Exploring Potential R&D Collaboration Partners through Patent Analysis based on Bibliographic Coupling and Latent Semantic Analysis

I. Park, Y. Jeong , B. Yoon, L. Mortara

The aim of the present research is to provide a new systematic methodology to explore potential R&D collaboration partners using patent information. The potential R&D collaboration partners are visualized as a patent assignee level-map based on technological similarity between patents by using network analysis. The proposed framework utilises two analytic methods to measure technological similarity. The first method, bibliographic coupling analysis, measures technological similarity based on the citation relationship using patent bibliographic information. Second, latent semantic analysis is utilized based on semantic similarity using patent textual information. The fuel cell membrane electrode assembly (MEA) technology field is selected and applied to illustrate the proposed methodology. The proposed approach allows firms, universities, research institutes, governments to identify potential R&D collaborators as a systematic decision-making support tool.

Keywords: collaborative R&D; R&D collaboration; R&D partner selection; bibliographic coupling; latent semantic analysis

Reference: PARK, I., JEONG, Y., YOON, B. & MORTARA, L. 2014.
Exploring potential R&D collaboration partners through patent analysis based on bibliographic coupling and latent semantic analysis.
Technology Analysis & Strategic Management, 1-23.

1. Introduction

As markets and technology are changing rapidly, the traditional practice of solely relying on in-house Research and Development (R&D) face both technological and business challenges. Open innovation encourages companies to undertake some R&D activities in partnership with others, in that it is beneficial for obtaining new, complementary knowledge, because no firm can self-supply all the knowledge necessary for its R&D activities (Chesbrough 2003). Thus, R&D collaboration is becoming increasingly used in large international enterprises, and has been considered a useful means of technology acquisition (Belderbos et al. 2004; Benfratello and Sembenelli 2002; Butcher and Jeffrey 2005; Das and Teng 2000; Nakamura 2003; Niedergassel and Leker 2011; Pisano 1990; Tyler and Steensma 1995; Finne 2003).

Research on R&D collaboration has explored the factors affecting R&D collaboration performance (Bailey et al. 1998; Belderbos et al. 2004; Bruce et al. 1995; Fritsch and Lukas 2001; Hakanson and Lorange 1991; Lhuillery and Pfister 2009; Miotti and Sachwald 2003). There has also been discussion on the determinants of success in R&D collaboration (Fritsch and Lukas 2001; Tether 2002). The effective partner selection is recognized as a core factor affecting collaboration performance (Ireland et al. 2002). Many studies have tried to identify factors that should be considered in partner selection (Geringer 1991; Geringer and Hebert 1991; Brouthers et al. 1995; Nielson 2003; Wu et al. 2009; Arranz and Fdez de Arroyabe 2008). However, studies on R&D collaboration partner selection have been rarely conducted although R&D collaboration partner selection is one of the important issues in the research area.

The conventional searching process for potential R&D collaboration partners has been based on expert opinions, human relationships, e-mail requests, or online communities (Jeon et al. 2011). These methods have some demerits, in that the data pool and scope of external sources is geographically limited, and dependent on word of mouth (Lee et al. 2010). Furthermore, the methods are time-consuming and labour-intensive, because these rely on the qualitative judgment of experts.

To overcome these limitations, several quantitative and systematic approaches using patent information to search for R&D partners are recently proposed (Jeon et al. 2011; Wang 2012; Geum et al. 2013; Yoon and Song 2014). A patent database is useful for firms to utilize to develop an R&D strategy since patents are the results of R&D and vast public resources that contain technical and market value. Thus, the present research utilizes both bibliographic and textual patent information to identify potential R&D collaborators.

The present study proposes a systematic framework for R&D collaborator exploration using the patent information. The proposed framework utilises two analytic methods to measure technological similarity. First, the bibliographic coupling analysis (BCA) is utilized to demonstrate the potential collaboration possibilities, taking advantage of immediate applicability, because it is constructed by the citing relationship; while co-citation is constructed by the cited relationship. Second, latent semantic analysis (LSA) is also applied to explore potential R&D collaboration partners. The method has the advantages of overcoming a limitation of the traditional keyword-based approach, like the vocabulary mismatch problem, and to reduce the experts' keyword selections process. The proposed framework has the advantage to exploring potential partners because both two analytic methods are used to find missing relevant documents in an information retrieval field.

The potential R&D collaboration partners are visualized as a patent assignee level-map based on technological similarity between patents by using network analysis. Two types of collaboration maps are provided to represent current collaboration state and potential collaborative possibilities. First, a R&D collaboration state map is to show the current R&D collaboration state before analysts explore the potential R&D collaborators. In the map, patent assignees are linked based on joint patenting relation because a joint patent which is co-assigned can describe the success of R&D collaboration with a certain degree of confidence (Kim and Song 2007). Second, a potential R&D collaboration partner map aims at exploring potential R&D collaborators. In the map, patent assignees are linked based on technological similarities measured by BCA or LSA.

The objective of the present research is to propose a new systematic quantitative method to explore potential R&D collaboration partners, using both patent bibliographic and textual information in an industry. The proposed approach is for the organizations that possess the granted patents, that is, the organizations that have some degree of technological capabilities. The information is visualized as a R&D collaboration state map and potential R&D collaboration partner map. As an exemplary case, the fuel cell membrane electrode assembly (MEA) technology field is selected to demonstrate the proposed approach.

The paper is organized as follows. The theoretical research background is explained in Section 2. The research methodology is proposed in Section 3, with the research concept and framework to explore potential R&D partners. Section 4 describes the results of applying the proposed approach to the exemplary case of the fuel cell MEA technology field. Section 5 provides interpretations and implications. Finally, Section 6 concludes with contributions and limitations, and suggests future research directions.

2. Theoretical Background

2.1. R&D Collaboration Partner Selection

As mentioned above, there are few researches on partner selection for R&D collaboration while many studies on R&D collaboration have been performed. However, many researches on partner selection methods have been conducted without focusing R&D collaboration. Most of the researches have proposed overall partner selection methodologies without distinguishing the specific types of partners (Hajidimitriou et al. 2002; Fischer et al. 2004; Zeng et al. 2006; Wang and Chen 2007; Solesvik and Encheva 2010; Niu et al. 2012; Solesvik and Gulbrandsen 2013; Fujiwara 2014). Otherwise, many researches have specified partners as a supplier, strategic alliance, supply chain partner, manufacturing partner, co-development alliance (Amid et al. 2006; Saen 2007; Jeon et al. 2011; Chen et al. 2008; Holmberg and Cummings 2009; Solesvik and Westhead 2010; Ding and Liang 2005; Chang et al. 2006; Huang et al. 2004; Emden et al. 2006; Feng et al. 2010). Several methodologies that have been applied to partner selection can be classified into four categories, (1) exact algorithms such as the Branch and Bound algorithm; (2) mathematical modelling and programming such as goal programming; (3) fuzzy decision-making and multi-attributive decision-making (MADM) algorithms e.g. analytic network process (ANP), analytical hierarchy process (AHP), fuzzy-AHP approach etc.; (4) heuristic and meta-heuristic algorithms such as genetic algorithms (GA) and ant colony optimizer (ACO) etc. (Niu et al. 2012). Since the methods are largely to derive ranking value of given candidate partners, the previous methods are not appropriate to explore potential partners.

Some researches focus on partner selection methods for R&D collaboration. Chen et al. (2010) established a mechanism for R&D strategic alliance partner selection by combining analytical hierarchy process (AHP) and fuzzy sets theory. However, the rest of studies on proposing R&D collaboration partner selection methodology utilized patent information. Jeon et al. (2011) proposed a systematic approach to searching for potential technology partners to solve a specific technical problem. Wang (2012) provided a framework for exploring potential R&D collaborators with complementarity in products consisting of multidisciplinary technologies using patent classification codes. Geum et al. (2013) presented a literature-based approach based on patent and science publication to identify strategic partners for collaborative R&D by designing indices. Yoon and Song (2014) proposed a systematic approach to exploring potential R&D collaboration partners by combining morphology analysis (MA) and generative topology map (GTM). Wang (2012) and Geum et al. (2013) proposed approaches based on structured bibliographic patent data. Jeon et al. (2011) and Yoon and Song (2014) suggested frameworks based on unstructured patent textual data. However, the present research utilizes both bibliographic and textual patent data, providing a visualization process to identify potential R&D collaboration partners. Additionally, it is suitable to explore potential R&D collaborators because the analytic methods, which are used to find missing relevant documents in information retrieval, are utilized.

2.2. Bibliographic Coupling

Bibliographic coupling (BC) (Kessler 1963), which is a similar concept to co-citation (CC) (Small 1973), is a similarity measure that uses citation analysis to establish a similarity relationship between documents. A coupling unit between two documents is an item of reference used by these two documents. If such an item exists, the two documents are bibliographically coupled. Their bibliographic coupling strength is the number of references they have in common. Moreover, a normalized bibliographic coupling strength is suggested by Glänzel and Czerwon (1995), since the number of references between two documents is different.

Conversely, two documents are said to be co-cited, when they both appear in the reference list of a third document. Co-citation frequency is defined as the frequency with which two documents are cited together. BC focuses on groups of papers that cite a source document; on the contrary, CC focuses on references that appear frequently in pairs. In other words, BC is constructed by the citing relationship, while CC is constructed by the cited relationship. The strength of co-citation can increase over time, as new documents that cite previous documents appear. Thus, although BC can be utilized immediately, CC is subject to provide insufficient information (Bichteler and Eaton 1980).

In general, BC is used to find related records in the citation database of the Web of Science and the World Wide Web (Dean and Henzinger 1999; Henzinger 2001; Atkins 1999). BC and CC are employed to find the relevant literatures that were not found by BC alone (Cleverdon 1967; Harter 1971; Swanson 1971; Small 1973, Braam et al. 1991; Chen et al. 2010; Chen et al. 2011), by exploring the research fronts with Clustering (Small and Griffith 1974; Persson 1994; Morris et al. 2003; Jarneving 2007). However, Yeh et al. (2013) used a BC approach to filter out irrelevant patent citations. Several research works compare the performance between BC and CC (Morris et al. 2003; van den Besselaar and Heimeriks 2006; Boyak and Klavans 2010). A combined approach of BC and CC for document retrieval is used by Bitcheler and Eaton (1980). Ma (2012) conducted research on author bibliographic coupling analysis. Furthermore, comparisons between BC and the text approach to find relevant literature have been conducted (Ahlgren and Jarneving 2008; Yan and Ding 2012).

2.3. Latent Semantic Analysis

Latent semantic analysis (LSA), which in the context of information retrieval is also called latent semantic indexing (LSI), and was first introduced by Dumais et al.

