Technological Forecasting & Social Change xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Concentric diversification based on technological capabilities: Link analysis of products and technologies

Hyunwoo Kim [a], Suckwon Hong [b], Ohjin Kwon [c,*], Changyong Lee [b,*]

[a] School of Science and Technology Management Policy, University of Science and Technology, 217 Gajeong-ro, Yuseong-gu, Daejeon 34113, Republic of Korea
[b] School of Management Engineering, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulsan 44919, Republic of Korea
[c] Korea Institute of Science and Technology Information, 66 Hoegi-ro, Dongdaemun-gun, Seoul 130-741, Republic of Korea

ARTICLE INFO

Article history:
Received 12 November 2016
Received in revised form 13 February 2017
Accepted 18 February 2017
Available online xxxx

Keywords:
Concentric diversification
Patent-product database
Text mining
Association rule mining
Link prediction

ABSTRACT

This research responds to the needs for concentric diversification by focusing on how firms can find new business opportunities based on their technological capabilities. We propose a systematic approach to identifying potential areas for concentric diversification at a product level via link analysis of products and technologies. For this, first, text mining is utilised to construct an integrated patent-product database from the US patent and trademark database. Second, association rule mining is employed to construct a product ecology network using directed technological relationships between products. Third, a link prediction analysis is conducted to identify potential areas for concentric diversification at a product level. Finally, three quantitative indicators are developed to assess the characteristics of the areas identified. Our case study employs a total of 850,676 patents and 328,288 products in the integrated patent-product database from 2010 to 2014 and shows that the proposed approach enables a wide-ranging search for potential areas for concentric diversification and the quick assessment of their characteristics, with statistically significant results. We believe that the proposed approach will be useful as a complementary tool for decision making for small and medium-sized high-tech companies that are considering entering new business areas, but which have little domain knowledge.

1. Introduction

Diversification is one of the pivotal strategies for organisations to recreate and enlarge their competencies (Ahuja and Lampert, 2001; Hamel and Prahalad, 1994). The existing literature suggests that concentric diversification based on technological capabilities is a relatively low-risk but profitable strategy since it is derived from the reproduction of core competencies that are related to existing products or services (Chen and Chang, 2012; Collis and Montgomery, 1995; Dutton, 1997; Markides, 1997; Zook and Allen, 2001). From a resource-based view, many researchers have verified that technological capabilities have a significant positive effect on the success of diversification (Silverman, 1999). However, while the results of such empirical analyses or case studies have been widely accepted in academia and in practice, a major question still remains as to how decision makers can best identify areas for concentric diversification.

Modelling and analysing technological capabilities for concentric diversification is a task beset with hazards including uncertainty, data unavailability, and the complexity of real world feedback. As such, industrial practitioners depend largely on expert-centric approaches (e.g. brainstorming and Delphi). While internal experts have professional knowledge and experience about corporate technologies, they may have little knowledge about the technologies involved in potential new business areas (Shin et al., 2013). Using external experts (e.g. consultants) may resolve this issue, but they frequently misjudge corporate technological and organisational capabilities and identify promising but inappropriate areas for the particular companies with which they are temporarily working. Hence, such expert-centric approaches need to be supported by high quality and well-organised information (Lee et al., 2013, 2015).

Highlighting the possible avenues for methodological adaptation, there have recently been certain shifts in the direction of research on concentric diversification, from case studies or empirical analysis to evidence-based quantitative approaches. Perhaps the most scientific approaches are offered by patent analysis as this provides global and reliable information about a wide range of technologies (Granstrand et al., 1997). Patent publications are considered valuable data sources as they are published according to international standards and contain information on almost 80% of technologies (Lee et al., 2015). So far, a variety of models and methods have been presented, such as patent citation analysis (Narin, 1994), keyword-based network analysis (Yoon and Park, 2004), keyword-based morphology analysis (Yoon and Park,

* Corresponding authors.
E-mail addresses: hwkim@kisti.re.kr (H. Kim), amoeba94@unist.ac.kr (S. Hong), dbajin@kisti.re.kr (O. Kwon), changyong@unist.ac.kr (C. Lee).

http://dx.doi.org/10.1016/j.techfore.2017.02.025
0040-1625/

2005, 2007; Yoon et al., 2014), and patent-based roadmapping processes (Lee et al., 2009).

