepsilon_i \quad \text{Eq. (3)}$$

Table 2 shows the result of the regression. According to the backward method, two models were tested. In the first model, both the logarithm of degree centrality and the logarithm of between centrality were used. In the second model, the logarithm of betweenness centrality was removed because it is less explanatory. The results suggest that the logarithm of degree centrality positively affects the logarithm of reach with high statistical significance. The p-value of the logarithm of degree centrality is 0.000 while those for the logarithm of betweenness centrality and constant are higher than 10%. The coefficient of the logarithm of degree centrality is positive. The model 2 tests the effect of only the logarithm of degree centrality on the logarithm of reach. The significance of the variable is very high, the p-value for degree centrality in model 2 is 0.000, too, while constant is insignificant with 10% of significance criterion. The explanation power of model 1 and model 2 are 0.199 and 0.193, respectively. That is, there is little difference between the explanation power of model 1 and model 2. It means that only the degree centrality affects the innovation performance but the combination of degree centrality and betweenness centrality does not.

According to the framework discussed in Section 3.2.2, H1.1 and H1.2 are supported if the innovation performance correlates with degree centrality and betweenness centrality, respectively. Furthermore, H1.3 is supported if the performance correlates positively with both of betweenness centrality and degree centrality. H1.4 is supported when the performance correlates positively with betweenness centrality and negatively with degree centrality, and H1.5 for its negative correlation with degree centrality and positive correlation with betweenness centrality. In this sense, the results seems that the core position yields better innovation because the innovation performance is dependent positively on degree centrality and negatively on betweenness centrality. However, this is not persuasive because the betweenness centrality is statistically insignificant in model 1 and the explanation power in model 2 is not so different from the one in model 1. Therefore, it is concluded that the structural embeddedness in service network gives the best innovation performance comparing to the other positions. That is, the analysis results support hypothesis 1.1.

**Table 2.** Cross Sectional Regression Results

| | Model 1 | Model 2 |
|---|---|---|
| Constant | 0.005<br>(0.291) | 0.168<br>(0.229) |
| logarithm of degree centrality | 1.134 ***<br>(0.251) | 1.011 ***<br>(0.211) |
| logarithm of betweenness centrality | -0.078<br>(0.086) | - |

Notes: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.
(se; standard error)

## 4.3 Time Effect Result for H2

In order to show whether a time lag is exist when the network position affects innovation performance of service systems, a cross sectional regression model is modified. The dependent variable in this model is again the logarithm of reach $R(it)$ of software service $i$ at a certain time period $t$ = 0. Independent variables are the logarithms of degree centrality $C_D(it)$ and betweenness centrality $C_B(it)$ of software services $i$ at time period $t$. For $Y_{it} = \log R(it)$, $X_{1it} = \log C_D(it)$ and $X_{2it} = \log C_B(it)$, model for this paper is represented as:

$$Y_{i0} = \sum_{t=0,1,,,36} \alpha_{it} X_{1it} + \sum_{t=0,1,,,36} \beta_{it} X_{2it} + \varepsilon_i \qquad \text{Eq. (4)}$$

In the time variable $t$, a unit period is a month and the whole study period is three years. That is, those variables are indicated by $t$ = 0 for current time, and $t$ = 1 for one month ago, $t$ = 2 for two months ago and to $t$ = 36 for three years ago, there are variables which indicate 0 of $t$ time to 41 of $t$ time. Six independent variables are set to compose one model so that there are total seven models. That is, independent variables from 0 to 5 of $t$ time and dependent variables of 0 of $t$ time formed model 1. Subsequently, independent variables from 6 to 11 of $t$ time with dependent variables are in model 2, 12 to 17 be model 3, and last 36 to 41 of $t$ time for independent variables with dependent variables can be model 7.