(1988), is a mathematical method for dimensionality reduction, because it transforms the original terms-documents vector space into a new coordinate system of conceptual topics and a lower dimensional space that captures the implicit higher-order structure, in the association of terms with documents (Deerwester et al. 1990). LSA tries to overcome a limitation of the traditional vector space model (VSM) (Salton 1971), which is the so-called vocabulary mismatch problem faced by information retrieval systems (Deerwester et al. 1990; Dumais 1995). Because of the tremendous diversity in the words experts use to describe the same object, lexical matching methods are necessarily incomplete and imprecise (Furnas et al. 1983). LSA assumes there is some underlying “latent” semantic structure in word usage data, which is partially obscured by the variability of word choice (Dumais et al. 1988). Singular value decomposition (SVD), factorization of a real or complex matrix in linear algebra, is used to estimate the latent structure, and to get rid of the obscuring “noise” in LSA. The LSA model scored as well as that of second-language English speakers, as evidenced by scores on the Test of English as a Foreign Language (TOEFL) in Landauer and Dumais’ research (1997).

LSA is used in various fields, such as automated document classification (Foltz and Dumais 1992), text summarization (Gong and Liu 2001), relationship discovery (Bradford 2005), automatic keyword annotation of images (Monay and Gatica-Perez 2003), and information visualization (Landauer et al. 2004). In this paper, LSA is applied as an information visualization method to surpass the limitation of keyword-based VSM.

3. Research Methodology

3.1. Research concept

This research aims to suggest an analytic framework to recommend R&D collaboration partners as a patent assignee level-map, through patent data visualization. Two types of collaboration maps are proposed to represent current R&D collaboration relation and potential R&D collaborative possibilities by using joint patents and technological similarities respectively.

Figure 1 shows the overall concept of the research. In both maps, a node shows assignee, and a node size means the technological capability, which is calculated by summing up the number of patents. Y-axis is the number of R&D collaboration partners. However, links have different meaning in respective maps. In R&D collaboration state map, patent assignees are linked based on joint patenting relation since joint patents, meaning that the successful result of R&D is shared by more than two organizations, are considered as collaboration outcome (Kim and Song 2007). In potential R&D collaboration map, patent assignees are linked based on technological similarities calculated by bibliographic coupling analysis (BCA) or latent semantic analysis (LSA). The width of links means the number of joint patents and the technological similarities between assignees.

In figure 1, solid lined rectangle of first period presents proposed methodology while dotted lined rectangle of second period represents additional process to check the validity of the results. Both R&D collaboration state map and potential R&D collaboration map is generated at the analysis time (first period in figure 1) since it is also important to comprehend current R&D collaboration state to explore potential R&D collaborators. The value of Y-axis in potential R&D collaboration map is same as R&D collaboration state map. In this research, the collected data is split into first and second periods and R&D collaboration state map is generated at the second period to verify the validity of the results. The appropriateness of partner matching for potential

R&D collaboration is explained by comparing a potential R&D collaboration maps at the first period with the R&D collaboration state map at the second period. Furthermore, results and implications from the potential R&D collaboration map by BCA and LSA are suggested, by comparison with each other.

Figure 1. Research Concept

3.2. Research Framework

The overall research framework, which corresponds to the first period in figure 1, consists of several steps like Figure 2. The first step is the data collection and pre-processing. In this step, both patent bibliographic information and documents are collected for bibliographic coupling analysis (BCA) and latent semantic analysis (LSA) in the target technology area. Second, the R&D collaboration state map is generated, using joint patent information. Third, a patent-patent matrix is generated, based on the bibliographic coupling relationship, or semantic similarity between patents. This step is an initial step for the potential R&D collaboration map generation. Fourth, an assignee-assignee matrix is generated, by aggregating the relation value from the patent-patent matrix by the assignee of patents. Fifth, a potential R&D collaboration map is generated, based on the assignee-assignee matrix, by using network analysis. Finally, the potential R&D collaboration partners are matched, from the maps based on BCA or LSA.

Figure 2. Research Framework

3.3. Data Collection and Pre-process

After the target technology field is selected, patent bibliographic information and patent documents are collected from the authorized patent database to search for

R&D collaboration partners. The bibliographic information includes patent registration number, assignee name, year of registration, and citing references, etc.

The collected patent documents follow a pre-processing procedure. The names of assignees are normalized, because some assignees deliberately register the different names (e.g. using abbreviations) to avoid patents being easily searched by others. Textual information from the abstract section in patent documents that presents the core concept of the invention is utilized for analysis; while the full text of a patent is not appropriate to analyze by using text-mining, because it includes lots of noise. Noise, such as punctuation marks, numbers, conjunctions and articles, is eliminated in the abstract. A stemming process which extracts the root of a word is conducted to treat the plural form and passive verb as the same word.

3.4. R&D Collaboration State Map Generation

An R&D collaboration state map is generated, using joint patent information with the predefined format, as in Figure 1. The collected data are split into two cumulative periods. R&D collaboration state map at the first period provides information on current R&D collaboration state. For example, Assignee C is the most actively collaborating actor at the first period because it has three R&D collaboration partners in figure 1. R&D collaboration state map at the second period is to verify the validity of the proposed method. For example, potential collaborators pairs in potential R&D collaboration map by BCA or LSA can be compared to the pairs in R&D collaboration state map at the second period.

3.5. Bibliographic Coupling Analysis

To generate a potential R&D collaboration map through bibliographic coupling analysis (BCA), an assignee-assignee matrix is generated from citation-based similarity. To this end, the patent-patent matrix is generated by the following steps.

(1) Bibliographic Coupling Matrix

A patent-patent matrix is constructed, based on the bibliographic coupling relation, which is calculated using bibliographic coupling strength. The original bibliographic coupling strength is defined as the number of common references. In general, the more references they both cite, the more common the technical background that they are both based on for development is (Kessler 1963). That is to say, the higher the bibliographic coupling strength between two patents, the higher their relevance (Huang et al. 2003). However, a normalized bibliographic coupling strength is needed, since the length of reference lists is different between two patents. In this research, the normalized coupling strength, which is shown by Glänzel and Czerwon (1995), is utilized to construct the patent-patent matrix. The normalized coupling strength (NCS) is defined as:

$$NCS_{ij} = \frac{r_{ij}}{\sqrt{n_i n_j}}, \qquad (1)$$

where, NCS_{ij} is the normalized coupling strength between patent i and j; r_{ij} is the number of references common to both i and j; n_i is the number of references in the reference list of patent i; and n_j is the number of references in the reference list of patent j. Figure 3 demonstrates the concept of bibliographic coupling and co-citation. Documents A and B are co-cited by C; meanwhile, A and B are bibliographic coupled with normalized coupling strength $\frac{2}{\sqrt{3\times 3}}$, since A and B share the references D and F, among the references {D, F, G} and {D, E, F} respectively.

Figure 3. Bibliographic coupling and co-citation

The assignee-assignee matrix is constructed by averaging the NCSs between assignees. The average normalized coupling strength (ANCS) is defined as:

$$\mathrm{ANCS}_{mn} = \frac{\sum NCS_{i_m j_n}}{p_m p_n}, \qquad (2)$$

where, ANCS_{mn} is the average normalized coupling strength between assignee m and n; $NCS_{i_m j_n}$ is the normalized coupling strength value between patents i and j when assignees m and n possess patent i and j respectively; p_m is the number of patents of assignee m, and p_n is the number of patents of assignee n. For example, ANCS_{XY} is calculated as a result of $(\mathrm{NCS}_{13} + \mathrm{NCS}_{14} + \mathrm{NCS}_{23} + \mathrm{NCS}_{24})/\ P_X\ P_Y$ in the case that assignee X has patents 1 and 2, assignee Y has patents 3 and 4.

(2) Bibliographic Coupling-based Map

A BCA-based potential R&D collaboration map is generated from the assignee-assignee matrix, by using network analysis. An appropriate threshold for the assignee-assignee matrix is chosen through sensitivity analysis, with considering the visibility of the map. A node and the size of a node mean an assignee and the technological capability of the assignee. Matched assignees with a link are potential R&D collaboration partners, since citation-based similarity between assignees is higher than a chosen threshold.

3.6. Latent Semantic Analysis

To generate the potential R&D collaboration map through latent semantic analysis (LSA), an assignee-assignee matrix based on semantic technological similarity is generated. To this end, a patent-patent matrix is generated, with the following steps.

(1) Latent Semantic Matrix

First of all, a term-patent document matrix is constructed, from the vast textual information, using a vector space model (Salton 1971). The term-document matrix is constructed based on tf-idf weighting (term frequency-inverse document frequency), which is widely used in information retrieval (Salton and McGill 1983). Tf-idf is defined as:

$$\mathrm{tf} - idf_{t,d} = tf_{t,d} \times \log\left(\frac{N}{df_t}\right), \qquad (3)$$

where, N is the number of patent documents in the corpus; $tf_{t,d}$ is the term frequency, which is the raw frequency of term t in document d; and df_t is the document frequency, which is the total number of documents containing the term t.

Singular value decomposition (SVD), which is a mathematical technique closely related to eigenvector decomposition and factor analysis, is used to estimate the latent structure, and to get rid of the obscuring "noise" in LSA (Dumais et al. 1988). Any rectangular term-document matrix, X, can be decomposed into the product of three other matrices:

$$\mathrm{X} = T_0 S_0 D_0{}', \qquad (4)$$

such that T_0 and D_0 have orthonormal columns, and S_0 is a diagonal. T_0 and D_0 are the matrices of left and right singular vectors, and S_0 is the diagonal matrix of singular values. SVD is unique up to certain row, column and sign permutations, and by convention, the diagonal elements of S_0 are constructed to be all positive, and ordered in decreasing magnitude. If singular values in S_0 are ordered by size, the first k largest may be kept, and the remaining smaller ones set to zero. The product of the resulting matrices is a matrix $\hat{X}$, which is approximately equal to matrix X, and closest in the least squares sense to X. The result is a reduced model:

$$X \approx \hat{X} = TSD', \tag{5}$$

which is the rank-k model with the best possible least squares-fit to X. Figure 4 is a schematic of the singular value decomposition and reduced singular value decomposition of the term-document matrix. In Figure 4, T_0, D_0, T, and D have orthogonal, unit-length columns, S_0 is the diagonal matrix of singular values, t is the number of rows of X, d is the number of columns of X, m (≤ min(t, d)) is the rank of X, and k is the chosen number of dimensions in the reduced model (k≤m). The choice of k is a critical problem in research. In practice, a value of k that yields good retrieval performance is used (Deerwester et al. 1990).

Figure 4. Schematic of the SVD, and reduced SVD of the matrix (Deerwester et al. 1990)

Second, the patent document-document matrix is structured by calculating cosine similarity from the latent semantic space through dimension reduction. The cosine similarity (Salton and McGill 1983) between patent documents is defined as:

$$\mathrm{sim}(D_1, D_2) = \frac{\vec{V}(D_1) \cdot \vec{V}(D_2)}{|\vec{V}(D_1)||\vec{V}(D_2)|}, \tag{6}$$

where, $\vec{V}(D_1)$ or $\vec{V}(D_2)$ is a vector of document 1 and document 2, respectively, and

$|\vec{V}(D_1)|$ or $|\vec{V}(D_2)|$ is the length of those vectors.