However, while previous studies have proved quite useful for various purposes, they are subject to certain limitations in terms of data, methodology, and practicality. First, in terms of data limitations, the outcomes of previous studies are not specific about the areas that are suitable for concentric diversification, but show general patterns of technological relationships at a macro level (Yoon et al., 2015). This is mainly because patents do not explicitly contain product or service information where the patented technologies can be applied (Chiu et al., 2008; Garcia-Vega, 2006; Nesta and Saviotti, 2005). Thus new data sources need to be secured to allow a product- or service-level analysis, covering a wide range of business areas. Second, with respect to methodology limitations, previous studies cannot assist timely decision making in concentric diversification since they have been limited to *ex post* evaluation which measures past performance, impacts or consequence (Breschi et al., 2003; Chen and Chang, 2012; Silverman, 1999; Wang et al., 2015). The areas for concentric diversification identified via previous methods are thus likely to be no more than the existing ones in which many competitors have already entered or shown interests. Another major drawback of previous methods is concerned with an inability to consider the direction of diversification. Note that even though a company owning item A has the possibility to diversify to item B, it does not guarantee the company which owns item B will be able to diversify to item A. Hence, any approach that is proposed needs to model the direction of diversification to enhance the feasibility of analysis results. Finally, from a practical standpoint, the verification of previous methods was usually omitted; and even if verification was carried out, it was done in a domain-specific qualitative way (Archibugi and Pianta, 1992; Breschi et al., 2003; Chen and Chang, 2012; Nasiriyar et al., 2013; Wang et al., 2015). Thus any approach that is proposed should establish external validity to be deployed in practice.

Considering these issues, we propose a systematic approach to identifying potential areas for concentric diversification at a product level via link analysis of products and technologies. The tenet of this research is that significant technological relationships between products extracted from large-scale quantitative databases can provide valuable information on the feasibility of concentric diversification from one area to another. At the heart of the proposed approach are (1) text mining techniques for constructing an integrated patent-product database from the US patent and trademark database; (2) association rule mining with conviction indicators for constructing a product ecology network using directed technological relationships between products; (3) link prediction analysis for identifying potential areas for concentric diversification at a product level; and finally (4) indicator analysis for assessing the characteristics of the areas identified. The approach we propose therefore incorporates the issues noted above into analysis for concentric diversification. We also develop a software system to automate our approach, allowing even those who are unfamiliar with the patent-product database and complex models to benefit from our research results.

We applied the proposed approach to support Korean high-tech companies in discovering their next-growth engines at the request of the Korea Institute of Science and Technology Information (KISTI). Our case study shows, with statistically significant results, that the proposed approach enables a wide-ranging search for potential areas for concentric diversification at the product-level and the quick assessment of their characteristics. We believe that the systematic process and quantitative outcomes our approach offers can facilitate decision making in concentric diversification, especially for small and medium-sized high-tech companies that are considering entering new business areas, but that have little domain knowledge.

This paper is organised as follows. Section 2 presents the background to our research, and Section 3 explains our research framework, which is then illustrated by a case study on semiconductors in Section 4. Finally, Section 5 presents our conclusions.

2. Background

2.1. Concentric diversification

Most diversification strategies fail to deliver value and most successful companies achieve their growth by expanding into logical adjacencies that have shared economies, and not from unrelated diversification or moves into "hot" markets (Chen and Chang, 2012; Collis and Montgomery, 1995; Dutton, 1997; Markides, 1997; Zook and Allen, 2001). Owing to such risks involved in diversification, concentric diversification – defined as entry to a new business area based on companies' core competencies related to their existing products or services (Ansoff, 1965) – has been considered a key strategy to sustain business growth. It is reported that the success rate of concentric diversification has reached around 70 to 90%, whereas approximately 90% of companies' attempts to diversify outside of their core competencies have failed (Zook and Allen, 2001). If it is properly implemented, concentric diversification may have advantages in terms of reducing R&D cost (Cantwell and Piscitello, 2000), reducing time to market (Cantwell and Piscitello, 2000; Granstrand, 1998), and creating synergies with other businesses (Patel and Pavitt, 1997; Valvano and Vannoni, 2003).

Concentric diversification has been the subject of many previous studies. The early studies on concentric diversification were based on case studies or empirical analysis. Many researchers have focused on the relationship between concentric diversification and firm performance. For instance, Rumelt (1974) carried out a pioneering study which found a relationship between concentric diversification and profitability. Palich et al. (2000) found an inverted-U shaped relationship between concentric diversification and firm performance with meta analysis drawn from 55 previous studies using a curvilinear model. Focusing more on technological and innovation competencies, Miller's (2004) longitudinal study with patent citation analysis of 227 firms which diversified from 1980 to 1992, also verified that concentric diversification has a positive effect on technological growth. This was reconfirmed in his follow-up study by citation-weighted patent analysis (Miller, 2006). Quintana-Garcia and Benavides-Velasco (2008) applied a generalized estimating equation (GEE) regression model to verify the relationship between concentric diversification and innovative competencies by using the USPTO patent database. Wang et al. (2015) verified that the relationship between technological capabilities and the success of diversification was altered depending on the market situations.