By making a comparison among seven models, model 3 got largest value of adjusted R squared. The explanation power of model 3 is 0.197, similar to results of first hypotheses. This model is related to independent variables from 12 of *t* time to 17 of *t* time. In this model, the p-value of degree centrality of 12 time is 0.061. Table 3 shows the results of regression of model 3. In this case, betweenness centrality is not significant relatively. These results seems to be show that degree centrality of twelve to seventeen months before, that is, around one year and a half before affects the performance.

According to the framework discussed as above, also H2.1 and H2.2 are supported if the innovation performance correlates with degree centrality and betweenness centrality, respectively. Furthermore, H2.3 is supported if the performance correlates positively with both of betweenness centrality and degree centrality. H2.4 is supported when the performance correlates positively with betweenness centrality and negatively with degree centrality, and H2.5 for its negative correlation with degree centrality and positive correlation with betweenness centrality. For second hypotheses, one model is selected by best value of R squared as mentioned, so model 3 could be chosen. From Table 3, betweenness centrality is not significant, but degree centrality is significant when $t = 12$. Therefore as first hypothesis, it is concluded that the structural embeddedness in service network gives the best innovation performance comparing to the other positions. That is, the analysis results support hypothesis 2.1. Furthermore, the structural embeddedness before one year in a service network would affect mostly the performance of service network.

**Table 3.** Time Effect Results

| Model 3 | Coefficient | Significant Probablitiy |
|---|---|---|
| constant | -0.015<br>(0.309) | 0.962 |
| logarithm of degree centrality, t=12 | 19.415 ***<br>(10.229) | 0.061 |
| logarithm of degree centrality, t=13 | -0.820<br>(17.183) | 0.962 |
| logarithm of degree centrality, t=17 | -17.457<br>(11.146) | 0.121 |
| logarithm of betweenness centrality, t=12 | -1.838<br>(1.173) | 0.121 |
| logarithm of betweenness centrality, t=13 | 1.894<br>(3.951) | 0.633 |
| logarithm of betweenness centrality, t=14 | -2.723<br>(4.049) | 0.503 |
| logarithm of betweenness centrality, t=15 | -2.253<br>(3.383) | 0.507 |
| logarithm of betweenness centrality, t=16 | 4.425<br>(3.167) | 0.166 |
| logarithm of betweenness centrality, t=17 | 0.361<br>(0.371) | 0.333 |

Notes: * p<0.10, ** p<0.05, *** p<0.01.

(se; standard error)

# Chapter 5. Discussion and Conclusion

Within this thesis, the hypotheses that network position affects innovation performance and whether a time lag exists were tested with a model of cross sectional regression. In the model, the dependent variable is innovation performance, measured with reach, or the number of Internet users visiting a domain per day. And the independent variable is the position in a service network, measured with the logarithms of degree centrality and betweenness centrality. Those independent variables have time variables at one month intervals. The results of first hypotheses that the network position affects innovation performance show that the network position of a software service impacts on the performance of the service. According to results, degree centrality considerably affects dependent variable, that is, innovation performance. On the other hand, betweenness centrality doesn't show effectively as degree centrality. Consequently, the structural embeddedness in service network gives the best innovation performance comparing to the other positions.

The results of second hypotheses suggest that a time lag exists when the network position affects the performance of service systems. According to the results, degree centrality about one year before affects the performance. It can be concluded that the structural embeddedness before one year in a service network gives the best innovation performance comparing to the other positions.

Findings have cast important implications both to academia and industry. Academic implication of findings is that the approach of this thesis provides innovation studies with a guideline how to study the relation between network position and innovation performance. It suggests that time lag should be considered for the study on relation with the network position and the innovation performance. Findings focus on the role of structural embeddedness in service network. The industrial implication is that the strategy of software service providers should be on considering the structural embeddedness rather than the other positions. One of the ways of strategy for the structural embeddedness is providing software service complementary to the target software service which users demand. Furthermore, the structural embeddedness before twelve months, about one year before, affects more positively on the performance.