Figure 5 shows the LSA procedure with an exemplary case. There are documents that include several terms. First, the term-document matrix X is structured. In this example, the term frequency is utilized for better understanding of the method because the example data set is small. In the proposed method, the tf-idf is utilized in this step instead of term frequency. Second, the matrix is decomposed into the product of three other matrices. In this case, k is chosen as 2, among the 5×5 diagonal matrix of singular values, which are sorted in descending order. Third, the latent semantic space is structured, through dimension reduction. Finally, the document-document matrix is constructed by calculating the cosine similarity between documents. In the result, the similarity score between D4 and D5 is 0.94, because these include the same common term 'car'. However, the similarity score between D2 and D3 is also very high, although both documents do not include common terms. Though there is no one-on-one word-matching between documents in vector space, it is analyzed that the constructed latent semantic space through LSA considers the context of the whole document corpus, in that 'astronaut' in D2 and 'cosmonaut' in D3 are synonyms. This is a remarkable point, which overcomes the limitation of the vector space model.

Figure 5. Exemplary case of latent semantic analysis

Finally, the extracted patent-patent document matrix is transformed to the assignee-assignee matrix, by calculating the average scores of patents that assignees own. The score between the same assignees is set to zero, since the score is meaningless, in that the aim of research is to explore R&D collaboration partners.

(2) Latent Semantic Analysis-based Map

The LSA-based potential R&D collaboration map is generated from the assignee-assignee matrix, by using network analysis. A node and the size of the node mean an assignee and the technological capability of the assignee. An appropriate threshold for the assignee-assignee matrix is chosen through sensitivity analysis, with considering the visibility of the map. Matched assignees with a link are potential R&D collaboration partners, since the contents-based technological similarity between assignees is higher than the chosen threshold.

4. Illustration

4.1. Data and Pre-process

The fuel cell membrane electrode assembly (MEA) technology field is selected as an exemplary study, because this technology field is facing an increasing trend of R&D collaboration and is receiving attention as an energy source for the future. Bibliographic and textual information of patents on fuel cell MEA are collected from the United States Patent and Trademark Office (USPTO) database. The collected 197 patents are registered by 65 assignees, from 1997 to 2006. NBER patent data- BR Bridge (Balasubramanian and Sivadasan 2008) bridging corporation information and patent information, and that investigated by the U.S. Census Bureau, is utilized to normalize the name of patent assignees.

The text-mining package ‘tm’ (Feinerer 2013) of R statistical software is utilized to conduct pre-processing, eliminating noise and stopwords, such as punctuation marks, numbers, conjunctions, and articles in the abstract section of patent documents, and stemming, extracting the root of words, as a preparatory stage for processing textual information.

4.2. R&D Collaboration state map

R&D collaboration state map is generated based on the joint patenting relation using the patents applied from 1995 to 2001 as figure 6. Six firms are mapped as current collaboration partners and three pairs of firms granted patents as a successful result of R&D collaboration while many firms are mapped as potential collaboration partners. NodeXL (Smith et al. 2010), a free, open-source network visualization template for MS Office, is utilized to visualize the information as a map.

Figure 6. R&D collaboration state map at the first period (1995-2001)

4.3. Potential R&D Collaboration Partner map through Bibliographic Coupling Analysis

The NBER data set is utilized to extract a patent-patent matrix with the bibliographic coupling relation, because the NBER patent dataset (Hall et al. 2001) includes the citation relationship. An assignee-assignee matrix (65 × 65) is constructed by averaging the normalized coupling strength (ANCS) values of patents by the assignee after calculating the normalized coupling strength (NCS). Table 1 shows an example of the assignee-assignee matrix. A potential R&D collaboration partner map based on the bibliographic coupling relation is generated as in Figure 7. The 20 firms that have the highest ANCSs are mapped where the threshold in the map is 0.18. Table 2 shows the pairs of potential R&D collaboration partners, based on bibliographic coupling relation with ANCS values, ranks, and locations of corporations, which are represented by the two letter ISO country code. The ANCS value is one, since the Forschungszentrum Julich GmbH and Prof. Dr. Rolf Hempelmann have respectively only one patent which is joint patent.

Figure 7. Potential R&D collaboration partner map, based on the bibliographic coupling relation (threshold = 0.18)

Table 1. Example of assignee-assignee matrix

Table 2. Pairs of potential R&D collaboration partners, based on the bibliographic coupling relation

4.4. Potential R&D Collaboration Partner map through Latent Semantic Analysis

The latent semantic analysis package 'lsa' (Wild et al. 2009) of R statistical software is utilized to extract a patent-patent matrix with semantic technological similarity. An assignee-assignee matrix (65×65) is constructed by calculating the average scores between assignees of cosine similarity values between patents. LSA is conducted, where k is 100, which is the highest matching performance, compared to the R&D collaboration state map at the second period. A potential R&D collaboration partner map based on semantic technological similarity is generated as in Figure 8. The 21 firms that have the highest similarities are mapped, where the threshold is 0.4 in the map. Table 3 shows the pairs of potential R&D collaboration partners based on semantic technological similarity with average cosine similarity values, ranks, and locations of corporations, which are represented by the two letter ISO country code. The average cosine similarity value is one, since the Forschungszentrum Julich GmbH and Prof. Dr. Rolf Hempelmann have respectively only one patent which is joint patent.

Figure 8. Potential R&D collaboration partner map, based on semantic technological similarity (k=100, threshold = 0.4)

Table 3. Pairs of potential R&D collaboration partners, based on latent semantic analysis

4.5. Comparison of results

To verify the validity of the results, R&D collaboration state map at the second period of figure 1 is generated as figure 9. Comparing between potential R&D collaboration partner map based on BCA (figure 7) and R&D collaboration state map at the second period (figure 9), five pairs of assignees, which are marked in italic bold letters (ranked at 1, 9, 10, 12, 14) in Table 2, are matched among the six pairs of assignees in figure 9. Comparing between potential R&D collaboration partner map based on LSA (figure 8) and R&D collaboration state map at the second period (figure 9), three pairs of assignees, which are marked with italic bold letters (ranked at 1, 8, 10) in Table 3, are matched within the 10 highest similarities among the six pairs of assignees in figure 9. Two types of potential R&D collaboration maps provide meaningful results in that the map by BCA matches four pairs among six real collaboration pairs within top 15 technological similarity scores and the map by LSA matches three pairs among six real collaboration pairs within top 10 technological similarity scores.

Figure 9. R&D collaboration state map at the second period (1995-2004)

5. Interpretations and Implications

To discuss the implications on R&D collaboration maps, R&D collaboration state maps at the respective periods are compared. Figure 10 shows the collaborative pairs of firms in the respective period of the R&D collaboration state map. The Hyundai Motor Company, the most remarkable, newly emerged as the most active collaboration firm in the second period, conducts collaborative R&D with Kia Motors Corporation and the Korea Institute of Science and Technology (KIST), in the fuel cell MEA technology field. Both two organizations can take advantages of close collaborative R&D, because both organizations are located in Korea, and in particular, Kia Motors is one of the subsidiaries of Hyundai Motors. The other collaborative pairs, California Institute of Technology - University of Southern California, Johnson Matthey Public Limited Company - Technical Fibre Products Limited, Forschungszentrum Julich GmbH - Prof. Dr. Rolf Hempelmann, Honda Motor Co., Ltd. - Tanaka Kikinzoku Kogyo K.K., are organizations located in the US, the UK, Germany, and Japan, respectively. Otherwise, they are global companies that have their headquarters in those countries. Thus, the existing R&D collaboration activities entirely rely on geographical proximity as the limitation of conventional potential R&D collaborator searching method is mentioned.

Figure 10. Change of R&D collaboration state

However, the proposed method demonstrates the results which can overcome the limitation of the conventional methods. The two potential R&D collaboration partner maps through BCA and LSA present potential R&D collaboration partner pairs overcoming the limitation of geographical proximity. The firms of seven pairs are located in different countries among 20 pairs in the potential R&D collaboration partner map through BCA; whereas, firms of 13 pairs are located in different countries, among 21 pairs in the potential R&D collaboration partner map through LSA, as shown shaded in Table 2 and Table 3.

When searching for collaboration partners based on patent information, which is public information that represents technological features, the degree of relying on geographical proximity is less than the existing collaborative activities. The pairs of partners through the LSA-potential R&D collaboration partner map relies less on geographical proximity than those of the BCA-potential R&D collaboration partner map. The results from the BCA-potential R&D collaboration partner map are more affected by geographical proximity, than those from the LSA-potential R&D collaboration partner map, since patents tend to cite already known references, and sometimes do self-citation. The results are summarized in Table 4. When comparing the R&D collaboration state maps, the states of R&D collaboration highly rely on geographical proximity since the maps are visualized based on the number of joint patents and it can show a real collaboration perspective and the limitation of conventional partner searching process. However, both the potential R&D collaboration partner maps suggest the potential R&D collaboration partners based on the technological similarities using patent citation and textual information without considering geographical proximity.

Table 4. Comparison of collaboration maps

The suggested organizations can be considered as potential strategic R&D collaboration partner candidates, in that the consideration of geographical proximity is decreased when it comes to selecting R&D partners, since the business environment is trending towards globalization. Several scholars have found geographical proximity to be less important for R&D collaborations. In a globalizing economy, innovation partners increasingly seem to collaborate across regional, and, even national boundaries (Vedovello 1997; McKelvey et al. 2003; Mora-Valentin et al. 2004; Saxenian 2006; Moodysson and Jonsson 2007; Ponds et al. 2010; Herrmann et al. 2012). Thus, the prospective partner pair information can be utilized as a useful source when searching R&D partners, to overcome the limitations of geographical location.

A simple application of the proposed method is to consider whether the matched partners are firms or universities/ research institutes, since in general, firms in the same industry are competitors, rather than partners. Nevertheless, there are no permanent collaboration partners and competitors in the radically-changed business environment, as Nalebuff and Brandenburger (1996) put forward the term co-opetition, in order to portray a business relationship that consists of both competition and cooperation. Thus, firms need the additional qualitative analysis stage, with considering those above pros and cons from various perspectives, such as the firms' technology development internal needs and strategic directions, and business relationship with competitors. After that, firms should build concrete collaborative strategies, such as informal interactions as meetings and conferences, establishing joint research collaborations, offering opportunities for knowledge exchange, and co-ordination for sustained interactions (Bruneel et al., 2010).

6. Conclusions

The present research contributes several perspectives. First, the proposed approach is able to intuitively comprehend the collaboration states, technology capabilities of partners, and potential collaborative matching by suggesting a visualized collaboration maps. Second, the quantitative methods using patent bibliographic and textual information can recommend potential collaboration partner pairs, which cannot be considered with existing qualitative methods, such as experts' opinion, or human relationships. Third, potential partners based on common technological interest are matched to overcome geographic proximity, which is a limitation of the conventional processes. Fourth, latent semantic analysis is utilized, to overcome the existing limitation of the keyword-based text-mining approach. In addition, the experts' keyword selection process is removed, by utilizing LSA. Finally, the proposed approach will be utilized as a systematic decision-making support tool for the department of technology strategy of firms, the department of technology transfer of universities or research institutes, and especially technology policy-makers in government institutions, for technology-based small and medium enterprise (SME), which is lacking in resources and information to suggest useful potential collaboration matching information.