Highlighting the possible avenues for methodological adaptation, recent studies focus more on quantitative data and scientific methods to assist decision making in concentric diversification. The dominant approach is based on patent analysis which provides global and reliable information about a wide range of technologies (Jaffe, 1986; Patel and Pavitt, 1997; Silverman, 1999). There has been a variety of research done with regards to diversification using patent databases. For instance, Yoon and Park (2004) proposed a text mining-based patent network to visualize technological relatedness and explore technological opportunities. Yoon and Park (2007) presented an integrated use of morphology analysis and conjoint analysis to identify and evaluate technology opportunities from patent documents. Lee et al. (2009) presented a patent-based technology-driven roadmapping process that starts from capability analysis and ends with business opportunity analysis. Seol et al. (2011) proposed an approach to exploring appropriate new business areas at the industry level using text mining techniques and data envelopment analysis.

However, while all these previous studies have proved valuable in using quantitative data and scientific methods, and in providing insights into concentric diversification, they are subject to certain limitations as mentioned. These drawbacks, which provide our underlying motivation, are fully addressed in this study. Table 1 summarises the difference between previous methods and the proposed approach.

Please cite this article as: Kim, H., et al., Concentric diversification based on technological capabilities: Link analysis of products and technologies, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.025

Table 1
Comparison of previous methods and the proposed approach.

Factor	Previous research	Current paper
Approach	*Ex post* evaluation	Predictive analysis
Level of analysis	Technology level (e.g. patent class)	Product level
Data	Mainly patent citations or co-classifications	Integrated patent-product database
Method	Centrality-focused network analysis	Association rule mining and link prediction analysis
Verification	Experts-based qualitative judgments	Statistical analysis as well as expert-based qualitative judgments
Results and implications	Relations between technologies	Potential areas for concentric diversification

2.2. Association rule mining

Association rule mining, which is also called market basket analysis or affinity analysis, is an unsupervised data mining technique that discovers significant relationships among items from a given data set (Agrawal et al., 1993). This method originated from the study of customer transactions databases to determine associations among items purchased. It generates association rules X → Y which indicate "what goes with what", for instance customers who bought item X also bought item Y.

Three general measures for association rules X → Y are as follows. Firstly, support X → Y is defined as the ratio of the number of transactions that include both items X and Y against the total number of transactions. This indicator represents the usefulness of discovered rules using the probability of co-occurrence of items X and Y. Secondly, confidence X → Y measures the ratio of the number of transactions containing item Y among transactions containing item X. This indicator represents the certainty of discovered rules using the conditional probability of Y given X. Finally, lift X → Y is calculated by dividing the confidence by the probability of occurrence of item Y. This indicator shows the statistical dependence between items X and Y.

The procedure of association rule mining consists of two steps. The first step involves searching for frequent co-occurred item sets by generating all the possible combinations of items over a user-specified threshold of support, *minsup*, and selecting the association rules whose confidence is higher than a user-specified threshold, *minconf*. The second step involves selecting the significant rules with a lift value greater than 1.

Over the last decade, association rule mining has been employed in a wide array of research fields including quantitative marketing (Liao and Chen, 2004; Rong et al., 2012), bioinformatics (Creighton and Hahash, 2003; Oyama et al., 2002; Wright et al., 2010), and technology and innovation management (He and Loh, 2010; Shih et al., 2010; Wang, 2012). Moreover, this method has been employed to capture significant relationships between technologies for various purposes such as technology trend analysis (Shih et al., 2010), technology impact analysis (Kim et al., 2011), and automatic technology classification (He and Loh, 2010). This study employs association rule mining with conviction indicators to construct a product ecology network through identifying significant directed technological relationships between products.

2.3. Link prediction

Link prediction is one major stream of research in network analysis for modelling the evolutionary mechanisms of dynamic networks and hidden structures of uncertain networks (Getoor and Diehl, 2005). This method predicts newly added or deleted links in future networks, as well as identifying unobserved or spurious links in current networks based on network properties such as attributes of nodes and the topological information of links (Lü and Zhou, 2011).

The models and methods for link prediction can be classified into three categories: Firstly, similarity-based approaches give a score to a pair of nodes using the similarity between them, given that a pair of nodes with higher similarity is more likely to be linked in future networks than that with lower similarity (Lü and Zhou, 2011). Secondly, maximum likelihood approaches assume some organising principles of the network structure (e.g. hierarchical structure), with the detailed rules and specific parameters obtained by maximising the likelihood of the observed structure (Clauset et al., 2008). Finally, probabilistic approaches abstract the underlying structure from the observed network and predict the missing links by using the trained model (Friedman et al., 1999).

Link prediction analysis has been employed in a wide array of research fields such as bioinformatics and biology (Lei and Ruan, 2013), quantitative marketing (Schafer et al., 2001), security (Clauset et al., 2008), and others. This method plays a particularly important role in the field of bioinformatics and biology as it can significantly reduce the high experimental costs involved in the analysis of biological interactions, such as protein-protein and disease-gene interaction (Lei and Ruan, 2013). Note that about 99% of the molecular interactions in human cells are unknown (Stumpf et al., 2008).