However, the hypothesis test for H2 has a serious multi-collinearity problem. The centrality appearing previous time period could affect the centrality appearing later. Thus, there was multi-collinearity problem among the variables because each variable means the time period the centrality is measured. Because of the multi-collinearity problem the statistical test does not determine the statistical significance of a certain variable. In order to resolve this shortcoming, the further study should redesign the experiment in which each model includes only one degree centrality and betweenness centrality at each time period, and the fitness of the models are compared to find the best fittest model.

Moreover, this thesis identified the network position only with degree centrality and betweenness centrality. But more indicators such as clustering coefficient (Newman, 2001) and node redundancy (Burt, 1992) should be measured in order to define the position of bridge and core.

According to this thesis, further research would be conducted to test which network would have better effect on performance, a cumulative network or a temporal network.

# Bibliography

Albert, R., and Barabási, A.-L. (2000), “Statistical mechanics of complex networks,” Rev. of Modern Physics, Vol. 74, pp. 47-97.

Albert, R., Jeong, H, and Barabási, A.-L. (1999), “Diameter of the World-Wide Web,” Nature, Vol. 401, pp. 130-131.

Armbrust, M., Fox, A., Griffith, R., et al. (2010), “A view of cloud computing,” Comm. of the ACM, Vol. 53, pp. 50-58.

Baek, S., Kim, K., and Altmann, J. (2014), “Role of platform providers in service networks: The case of Salesforce.com AppExchange,” The IEEE Conference of Business Informatics, Geneva, Switzerland, 14-17 July.

Bonacich, P. (1987), “Power and centrality: A family of measures,” Am. J. of Sociology, Vol. 92, pp. 1170-1182.

Burt, R.S. (1992), “Structural Holes: The social structure on competition,” Harvard Univ. Pr., MA.

Campbell-Kelly, M. (2009), “Historical reflections: The rise, fall, and resurrection of software as a service,” Comm. of the ACM, Vol. 52, pp. 28-30.

Chesbrough, H.W. (2003), “Open Innovation: The new imperative for creating and profiting from technology,” Harvard Bus. Sch. Pr., MA.

Chesbrough, H.W., Vanhaverbeke, W., West, Joel. (2005), “Open Innovation: A new paradigm for understanding industrial innovation,” Oxford University Press.

Cusumano, M. (2010), “Cloud computing and SaaS as new computing platforms,” Comm. of the ACM, Vol. 53, pp. 27-29.

Everard, A., and Henry, R. (2002). “A social network analysis of interlocked directorates in electronic commerce firms,” Electronic Commerce Res. and App., Vol. 1, pp. 225-234.

Festinger, L., Schachter, S., and Back, K. (1963), “Social pressures in informal groups: A study of human factors in housing.” Stanford Univ. Pr.

Freeman, L.C. (1979), “Centrality in social networks: Conceptual clarification,” Social Networks, Vol. 1, pp. 215-239.

Freeman, C. (1991), “Networks of innovators: A synthesis of research issues,” Research Policy, Vol. 20, pp. 499-514.

Gawer, A., and Cusumano, M.A. (2002), “Platform Leadership: How Intel, Microsoft, and Cisco Drive Industry Innovation,” Harvard Bus. Sch. Pr., MA.

Girvan, M., and Newman, M.E.J. (2002), “Community structures in social and biological networks,” PNAS, Vol. 99, pp. 7821-7826.

Granovetter, M.S. (1973), “The strength of weak ties,” Am. J. of Sociology, Vol. 78, pp. 1360-1380.

Grewal, R., Lilien, G.L., and Mallapragada, G. (2006), “Location, location, location: How network embeddedness affects project success in open source systems,” Management Science, Vol. 52, pp. 1043-1056.

Haines, M.N., and Rothenberger, M.A. (2010), “How a service-oriented architecture may change the software development process," Comm. of the ACM, Vol. 53, pp. 135-140.

Hansen, M.T. (1999), "The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits," Administrative Science Q., Vol. 44, pp. 82-111.