There are several limitations to the research. First, the recommended pairs based on technological similarity have possibilities that include not only collaboration partners, but also competitors. However, in the contemporary business environment, there are no permanent collaboration partners and competitors. Thus, additional qualitative in-depth analysis should be conducted, after identifying potential collaboration partners based on technological similarity. Second, the proposed approach tried to search the potential collaboration partners only in the same industry. Even though firms are traditionally understood to collaborate primarily for sharing costs and the risks of R&D with others in the same industry (Douglas 1990), industrial firms sometimes collaborate more to create new technological options, and access complementary research strengths, which are unavailable to firms in the other industries (Vonortas 1997).

In future research to surmount the above limitations, several factors should be investigated, to conduct additional in-depth analysis for sorting out the partners and competitors. The approach to obtain synergistic effect between heterogeneous industries by collaboration should be necessary. Research could be discussed on the possibility of fusion between industries, based on collaborative R&D.

Reference

Ahlgren, P. and B. Jarneving. 2008. Bibliographic coupling, common abstract stems and clustering: A comparison of two document-document similarity approaches in the context of science mapping. *Scientometrics* 76, no. 2: 273-90.

Amid, A., S.H. Ghodsypour, and C. O'brien. 2006. Fuzzy multi objective linear model for supplier selection in a supply chain. *International Journal of Production Economics* 104, no. 2: 394-407.

Arranz, N. and J.C. Fdez de Arroyabe. 2008. The choice of partners in R&D cooperation: an empirical analysis of Spanish firms. *Technovation* 28, no. 1:88-100.

Atkins, H. 1999. The ISI Web of Science- links and electronic journals. D-Lib Magazine 5, no. 9 Available at: http://www.dlib.org/dlib/september99/atkins/09atkins.html.

Bailey, W.J., R. Masson and R. Raeside. 1998. Choosing successful technology development partners: A best-practice model. *International Journal of Technology Management* 15, no. 1: 124-38.

Balasubramanian, N. and J. Sivadasan. 2008. NBER Patent Data-BR Bridge: User Guide and Technical Documentation. Working Paper, Univeristy of Michigan.

Belderbos, R., M. Carree, and B. Lokshin. 2004. Cooperative R&D and firm performance. *Research Policy* 33, no. 10:1477-92.

Benfratello, L. and A. Sembenelli. 2002. Research joint ventures and firm level performance. *Research Policy* 31, no. 4:493-507.

Bichtelet, J. and Eaton, E.A. 1980. The combined use of bibliographic coupling and cocitation for document retrieval. *Journal of the American Society for Information Science* 31, no. 4: 278-82.

Boyack, K.W. and R. Klavans. 2010. Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?. *Journal of the American Society for Information Science and Technology* 61, no. 12: 2389-2404.

Braam, R.R., H.F. Moed, and A.F. van Raan. 1991. Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for Information Science* 42, no. 4: 233-51.

Bradford, R.B. 2005. Efficient discovery of new information in large text databases. Intelligence and Security Informatics, Springer Berlin Heidelberg, 374-80.

Brouthers, K.D., L.E. Brouthers, T.J. Wilkinson. 1995. Strategic alliances: choose your partners. *Long Range Planning* 28, no. 3:18-25.

Bruce, M., L. Fiona, L. Dale and W. Dominic. 1995. Success factors for collaborative product development: A study of suppliers of information and communication technology. *R&D Management* 25, no. 1: 33-44.

Bruneel, J., P. d'Este, and A. Salter. 2010. Investigating the factors that diminish the barriers to university-industry collaboration. *Research Policy* 39, no. 7: 858-68.

Butcher, J. and P. Jeffrey. 2005. The use of bibliometric indicators to explore industry-academia collaboration trends over time in the field of membrane use for water treatment. *Technovation* 25, no. 11:1273-80.

Chang, S.L., R.C. Wang, and S.Y. Wang. 2006. Applying fuzzy linguistic quantifier to select supply chain partners at different phases of product life cycle. *International Journal of Production Economics* 100, no. 2: 348-59.

Chen, D.Z., M.H. Huang, H.C. Hsieh, and C.P. Lin. 2011. Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology. *Journal of Informetrics* 5, no. 3: 400-12.

Chen, D.Z., Y.S. Sung, and C.H. Kuan. 2010. Identifying core patents by citations, bibliographic coupling and co-citation. Proceedings of the Eleventh International Conference on Science and Technology Indicators, 69-70.

Chen, S.H., H.T. Lee, and Y.F. Wu. 2008. Applying ANP approach to partner selection for strategic alliance. *Management Decision* 46, no.3: 449-65.

Chen, S.H., P.W. Wang, C.M. Chen, and H.T. Lee. 2010. An analytic hierarchy process approach with linguistic variables for selection of an R&D strategic alliance partner. *Computers & Industrial Engineering* 58, no. 2: 278-287.

Chesbrough, H. 2003. Open innovation: How companies actually do it. *Harvard Business Review* 81, no. 7: 12-14.

Cleverdon, C. 1967. The Cranfield tests on index language devices. Aslib Proceedings 19, no.6: 173-94.

Das, T.K. and B.S. Teng. 2000. A resource-based theory of strategic alliances. *Journal of Management* 26, no. 1: 31-60.

Dean, J. and M.R. Henzinger. 1999. Finding related pages in the World Wide Web. Proceedings of the Eigtth International World Wide Web Conference (WWW8), 389-401.

Deerwester, S., S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman. 1990. Indexing Latent Semantic Analysis. *Journal of the American Society for Information Science* 41, no. 6: 391-407.

Ding, J.F., and G.S. Liang. 2005. Using fuzzy MCDM to select partners of strategic alliances for liner shipping. *Information Sciences* 173, no. 1: 197-225.

Douglas, J. 1990. Consortia for R&D Advantage. *EPRI Journal* 15: 4-15.

Dumais S.T. 1995. Using LSI for information filtering:TREC-3 experiments. Proceedings of the 3rd Text REtrieval Conference(TREC3), 219-30.

Dumais. S.T., G.W. Furnas, T.K. Landauer, S. Deerwester, R. Harshman. 1988. Using latent semantic analysis to improve access to textual information. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 281-85.

Emden, Z., R.J. Calantone, and C. Droge. 2006. Collaborating for new product development: selecting the partner with maximum potential to create value. *Journal of product innovation management* 23, no. 4: 330-41.

Feinerer, I. 2013. Introduction to the tm Package Text Mining in R. tiré de http://cran. r-project. org/web/packages/tm/vignettes/tm. pdf, consulté le, 18.

Feng, B., Z.P. Fan, and J. Ma. 2010. A method for partner selection of codevelopment alliances using individual and collaborative utilities. *International journal of production economics* 124, no. 1: 159-70.

Fischer, M., H. Jahn, and T. Teich. 2004. Optimizing the selection of partners in production networks. Robotics and Computer-Integrated Manufacturing 20, no. 6: 593-601.

Foltz, P.W. and S.T. Dumais. 1992. Personalized Information Delivery: An analysis of information filtering methods. Communications of the ACM 34, no. 12: 51-60.

Fritsch, M. and R. Lukas. 2001. Who cooperates on R&D? *Research Policy* 30, no. 2: 297-312.

Fujiwara, T. 2014. Real options analysis on strategic partnerships of biotechnological start-ups. *Technology Analysis & Strategic Management* ahead-of-print : 1-22.

Furnas, G.W., T.K. Landauer, L.M. Gomez, and S.T. Dumais. 1983. Statistical semantics: Analysis of the potential performance of key-word information systems. *Bell System Technical Journal* 62, no. 6: 1753-1806.

Geringer, J.M. 1991. Strategic determinants of partner selection criteria ininter-national joint ventures. *Journal of International Business Studies* 22, no. 1: 41-62.

Geringer, J.M., L. Hebert. 1991. Measuring performance of international joint ventures. *Journal of International Business Studies* 22, no. 2: 249-63.

Geum, Y., S. Lee, B. Yoon, and Y. Park. 2013. Identifying and evaluating strategic partners for collaborative R&D: Index-based approach using patents and publications. *Technovation* 33, no. 6: 211-24.

Glänzel, W. and H.J. Czerwon. 1995. A new methodological approach to bibliographic coupling and its application to research-front and other core documents. Proceedings of 5th International Conference on Scientometrics and Informetrics, 167-76.

Gong, Y., and X. Liu. 2001. Creating Generic Text Summaries. Proceedings of Sixth International Conference on Document Analysis and Recognition, 903-7.

Hajidimitriou, Y.A., and A.C. Georgiou. 2002. A goal programming model for partner selection decisions in international joint ventures. *European Journal of Operational Research* 138, no. 3: 649-62.

Hakanson, L. and P. Lorange. 1991. R&D based co-operative ventures. In Corporate and Industry Strategies for Europe, ed. L.G. Matsson and B. Styme, 235-63. Amsterdam: North-Holland.

Hall, B.H., A.B. Jaffe, and M. Trajtenberg. 2001. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. No. w8498. National Bureau of Economic Research.

Harter, S.P. 1971. The Cranfield II relevance assessments: A critical evaluation. *The Library Quarterly* 41, no. 3, 229-43.

Henzinger, M.R. 2001. Hyperlink analysis for the Web. *IEEE Internet Computing* 5, no. 1: pp.45-50.

Herrmann, A.M., J.L. Taks, and E. Moors. 2012. Beyond Regional Clusters: On the Importance of Geographical Proximity for R&D Collaborations in a Global Economy—the Case of the Flemish Biotech Sector. *Industry and Innovation* 19, no. 6: 499-516.

Holmberg, S.R., and J.L. Cummings. 2009. Building successful strategic alliances: Strategic process and analytical tool for selecting partner industries and firms. *Long Range Planning* 42, no. 2: 164-93.

Huang, M.H., L.Y. Chiang, and D.Z. Chen. 2003. Constructing a patent citation map using bibliographic coupling: A study of Taiwan's high-tech companies. *Scientometrics* 58, no. 3: 489-506.

Huang, X.G., Y.S. Wong, and J.G. Wang. 2004. A two-stage manufacturing partner selection framework for virtual enterprises. *International Journal of Computer Integrated Manufacturing* 17, no. 4: 294-304.

Ireland, R.D., M.A. Hitt, and D. Vaidyanath. 2002. Alliance management as a source of competitive advantage. *Journal of Management* 28, no. 3: 413-46.

Jarneving, B. 2007. Bibliographic coupling and its application to research-front and other core documents. *Journal of Informetrics* 1, no. 4: 287-307.

Jeon, J., C. Lee, and Y. Park. 2011. How to use patent information to search potential technology partners in open innovation. *Journal of Intellectual Property Rights* 16, no. 5: 385-93.

Jeon, J., H. Lee, and Y. Park. 2011. Implementing technology roadmapping with supplier selection for semiconductor manufacturing companies. *Technology Analysis & Strategic Management* 23, no.8: 899-918.