However, despite its potential utilities, the application of link prediction analysis to technology and innovation research has not yet been explored. This method can offer the following advantages for identifying potential areas for concentric diversification based on technological capabilities. Firstly, expert-centric approaches employed in practice have become extremely time-consuming and labour-intensive, as the number of products and services extant, and the complexity of technological knowledge mount. Time, cost, and effort associated with expert-centric approaches can be reduced significantly by supporting good quality and well-organised information. Secondly, if this method is integrated into objective and reliable databases, the results of link prediction analysis can provide insight into future prospects and/or the hidden structure of relationships between products/services and technologies, thereby supporting timely decision making with regard to concentric diversification. Considering these factors, we employ a link prediction analysis with an integrated patent-product database to identify potential areas for concentric diversification.

3. Data and methodology

3.1. Data

We based our research on the integrated patent-product database constructed by KISTI (2012) from the United States Patent and Trademark Office (USPTO) database. The quality of this database has been verified by several domain experts and improved over the last four years,[1] serving as a primary data source for KISTI's technology intelligence system, which is called the technology opportunity discovery (TOD) system (http://tod.kisti.re.kr).

The structures and characteristics of this database are as follows. First, it uses a set of feasible product name entities extracted from the 'Goods and Services' field of the US trademark database through data parsing techniques. The product name entities obtained in this way are considered to be valid since the owner of a trademark for a product should verify its commercial use in the market according to the US Trademark Law. KISTI's product database also includes variants of product names, as they can be written using different expressions: for instance, "AC-DC converter" can be written as "AC-to-DC converter" or "AC/DC converter". Second, it employs the US patents to analyse international technologies because the US is the world's largest patent market where the majority of patents submitted to the USPTO are also submitted in other countries (Lee et al., 2015). Finally, the relationships

[1] For more detailed information, see KISTI (2012).

Please cite this article as: Kim, H., et al., Concentric diversification based on technological capabilities: Link analysis of products and technologies, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.025

between patents and products are constructed via matching product name entities extracted from the US trademark database to product name candidates extracted from the title, abstracts, and claims of US patents, by using text mining and keyword recognition techniques. Specifically, keyword recognition is done via tagging, parsing, and head word classification by using the Stanford parser. The patent description part was excluded from this task since it includes enormous noise data and reduces reliability in matching products and patents.

Table 2 shows the structure of the integrated patent-product data employed in this research. In the table, the data fields of patent and representative products are employed to construct a product ecology network and to identify potential areas for concentric diversification, while the remaining fields are used to assess the characteristics of potential areas for concentric diversification. While this study uses the USPTO database to construct an integrated patent-product database, other types of databases such as the European Patent Office (EPO) and Securities and Exchange Commission (SEC) 10-k databases, can also be employed.

3.2. Methodology

We examine the overall process of the proposed approach, giving a brief explanation of each step at the same time. The suggested approach employs various methods such as text mining, association rule mining, link prediction analysis, and indicator analysis. Given the complexities involved, the proposed approach is designed to be executed in three discrete steps: constructing a product ecology network; identifying potential areas for concentric diversification; and finally, assessing the characteristics of potential areas for concentric diversification. Fig. 1 depicts the overall process of the proposed approach with inputs, methods, and outputs of each step.

3.2.1. Step 1: Construction of a product ecology network

This step constructs a product ecology network that shows the relationships between products based on technological capabilities via three sub-steps. Firstly, association rule mining with conviction indicators is conducted to identify significant directed relationships between products from the patent-product database. The conviction indicator is adopted in this study as an alternative to confidence indicators which cannot capture directions of associations adequately (Brin et al., 1997). As shown in Eq. (1), this indicator compares the probability of X appearing without Y if they were independent of the actual frequency of the appearance of X without Y, where X and Y are the antecedent and consequent of an association rule. That is, the conviction value of 1.2 shows that the rule X → Y is incorrect 20% more often if the association between X and Y were purely random chance. In this respect, it is similar to lift indicators – that measure how many times more often X and Y occur together than would be expected if they were statistically independent – but different in that it also uses the information of the absence of the consequent, thereby being directed (i.e. conviction X → Y ≠ conviction Y → X). Hence, the conviction indicator is considered more appropriate than confidence and lift indicators in concentric diversification research since even though a company owning product X has the possibility to diversify to product Y, it does not guarantee the company which owns product Y will be able to diversify to product X.

$$\text{conviction}(X \rightarrow Y) = \frac{1-P(X \cap Y)}{1-P(Y|X)} \tag{1}$$

Secondly, given a prescribed cut-off value, the association rules between products are transformed to a product relation matrix, where each row and column indicates a product name. Here, if the conviction indicator for products X and Y is greater than the cut-off value, the corresponding element (X, Y) of the product relation matrix has the conviction value for the rule X → Y, otherwise 0.