Hargadon, A.B. (2002), "Brokering knowledge: Linking, learning and innovation," Research in Organizational Behavior, Vol. 24, pp. 41-85.

Hwang, J., Altmann, J., Kim, K. (2009), "The Structural Evolution of the Web 2.0 Service Network," TEMEP Discussion Paper, No. 2009:14

Kim, K., and Altmann, J. (2013), "Evolution of the Software-as-a-Service innovation system through collective intelligence," International J. of Cooperative Information Systems, Vol. 22, p. 1340006.

Kim, K., Hwang, J., and Altmann, J. (2011), "Does Web2.0 foster innovation? An analysis of the openness of the Web2.0 service network," The 44th HICSS, Kauwai, HI, 4-7 January.

Krackhardt, D., and Stern, R.N. (1988), "Informal networks and organizational crises: An experimental simulation," Social Psychology Q., Vol. 51, pp. 123-140.

Lewis, T.G. (2009), "Network Science: Theory and Applications," Wiley.

Lopes, A.J., Pineda, R. (2013), "Service systems engineering applications," CSER, Atlanta, GA, 19-22 March.

Maglio, P.P., Vargo, S.L., Caswell, N., Spohrer, J. (2009), "The service system is the basic abstraction of service science," The 41th HICSS, Waikoloa, HI, 7-10 January.

Michael, O. (2009), "Mashup Patterns: Designs and examples for the modern enterprise," Pearson Education, Books.

Milgram, S. (1967), "The small-world problem," Psychology Today, Vo.l. 1, pp. 61-67.

Newman, M.E.J. (2001), "Clustering and preferential attachment in growing networks," Physical Review E., Vol. 64, p. 025102.

Newman, M.E.T. (2001), "Scientific collaboration networks. Ⅱ. Shortest paths, weighted networks, and centrality," Physical Review E., Vol. 64, p. 016132.

Ogrinz, M. (2009), "Mashup Patterns: Designs and examples for the modern enterprise," Addison-Wesley.

Opsahl, T., Agneessens, F., and Skvoretz, J. (2010), "Node centrality in weighted networks: Generalizing degree and shortest paths," Social Networks, Vol. 32, pp. 245-251.

Papazoglou, M.P., and Georgakopoulos, D. (2003), "Introduction: Service-Oriented Computing," Comm. of the ACM, Vol. 46, pp. 24-28.

Sasidharan, S., Santhanam, R., Brass, D.J., et al. (2011), "The effects of social network structure on enterprise systems success: A longitudinal multilevel analysis," Information Systems Research, Vol. 23, pp. 658-678.

Scott, J. (1991), "Social Network Analysis: A Handbook," Sage Publication, UK.

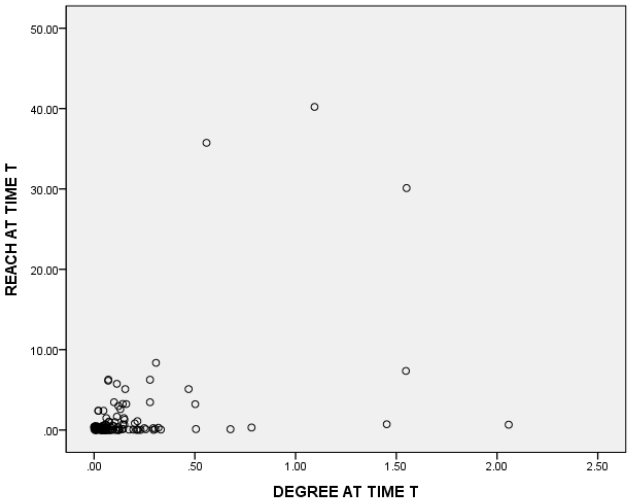
Tsai, W. (2001), "Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance," Academy of Management Journal, Vol. 44, pp. 996-1004.