Kessler, M.M. 1963. Bibliographic coupling between scientific papers. *American Documentation* 14, no. 1: 10-25.

Kim C. and J. Song 2007. Creating new technology through alliances: An empirical investigation of joint patents. *Technovation* 27, no. 8: 461-470.

Landauer, T.K. and S.T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review* 104, no.2: 211-40.

Landauer, T., D. Laham, and M. Derr. 2004. From paragraph to graph: Latent semantic analysis for information visualization. Proceedings of the National Academy of Science 101: 5214-19.

Lee, S., Y. Geum, B. Yoon, M. Kim, and J. Shin. 2010. Strategic partner selection for collaborative R&D: Literature-based technology intelligence. Proceedings of R&D management conference, Manchester, UK (Vol. 30).

Lhuillery, S. and E. Pfister. 2009. R&D cooperation and failures in innovation projects: Empirical evidence from French CIS data. *Research Policy* 38, no. 1: 45-57.

Ma, R. 2012. Author bibliographic coupling analysis: A test based on a Chinese academic database. *Journal of Informetrics* 6, no. 4: 532-42.

McKelvey, M., H. Alm, and M. Riccaboni. 2003. Does co-location matter for formal knowledge collaboration in the Swedish biotechnology pharmaceutical sector?. *Research Policy* 32, no. 3: 483-501.

Miotti, L. and F. Sachwald. 2003. Co-operative R&D: why and with whom?: An integrated framework of analysis. *Research Policy* 32, no. 8: 1481-99.

Monay, F., and D. Gatica-Perez. 2003. On image auto-annotation with latent space models. Proceedings of the 11th ACM international conference on Multimedia, ACM, 275-78.

Moodysson, J. and O. Jonsson. 2007. Knowledge collaboration and proximity: the spatial organization of biotech innovation projects. *European Urban and Regional Studies* 14, no. 2: 115-31.

Mora-Valentin, E.M., A. Montoro-Sanchez, and L.A. Guerras-Martina. 2004. Determining factors in the success of R&D cooperative agreements between firms and research organizations. *Research Policy* 33, no. 1: 17-40.

Morris, S.A., G.G. Yen, Z. Wu, and B. Asnake. 2003. Time line visualization of research fronts. *Journal of the American Society for Information Science and Technology* 54, no. 5: 413-22.

Nakamura, M. 2003. Research alliances and collaborations: introduction to the special issue. *Managerial and Decision Economics* 24, no. 2-3: 47-9.

Nalebuff, B. and A. Brandenburger. 1996. Co-opetition. Profile Books Limited.

Nielsen, B. 2003. An empirical investigation of the drivers of international strategic alliance formation. *European Management Journal* 21, no. 3: 301-22.

Niu, S.H., S.K. Ong, and A.Y.C. Nee. 2012. An enhanced ant colony optimiser for multi-attribute partner selection in virtual enterprises. *International Journal of Production Research* 50, 8: 2286-303.

Persson, O. 1994. The intellectual base and research fronts of JASIS 1986-1990. *Journal of the American Society for Information Science* 45, no. 1: 31-8.

Pisano, G.P. 1990. The R&D boundaries of the firm: An empirical analysis. *Administrative Science Quarterly* 35, no. 1: 153-76.

Ponds, R., F. Van Oort, and K. Frenken. 2010. Innovation, spillovers and university-industry collaboration: an extended knowledge production function approach, *Journal of Economic Geography* 10, no. 2: 231-55.

Saen, R.F. 2007. Suppliers selection in the presence of both cardinal and ordinal data. *European Journal of Operational Research* 183, no.2: 741-7.

Salton, G. 1971. The SMART retrieval system—experiments in automatic document processing. Englewood Cliffs, NJ: Prentice Hall.

Salton, G. and M. McGill. 1983. Introduction to modern information retieval, McGraw-Hill.

Saxenian, A.L. 2006. The new argonauts: Regional advantage in a global economy, Cambridge, MA: Harvard University Press.

Small, H. 1973. Co-citation in the scientific liternature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science* 24, no. 4: 265-69.

Small, H.G. and B.C. Griffith. 1974. The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies* 4, no. 1: 17-40.

Smith, M., N. Milic-Frayling, B. Shneiderman, E. Mendes Rodrigues, J. Leskovec, and C. Dunne. 2010. NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010. http://nodexl.codeplex.com/ from the Social Media Research Foundation, http://www.smrfoundation.org.

Solesvik, M. Z., and M. Gulbrandsen.2013. Partner Selection for Open Innovation. Technology Innovation Management Review April 2013: Open Innovation and Entrepreneurship.

Solesvik, M. Z., and S. Encheva. 2010. Partner selection for interfirm collaboration in ship design. *Industrial Management & Data Systems* 110, no.5: 701-17.

Solesvik, M.Z., and P. Westhead. 2010. Partner selection for strategic alliances: case study insights from the maritime industry. *Industrial Management & Data Systems* 110, 6: 841-60.

Swanson, D.R. 1971. Some unexplained aspects of the Cranfield tests of indexing performance factors. *The Library Quarterly* 41, no. 3: 223-28.

Tether, B.S. 2002. Who co-operates for innovation, and why: An empirical analysis. *Research Policy* 31, no. 6: 947-67.

Tyler, B.B., H.K. Steensma. 1995. Evaluating technological collaborative opportunities: a cognitive modelling perspective. *Strategic Management Journal* 16: 43-70.

Van den Besselaar, P. and G. Heimeriks. 2006. Mapping research topics using word-reference co-occurrences: A method and an exploratory case study. *Scientometrics* 68, no. 3: 377-93.

Vedovello, C. 1997. Science parks and university-industry interaction: geographical proximity between the agents as a driving force. *Technovation* 17, no. 9: 491-502.

Vonortas, N.S. 1997. Cooperation in Research and Development, Boston/ Dordrecht/ London: Kluwer Academic Publishers.

Wang, T.C., and Y.H. Chen. 2007. Applying consistent fuzzy preference relations to partnership selection. *Omega* 35, no. 4: 384-8.

Wild, F., D. Rstem, and M.F. Wild. 2009. The lsa Package.

Wu, W.Y., H.A. Shih, and H.C. Chan. 2009.The analytic network process for partner selection criteria in strategic alliances. *Expert Systems with Applications* 36, no. 3: 4646-53.

Yan, E. and Y. Ding. 2012. Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other. *Journal of the American Society for Information Science and Technology* 63, no. 7: 1313-26.

Yeh, H.Y., Y. S. Sung, H.W. Yang, W.C. Tsai, and D.Z. Chen. 2013. The bibliographic coupling approach to filter the cited and uncited patent citations: a case of electric vehicle technology. *Scientometrics* 94, no. 1: 75-93.

Yoon, B., B. Song. 2014. A systematic approach of partner selection for open innovation. *Industrial Management & Data Systems* 114, no. 7

Zeng, Z.B., Y. Li, and W. Zhu. 2006. Partner selection with a due date constraint in virtual enterprises. *Applied Mathematics and Computation* 175, no. 2: 1353-65.

Table 5. Example of assignee-assignee matrix

	Assig.1	Assig.2	Assig.3	···	···	Assig.64	Assig.65
Assig.1	0	0.004379	0.007468	···	···	0.010062	0
Assig.2	0.004379	0	0.000368	···	···	0.005415	0
Assig.3	0.007468	0.000368	0	···	···	0.005005	0
···	···	···	···	0	···	···	···
···	···	···	···	···	0	···	···
Assig.64	0.010062	0.005415	0.005005	···	···	0	0
Assig.65	0	0	0	···	···	···	0

Table 6. Pairs of potential R&D collaboration partners based on bibliographic coupling relation

Rank	Assignee A	Loc.	Assignee B	Loc.	ANCS
1	***Forschungszentrum Julich GmbH***	***DE***	***Prof. Dr. Rolf Hempelmann***	***DE***	***1***
2	MTI MicroFuel Cells Inc.	US	The Gillette Company	US	0.51388
3	Giner Electro Chemical Systems, LLC	US	The Gillette Company	US	0.5
4	Institute of Nuclear Energy Research	TW	Korea Institute of Energy Research	KR	0.4082
5	Giner Electro Chemical Systems, LLC	US	MTI MicroFuel Cells Inc.	US	0.3707
6	Giner Electro Chemical Systems, LLC	US	University of Southern California	US	0.3638
7	3M Innovative Properties Company	US	Minnesota Mining and Manufacturing Company	US	0.3549
8	Danish Power Systems APS	DK	Kia Motors Corporation	KR	0.3535
9	***Hyundai Motor Company***	***KR***	***Korea Institute of Science and Technology***	***KR***	***0.3333***
10	***Hyundai Motor Company***	***KR***	***Kia Motors Corporation***	***KR***	***0.3333***
11	MTI MicroFuel Cells Inc.	US	University of Southern California	US	0.2512
12	***Johnson Matthey Public Limited Company***	***GB***	***Technical Fibre Products Limited***	***GB***	***0.25***
13	The Gillette Company	US	University of Southern	US	0.2425

			California		
14	***California Institute of Technology***	***US***	***University of Southern California***	***US***	***0.2415***
15	NuVant Systems, LLC	US	Samsung Electronics Co., Ltd.	KR	0.2357
16	David Fuel Cell Components, S.L.	ES	IRD Fuel Cell A/S	DK	0.2132
17	Aisin Seiki Kabushiki Kaisha	JP	Ion Power, Inc.	US	0.2053
18	Polyfuel, Inc.	US	The Gillette Company	US	0.1893
19	Industrial Technology Research Institute	TW	Ion Power, Inc.	US	0.1890
20	Firma Carl Freudenberg	DE	Protonex Technology Corporation	US	0.1826

Table 7. Pairs of potential R&D collaboration partners based on latent semantic analysis

Rank	Assignee A	Loc.	Assignee B	Loc.	Cos Sim.
1	***Forschungszentrum Julich GmbH***	***DE***	***Prof. Dr. Rolf Hempelmann***	***DE***	***1***
2	Japan Gore-Tex, Inc.	JP	Permelec Electrode Ltd.	JP	0.6758
3	Industrial Technology Research Institute	TW	Kabushiki Kaisha Toyota Chuo Kenkyusho	JP	0.6688
4	Kabushiki Kaisha Toyota Chuo Kenkyusho	JP	N. E. Chemcat Corporation	JP	0.6578
5	Korea Institute of Energy Research	KR	University of Southern California	US	0.6548
6	Industrial Technology Research Institute	TW	Tanaka Kikinzoku Kogyo K.K.	JP	0.6203
7	Kabushiki Kaisha Toyota Chuo Kenkyusho	JP	Tanaka Kikinzoku Kogyo K.K.	JP	0.5852
8	***Hyundai Motor Company***	***KR***	***Kia Motors Corporation***	***KR***	***0.5259***
9	Industrial Technology Research Institute	TW	University of Southern California	US	0.5236
10	***Hyundai Motor Company***	***KR***	***Korea Institute of Science and Technology***	***KR***	***0.4889***
11	Industrial Technology Research Institute	TW	Southwest Research Institute	US	0.4676
12	Danish Power Systems APS	DK	N. E. Chemcat Corporation	JP	0.4584