Finally, a product ecology network is generated, as exemplified in Fig. 2. In the network, nodes and links indicate products and their technological relationships; the size of a node and the thickness of a link are proportional to the frequency of the corresponding product's occurrence in the database and the conviction value for the relevant products, respectively.

3.2.2. Identification of potential areas for concentric diversification

The areas for concentric diversification that can be derived from the product ecology network are likely to be the areas in which many competitors have already entered or shown interests, since the relationships between products are constructed based on *ex post* evaluation. In other words, in order to be practically useful, the areas should be new to a target firm that is currently seeking ideas for concentric diversification. For this reason, this step identifies potential areas for concentric diversification at a product level via link prediction analysis with Katz indicators.

The Katz indicator is one of the similarity-based link prediction indicators which measure the similarity of the pair of nodes based on the paths between them (Katz, 1953). Unlike other similarity-based link prediction indicators (e.g. Adamic-Adar indicator and resource allocation indicator), this indicator can consider the directions of networks, so is considered to be more appropriate for concentric diversification research. As shown in Eq. (2), The Katz indicator sums up all existing paths between two nodes: x and y. Moreover, the contribution of longer paths are penalised during the calculation by a damping factor of β^l, where l is the length of the path and $|paths^{(l)}_{x,y}|$ are all existing paths from node x to y with the length l.

$$\text{Katz}(x,y) = \sum_{l=1}^{\infty} \beta^l * \left|\text{paths}^{(l)}_{x,y}\right| \tag{2}$$

It is proved that the Katz indicator can also be calculated in a matrix format by finding $(I-\beta\mathbf{A})^{-1}-\mathbf{I}$, where $\mathbf{I}$ is an identity matrix and $\mathbf{A}$ is the adjacency matrix (i.e. product relation matrix) of the network.

Considering these, first, the Katz indicators are calculated for all the possible combinations of products, with the product relation matrix

Table 2
Structure of the integrated patent-product database.

Patent	Products	Patent information			
		Publication year	Assignee	…	Classes
PN_1	$Prod_{1,1}$; $Prod_{1,2}$;…; $Prod^{1}_{1,p}$	PY_1	PA_1	…	$PC_{1,1}$; $PC_{1,2}$;…; $PC^{1}_{1,c}$
PN_2	$Prod_{2,1}$; $Prod_{2,2}$;…; $Prod^{2}_{2,p}$	PY_2	PA_2	…	$PC_{2,1}$; $PC_{2,2}$;…; $PC^{2}_{2,c}$
PN_3	$Prod_{3,1}$; $Prod_{3,2}$;…; $Prod^{3}_{3,p}$	PY_3	PA_3	…	$PC_{3,1}$; $PC_{3,2}$;…; $PC^{3}_{3,c}$
PN_4	$Prod_{4,1}$; $Prod_{4,2}$;…; $Prod^{4}_{4,p}$	PY_4	PA_4	…	$PC_{4,1}$; $PC_{4,2}$;…; $PC^{4}_{4,c}$
PN_5	$Prod_{5,1}$; $Prod_{5,2}$;…; $Prod^{5}_{5,p}$	PY_5	PA_5	…	$PC_{5,1}$; $PC_{5,2}$;…; $PC^{5}_{5,c}$
…		…	…	…	…
PN_{n-4}	$Prod_{n-4,1}$; $Prod_{n-4,2}$;…; $Prod^{n-4}_{n-4,p}$	PY_{n-4}	PA_{n-4}	…	$PC_{n-4,1}$; $PC_{n-4,2}$;…; $PC^{n-4}_{n-4,c}$
PN_{n-3}	$Prod_{n-3,1}$; $Prod_{n-3,2}$;…; $Prod^{n-3}_{n-3,p}$	PY_{n-3}	PA_{n-3}	…	$PC_{n-3,1}$; $PC_{n-3,2}$;…; $PC^{n-3}_{n-3,c}$
PN_{n-2}	$Prod_{n-2,1}$; $Prod_{n-2,2}$;…; $Prod^{n-2}_{n-2,p}$	PY_{n-2}	PA_{n-2}	…	$PC_{n-2,1}$; $PC_{n-2,2}$;…; $PC^{n-2}_{n-2,c}$
PN_{n-1}	$Prod_{n-1,1}$; $Prod_{n-1,2}$;…; $Prod^{n-1}_{n-1,p}$	PY_{n-1}	PA_{n-1}	…	$PC_{n-1,1}$; $PC_{n-1,2}$;…; $PC^{n-1}_{n-1,c}$
PN_n	$Prod_{n,1}$; $Prod_{n,2}$;…; $Prod^{n}_{n,p}$	PY_n	PA_n	…	$PC_{n,1}$; $PC_{n,2}$;…; $PC^{n}_{n,c}$

Please cite this article as: Kim, H., et al., Concentric diversification based on technological capabilities: Link analysis of products and technologies, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.025