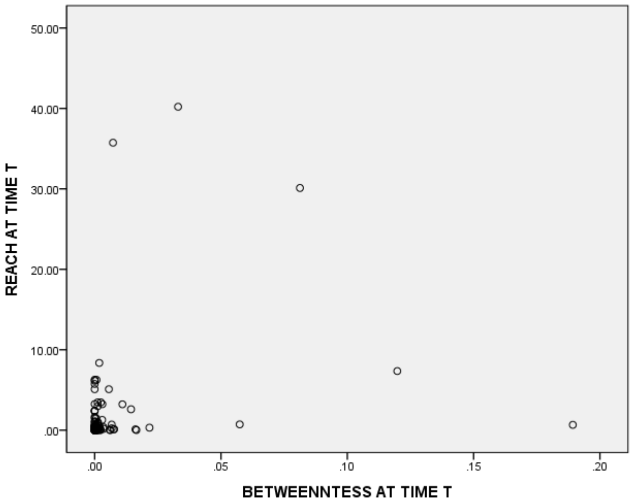
Valverde, S., and Sole, R.V. (2007), "Self-organization versus hierarchy in open source," Physical Review E, Vol. 76, p. 046118.

Wagner, C.S., and Leydesdorff, L. (2005), "Network structure, self-organization, and the growth of international collaboration in science," Research Policy, Vol. 34, pp. 1608-1618.

# Appendix 1: Scattograms

Scattograms of reach vs. degree centrality (top) and reach vs. betweenness centrality (bottom) in linear-linear scales



# Abstract (Korean)

최근 서비스 시스템은 새로운 서비스를 창조하려는 이들에게 서비스 자원을 공유함으로써 자유로운 혁신을 가능하게 하는 개방형 혁신 플랫폼을 제공한다. 개방형 혁신을 추구하는 서비스 시스템에 관한 선행연구가 연결망 관점에서 혁신 행위자들 사이의 협력관계의 특성을 밝혀내는 데 기여한 것은 사실이지만, 서비스 시스템의 연결망 위치가 혁신성에 기여를 하는데 정적으로만 증진시키는 것에만 관심을 두었다는 것이다. 본 논문은 개별 서비스 연결망의 위치가 혁신 성과에 영향을 미치고, 그 관계에는 시간차가 존재한다는 것을 밝히려 한다. 이에 따라서 먼저 연결망 위치가 각 서비스의 혁신성과에 미치는 미칠 것인지를 보고, 어느 위치가 혁신 성과에 긍정적 영향을 미치는지를 이해하고자 한다. 더 나아가, 연결망 위치가 혁신성과에 영향을 미칠 때 시간차가 존재하는 지를 보고자 한다. 이를 위하여, 개방 API를 통해서 개방형 혁신을 위해 사용된 서비스 시스템들을 조사했다. 각 서비스 시스템에서의 연결점과 연결선은 각각의 소프트웨어 서비스와 두 소프트웨어 서비스들이 합성 서비스를 위해 협력함을 나타낸다. 높은 성과를 나타내고 중앙적 위치의 특징을 가진 여러 소프트웨어 서비스들이 선택되었다. 서비스의 연결망 위치가 성과에 영향을 미친다는 가설을 검증하기 위해 다변량 선형 회귀를 수행했다. 본 연구에서는, www.programmable.com에서 서비스 연결망 자료를 구하고, www.alexa.com에서 혁신성과 자료를 구하였다. 분석 결과는, 연결망 위치가 혁

신 성과에 영향을 미친다는 것을 확인할 수 있었고, 그 영향이 발현되는 데는 연결성에 따라 시간차가 발생한다는 것을 알았다. 이 연구결과는, 학문적 및 실무적 관점에서 연결망에서 위치 선정 전략을 구상하는 데 유용하게 쓰일 것으로 기대된다.

**주요어** : 개방성 혁신, 서비스 시스템, Software-as-a-Service, 소프트웨어 서비스, 복합 서비스, 네트워크 포지션.

**학 번** : 2011-23442.