13	Institute of Nuclear Energy Research	TW	Permelec Electrode Ltd.	JP	0.4461
14	Energy Partners, L.C.	US	The Texas A&M University System	US	0.4460
15	Energy Partners, L.C.	US	Hydrogenics Corporation	CA	0.4434
16	IRD Fuel Cell A/S	DK	Tanaka Kikinzoku Kogyo K.K.	JP	0.4386
17	Kabushiki Kaisha Toyota Chuo Kenkyusho	JP	Kia Motors Corporation	KR	0.4263
18	Institute of Nuclear Energy Research	TW	Ion Power, Inc.	US	0.4207
19	N. E. Chemcat Corporation	JP	Tanaka Kikinzoku Kogyo K.K.	JP	0.4129
20	Kia Motors Corporation	KR	N. E. Chemcat Corporation	JP	0.4061
21	E. I. du Pont de Nemours and Company	US	Tanaka Kikinzoku Kogyo K.K.	JP	0.4026

Table 8. Comparison of collaborations maps

Types of collaboration map	Perspective	Impact of proximity
R&D collaboration state map	Real collaboration	High
BCA-potential R&D collaboration partner map	Citation pattern similarity	Medium
LSA-Potential R&D collaboration partner map	Content similarity	Low

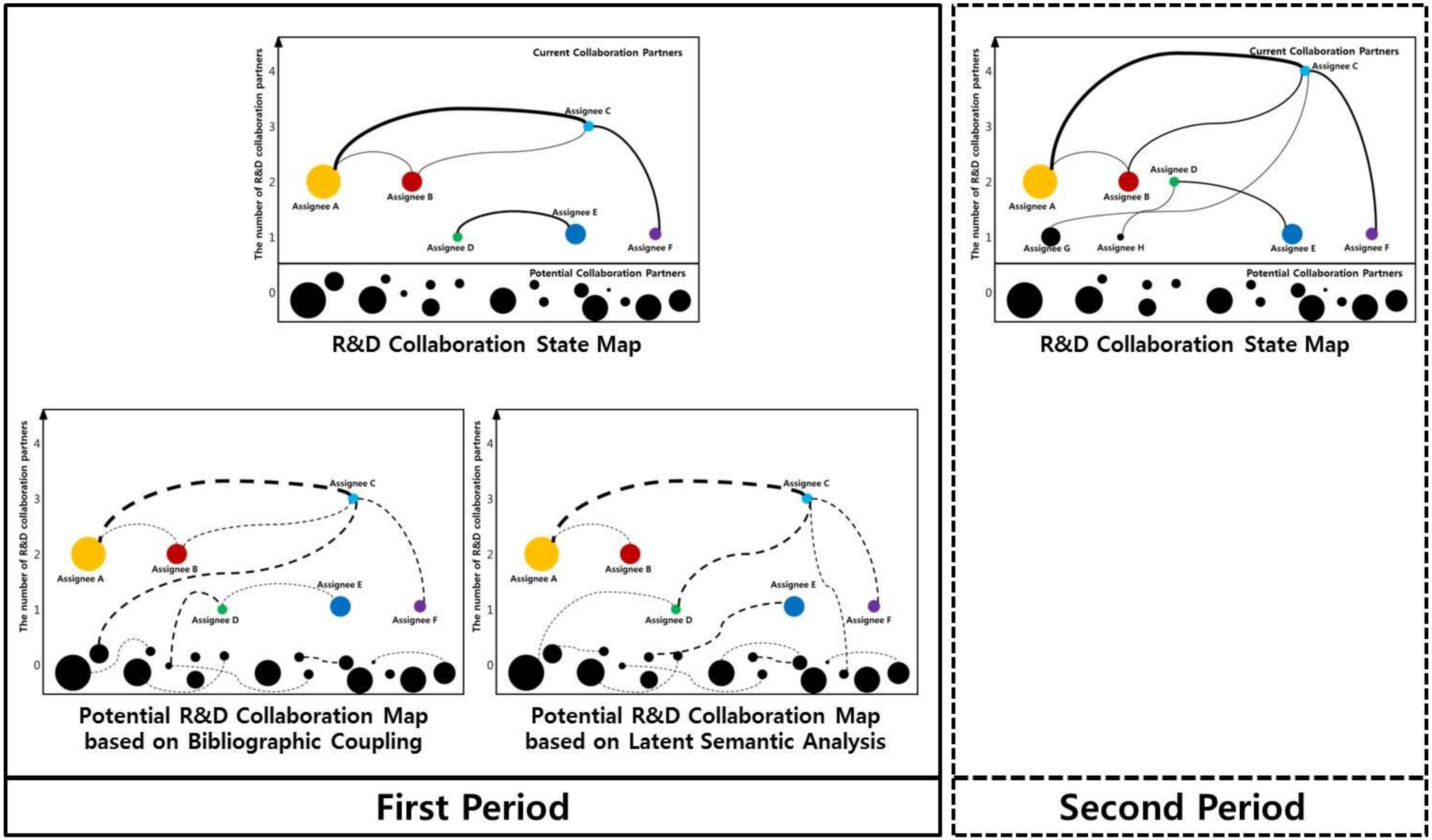

Figure 7. Research Concept

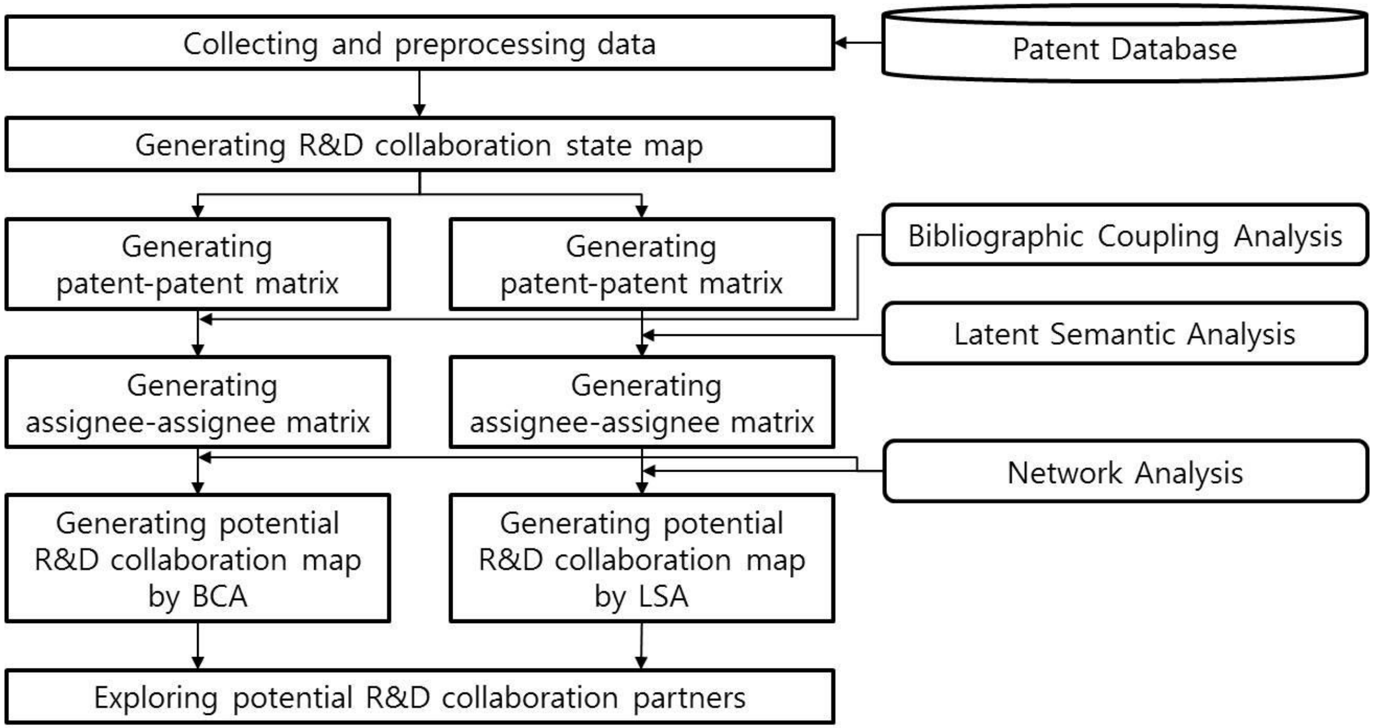

Figure 8. Research Framework

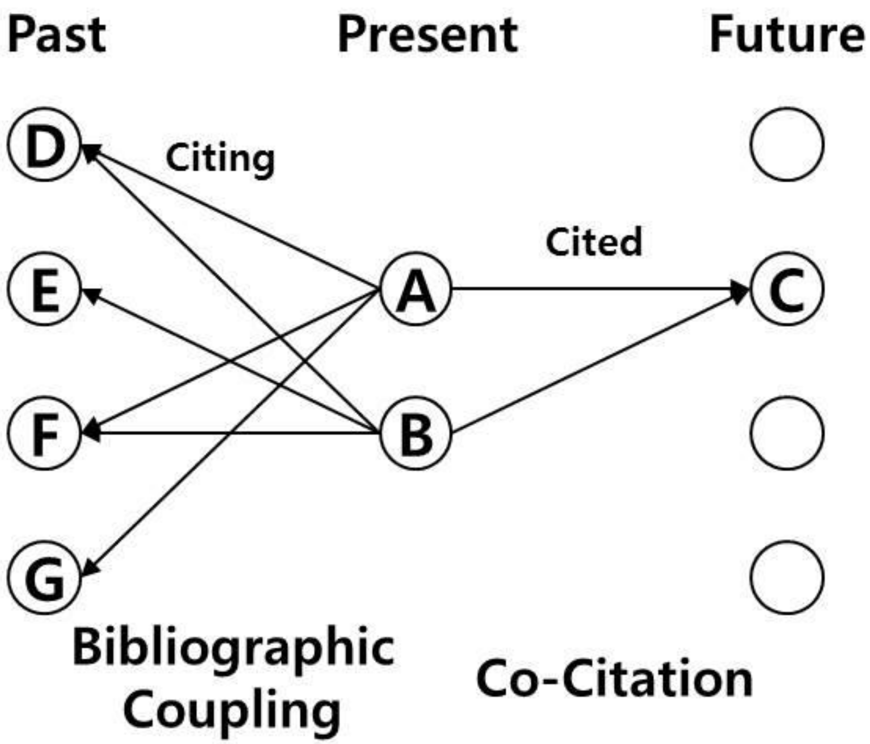

Figure 9. Bibliographic coupling and co-citation

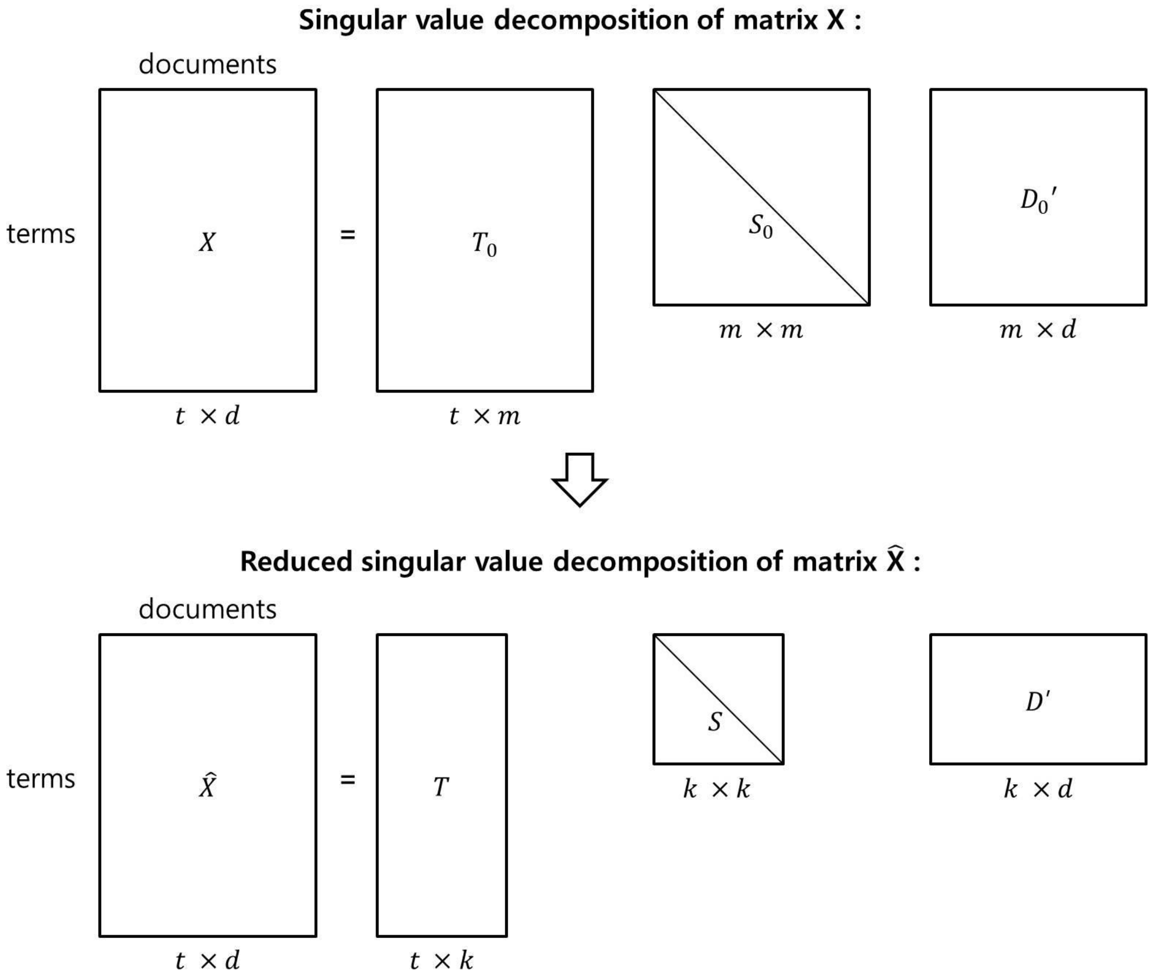

Figure 10. Schematic of the SVD and reduced SVD of matrix [Deerwester et al., 1990]

Doc. 1	cosmonaut, moon, car