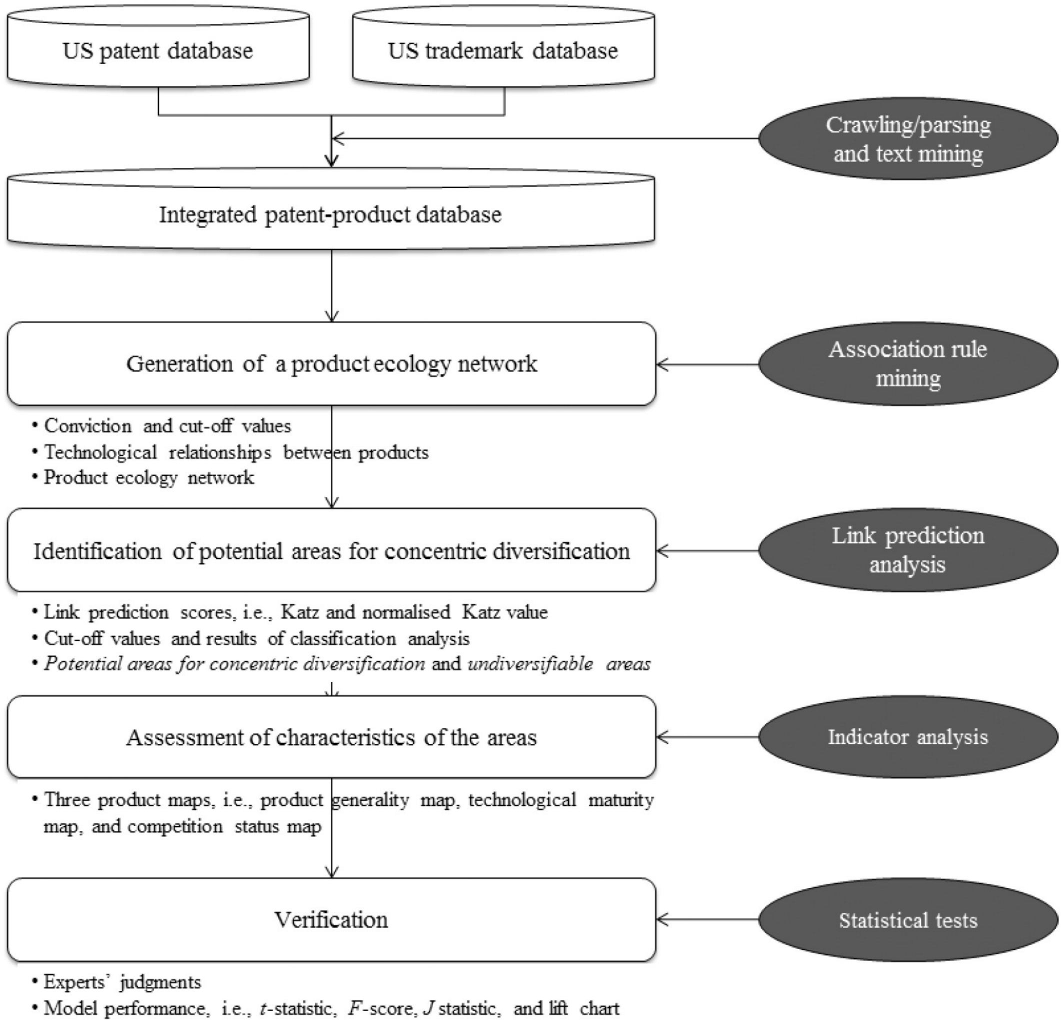

Fig. 1. Overall process of the proposed approach.

created and identity matrix in the same dimension. Second, the elements on the primary diagonal of the resulting matrix are then deleted as they represent links within the same node; which is meaningless in terms of concentric diversification. Third, as shown in Eq. (3), the normalisation of values is conducted for remaining elements so that comparisons of Katz indicators become easier. Here, the normalised Katz values range from 0 to 1. Finally, pairs with the normalised Katz values higher than or equal to a prescribed cut-off are classified as *potential areas for concentric diversification,* while remaining pairs are classified as *undiversifiable areas*. That is, *potential areas for concentric diversification* is the set of pairs that are predicted to occur in the future product ecology network and can be interpreted as possible concentric diversification in the future, whereas *undiversifiable areas* is the set of pairs that are predicted not to occur in the future product ecology network.