Doc. 2	astronaut, moon
Doc. 3	cosmonaut
Doc. 4	car, truck
Doc. 5	car
Doc. 6	truck

Document Corpus

Document

Term

	D1	D2	D3	D4	D5	D6
cosmonaut	1	0	1	0	0	0
astronaut	0	1	0	0	0	0
moon	1	1	0	0	0	0
car	1	0	0	1	1	0
truck	0	0	0	1	0	1

Term-Document Matrix: X (5×6)

Singular Value Decomposition (SVD)

	dim1	dim2	dim3	dim4	dim5
cosmonaut	-0.44	-0.3	0.57	0.58	0.25
astronaut	-0.13	-0.33	-0.59	0	0.73
moon	-0.48	-0.51	-0.37	0	-0.61
car	-0.7	0.35	0.15	-0.58	0.16
truck	-0.26	0.65	-0.41	0.58	-0.09

T_0 (5×5)

1.62	0.00	0.00	0.00	0.00
0.00	1.59	0.00	0.00	0.00
0.00	0.00	1.28	0.00	0.00
0.00	0.00	0.00	1.00	0.00
0.00	0.00	0.00	0.00	0.39

S_0 (5×5)

	D1	D2	D3	D4	D5	D6
dim1	-0.75	-0.28	-0.2	-0.45	-0.33	-0.12
dim2	-0.29	-0.53	-0.19	0.63	0.22	0.41
dim3	0.28	-0.75	0.45	-0.2	0.12	-0.33
dim4	0	0	0.58	0	-0.58	0.58
dim5	-0.53	0.29	0.63	0.19	0.41	-0.22

D_0' (5×6)

Dimension Reduction (k=2)

	D1	D2	D3	D4	D5	D6
dim1	-1.62	-0.60	-0.40	-0.97	-0.71	-0.26
dim2	-0.46	-0.84	-0.30	1.00	0.35	0.65

Cosine Similarity

	D1	D2	D3	D4	D5	D6
D1	1	0	0	0	0	0
D2	0.78	1	0	0	0	0
D3	0.95	0.93	1	0	0	0
D4	0.47	-0.17	0.17	1	0	0
D5	0.74	0.15	0.49	0.94	1	0
D6	0.11	-0.53	-0.20	0.92	0.75	1

Figure 11. Exemplary case of latent semantic analysis

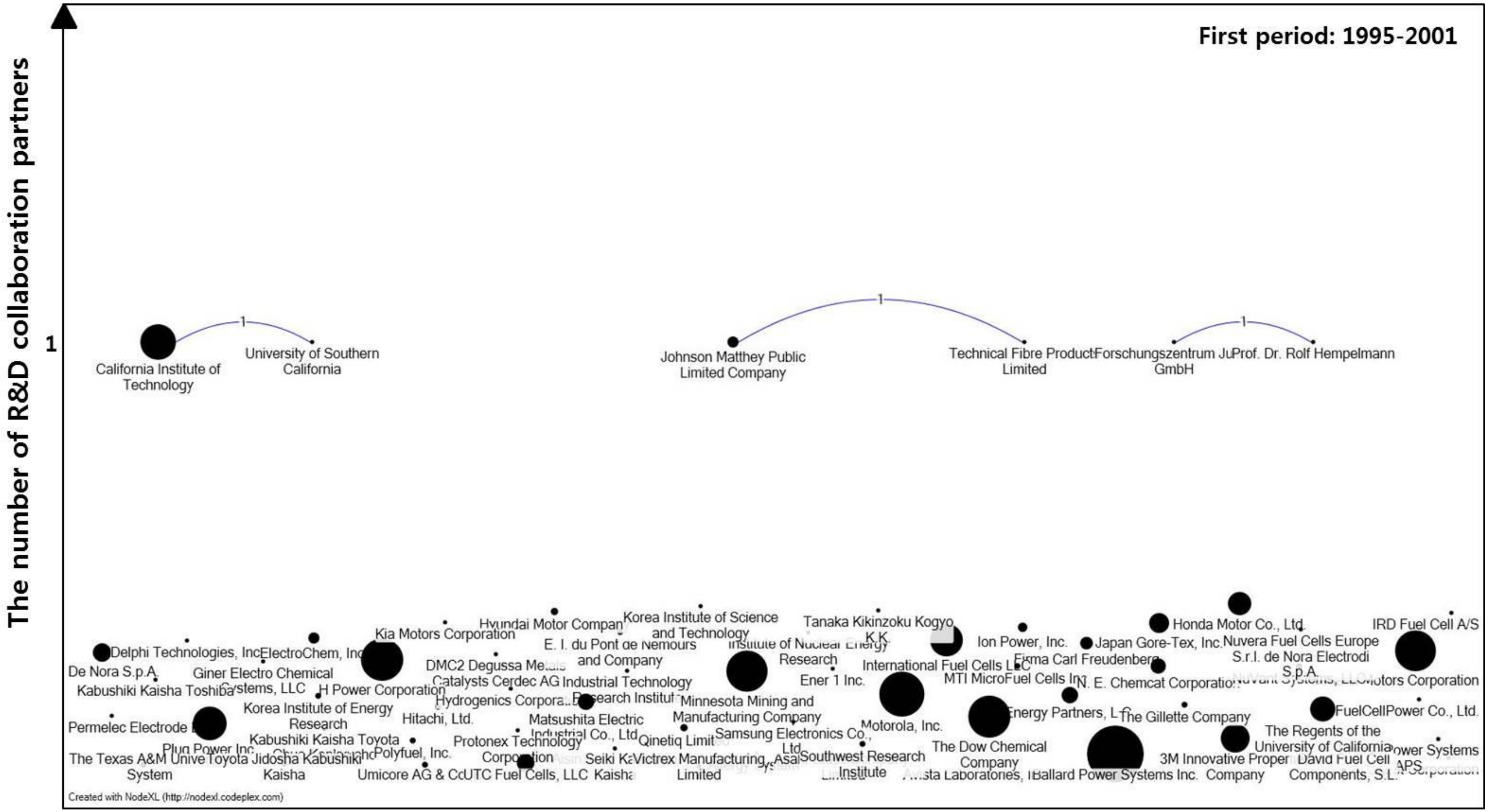

Figure 12. R&D collaboration state map at the first period (1995-2001)