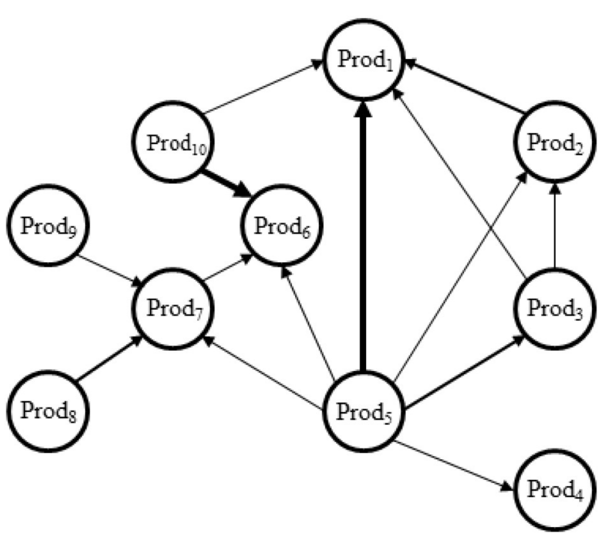

Fig. 2. Example of a product ecology network.

$$\mathrm{Katz}_{\mathrm{normalized}} = \frac{\mathrm{Katz} - \min(\mathrm{Katz})}{\max(\mathrm{Katz}) - \min(\mathrm{Katz})}. \tag{3}$$

3.2.3. Assessment of characteristics of potential areas for concentric diversification

The potential areas for concentric diversification identified in the preceding step require different attention levels across different organisational contexts. It is important for organisations to evaluate and prioritise the potential areas for more detailed investigation. In this respect, patent indicator analysis has been found useful in fostering value creation processes and generating strategic information in a company (Grimaldi et al., 2015).

We develop three indicators describing product generality, technological maturity, and competition status. First, product generality for product *i* (PG_i), measures the number of potential areas that could be further diversified from product *i*, and is calculated by the number of *potential areas for concentric diversification* that product *i* directs to others

Please cite this article as: Kim, H., et al., Concentric diversification based on technological capabilities: Link analysis of products and technologies, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.025

Table 3
Parts of the product relation matrix.

	Semiconductor	Memory cell	Sense amplifier	…	Control circuit	Display	LED
Semiconductor	0	1.952	0	…	1.298	0	0
Memory cell	1.317	0	1.345	…	0	0	0
Sense amplifier	0	1.740	0	…	0	0	0
…	…	…	…	…	…	…	…
Control circuit	3.160	3.983	0	…	0	1.722	0
Display	8.888	0	0	…	0	0	3.318
LED	0	0	0	…	0	1.476	0

obtained in the preceding step. Second, technological maturity for product i (TM_i) evaluates the extent to which technologies for product i are developed, and is computed by counting the number of patents related to product i in the patent-product database. Finally, the competition status for product i (CS_i) shows how fierce the market for product i is, and is the number of assignees possessing patents related to product i. The values of these indicators are also normalised in the same manner, as shown in Eq. (3), thereby ranging from 0 to 1.

4. Empirical analysis and results

4.1. Data

We employed a total of 850,676 patents and 328,288 products in the integrated patent-product database from 2010 to 2014, considering the rates of development of technologies and their relevant products. The data was also divided into two categories by time periods (2010 to 2012 and 2010 to 2014). Specifically, the data from 2010 to 2012 was used to construct the product ecology network and identify potential areas for concentric diversification, whereas the whole data from 2010 to 2014 was used to assess the characteristics of the potential areas identified and assess the model performance.

4.2. Concentric diversification based on technological capabilities

4.2.1. Construction of a product ecology network

We constructed a product ecology network to represent the directed relationships between products based on technological capabilities. Significant relationships between products were first identified from the patent-product database via association rule mining with conviction indicators, as reported in Table 3. Specifically, the association rules with conviction values over 1.2 were chosen after frequent product sets were generated by using the Apriori algorithm. Here, the value of conviction indicators is 1 when the products in antecedent and consequent of the association rule are independent, and shows the values far from 1 when the rule has a logical implication. A total of 13,494 relationships between 6712 products were found with the conviction value of 1.2 (i.e. the rule X → Y is incorrect 20% more often if the association between X and Y were purely random chance). The resulting product relation matrix is a 6713 by 6713 matrix with the product name entities in the first row and column.