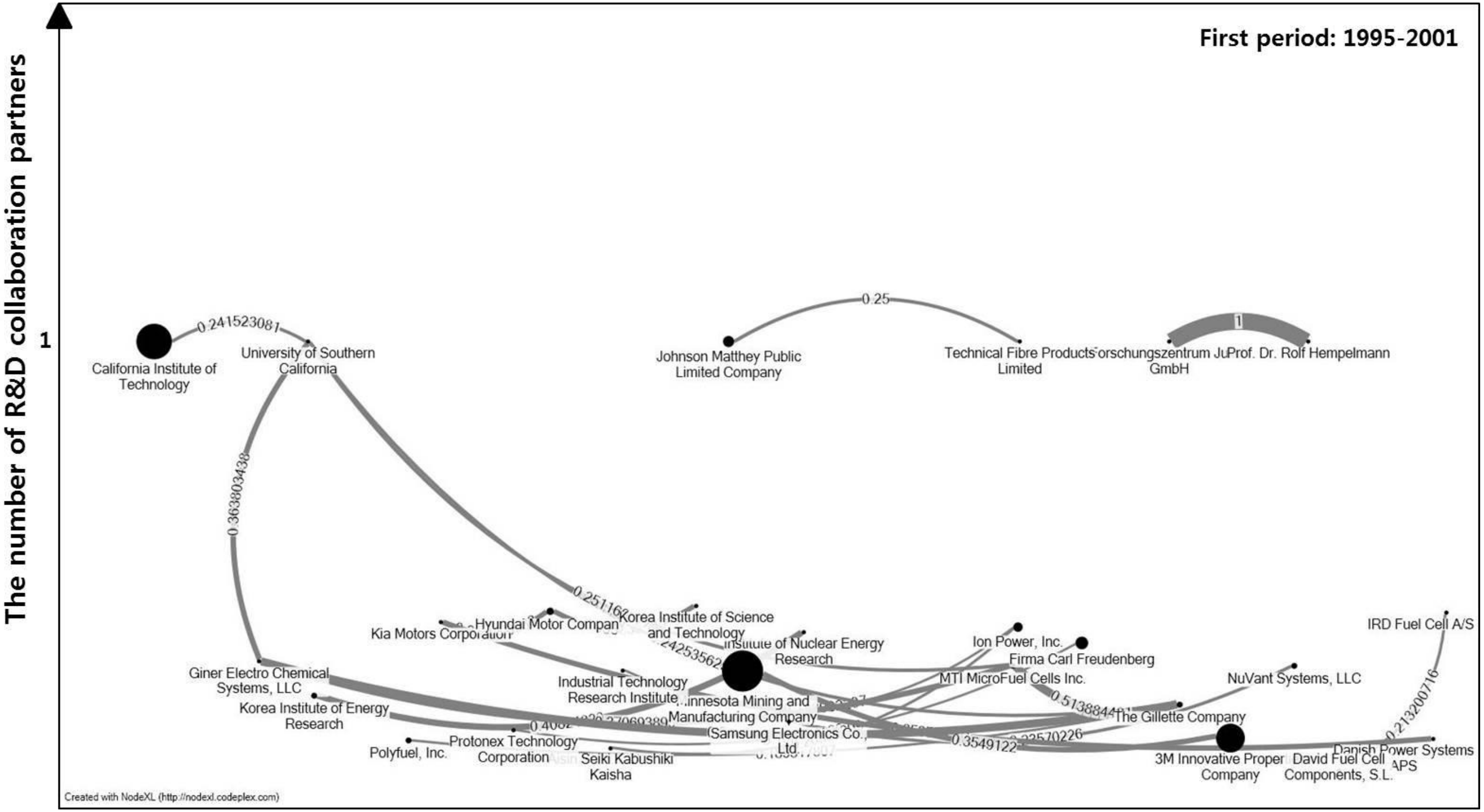

Figure 13. Potential R&D collaboration partner map based on bibliographic coupling relation (threshold = 0.18)

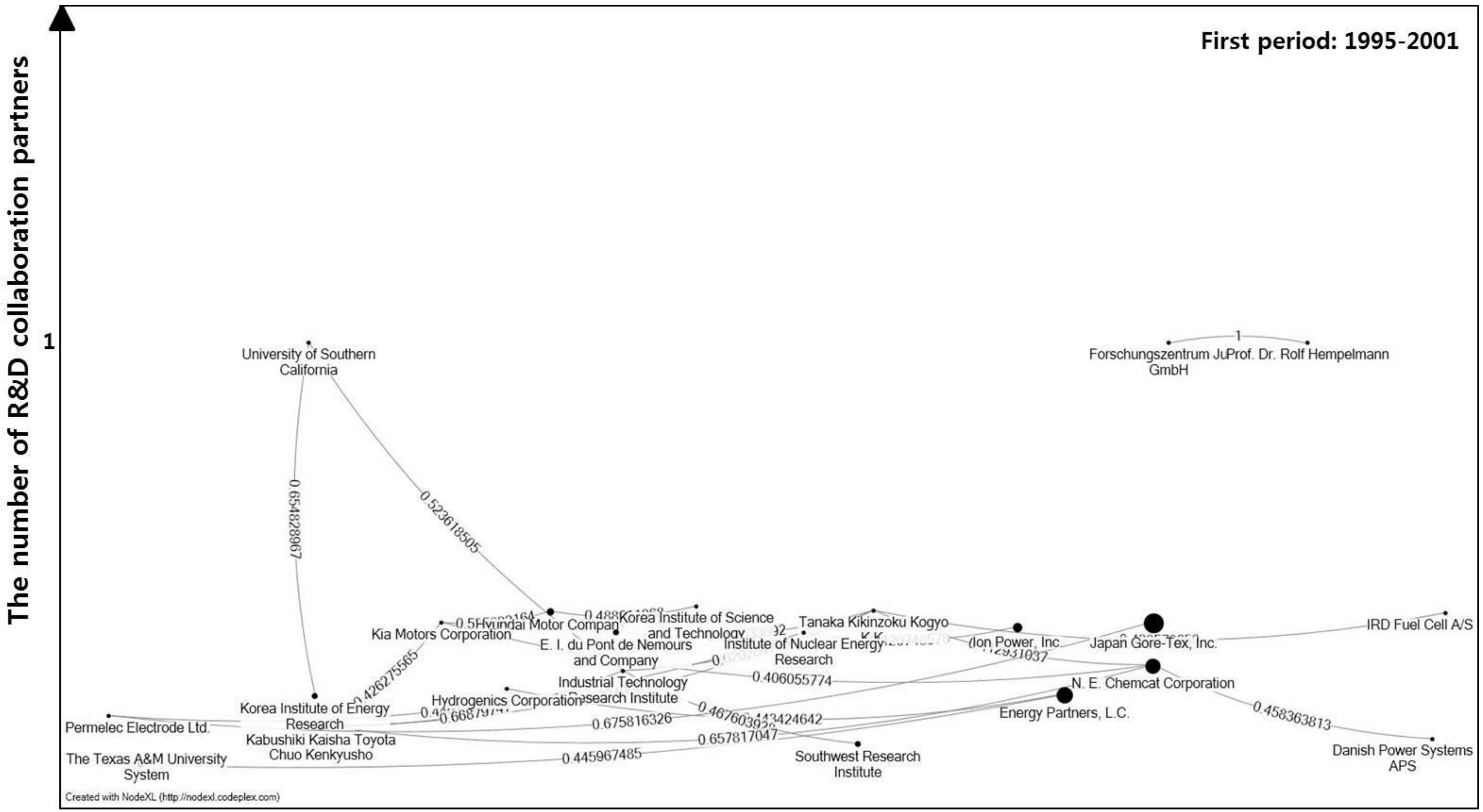

Figure 14. Potential R&D collaboration partner map based on semantic technological similarity (k=100, threshold = 0.4)

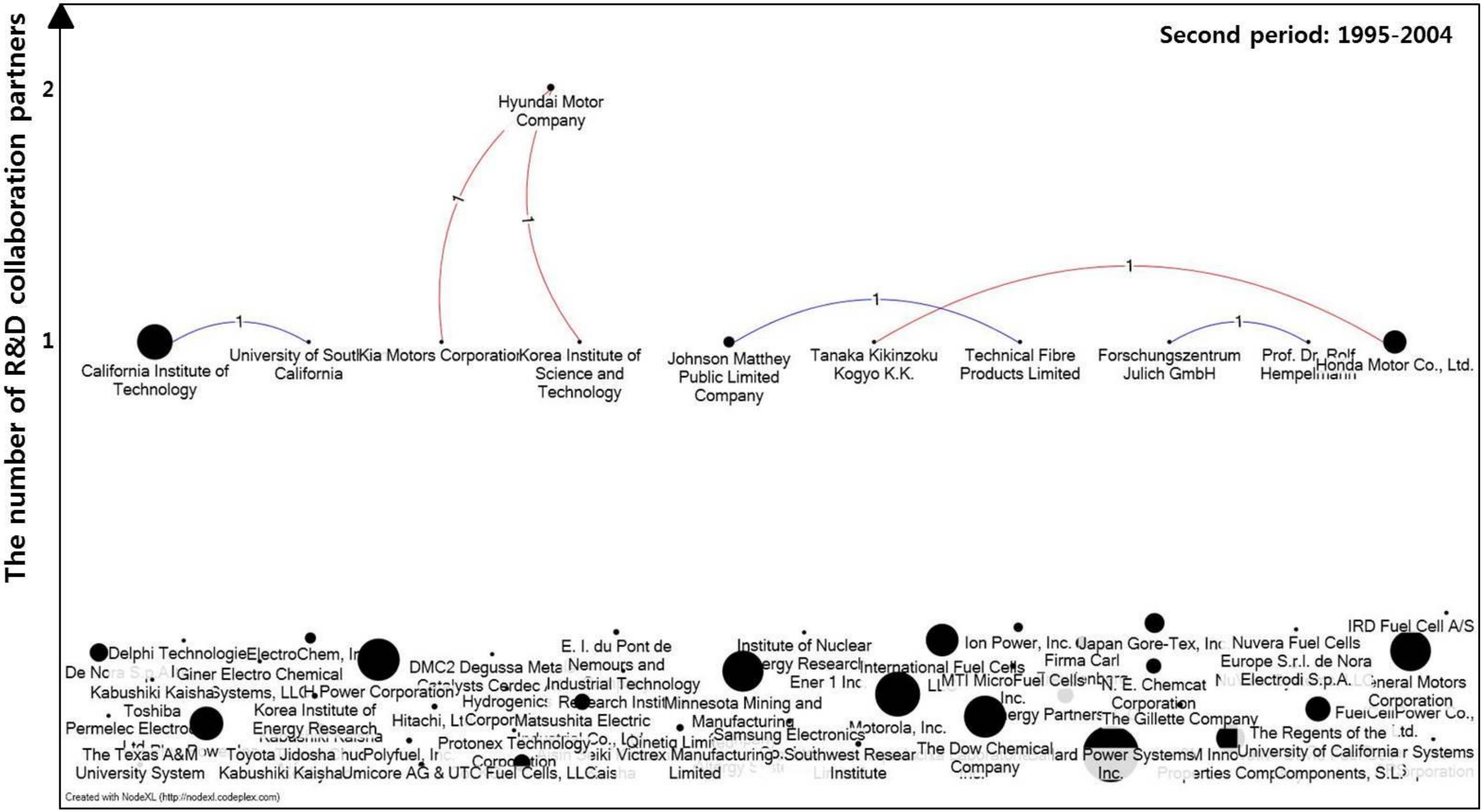

Figure 15. R&D collaboration state map at the second period (1995-2004)

1995~2001	
California Institute of Technology	University of Southern California
Johnson Matthey Public Limited Company	Technical Fibre Products Limited
Forschungszentrum Julich GmbH	Prof. Dr. Rolf Hempelmann

→

1995~2004	
California Institute of Technology	University of Southern California
Johnson Matthey Public Limited Company	Technical Fibre Products Limited
Forschungszentrum Julich GmbH	Prof. Dr. Rolf Hempelmann
Hyundai Motor Company	Kia Motors Corporation
Hyundai Motor Company	Korea Institute of Science and Technology
Honda Motor Co., Ltd.	Tanaka Kikinzoku Kogyo K.K.

Figure 16. Change of R&D collaboration state