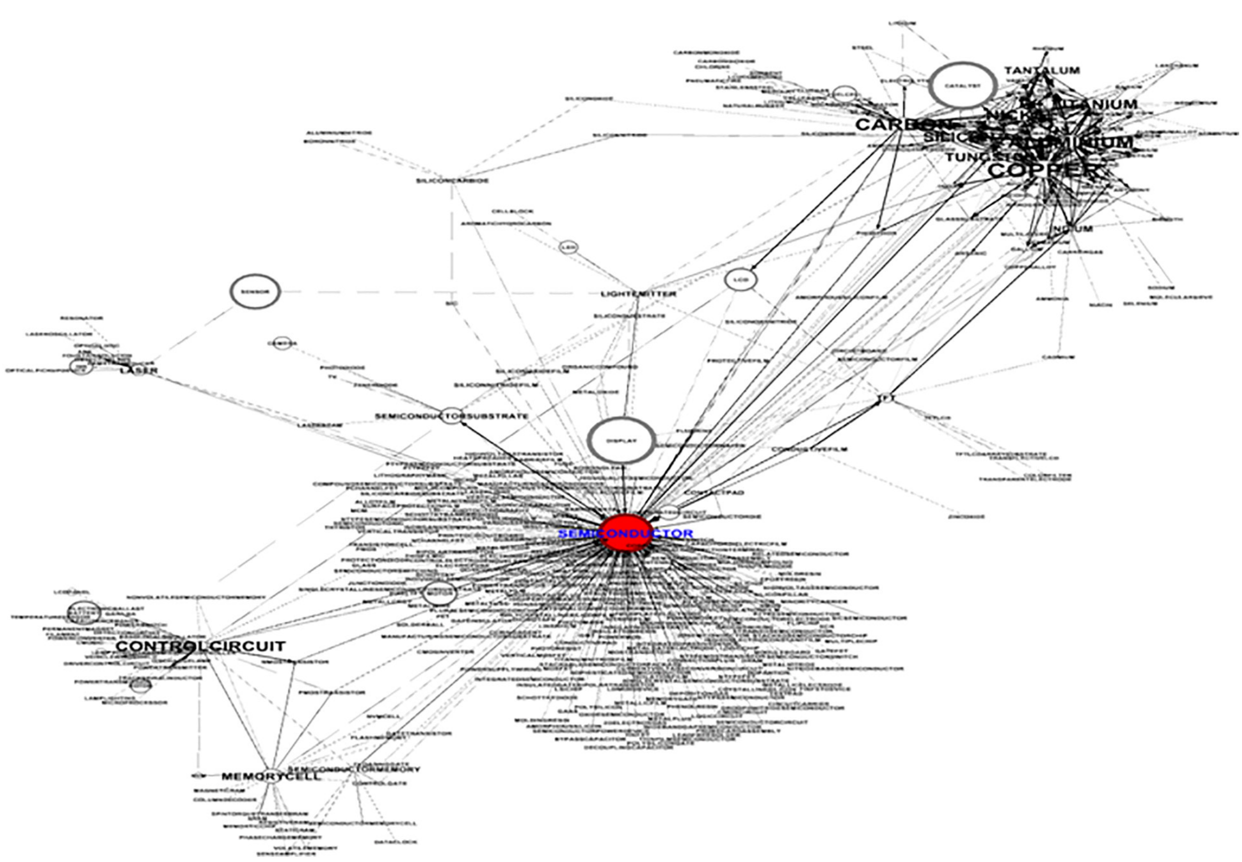

Fig. 3. Product ecology network (2010−2012).

Please cite this article as: Kim, H., et al., Concentric diversification based on technological capabilities: Link analysis of products and technologies, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.025

Table 4
Parts of the results of link prediction and classification analysis.

Source product	Target product	$Katz_{norm}$	Classification	Link in 2010–2014
Data driver	LCD	0.860	PA	O
Memory cell	Semiconductor	0.846	PA	O
Semiconductor	Memory cell	0.836	PA	O
Torque converter	Power train	0.832	PA	O
Computer	Card reader	0.825	PA	X
Display	Keyboard	0.817	PA	X
Computer	Motherboard	0.816	PA	X
Electric motor	Drive wheel	0.812	PA	O
Solar cell	Resin binder	0.811	PA	O
2D image sensor	Optical reader	0.811	PA	O
…	…	…	…	…
Motor	Gear set	0.030	UA	X
Display	Printer	0.029	UA	X
Display	LCD	0.017	UA	O
Wire mesh	Circuit board	0	UA	O

*PA: *Potential areas for concentric diversification*; UA: *Undiversifiable areas*.

The product ecology network was then developed, which are not reported in its entirety owing to lack of space. Fig. 3 presents the parts of the product ecology network that focus on semiconductors. Semiconductors are considered to be a good example for the following reasons: First, this product has played a role of a technology enabler for diverse electronic technologies. Second, semiconductors entail continuous growth but high risks, and thus careful diversification of business is required to reduce the high volatility of the industry. Third, owing to the short life cycles of products related to semiconductors, identifying promising areas for concentric diversification in the future help organisations respond to market changes properly. Finally, semiconductors account for a significant part in our integrated patent-product database, occurring 20,944 times with 16,565 different products.

In the network, the size of a node, the thickness of a link, and the size of a node label indicate the number of occurrences of the corresponding product in the patent-product database, the magnitude of conviction value of the corresponding association rule, and the magnitude of out-degree centrality of the corresponding product, respectively. As there are many applications for semiconductors, many relationships are found in the network; the in-degree and out-degree centrality values for semiconductors are 245 and 17. For example, displays, LCDs, and contact pads have linkages *to* semiconductors, whereas the semiconductor substrate and thin film transistor have linkages *from* semiconductors. The product ecology network constructed in this way was used to identify potential areas for concentric diversification.

4.2.2. Identification of potential areas for concentric diversification

Potential areas for concentric diversification were discovered in this step by using link prediction analysis. With the product relation matrix created and the identity matrix with the same dimensions, Katz

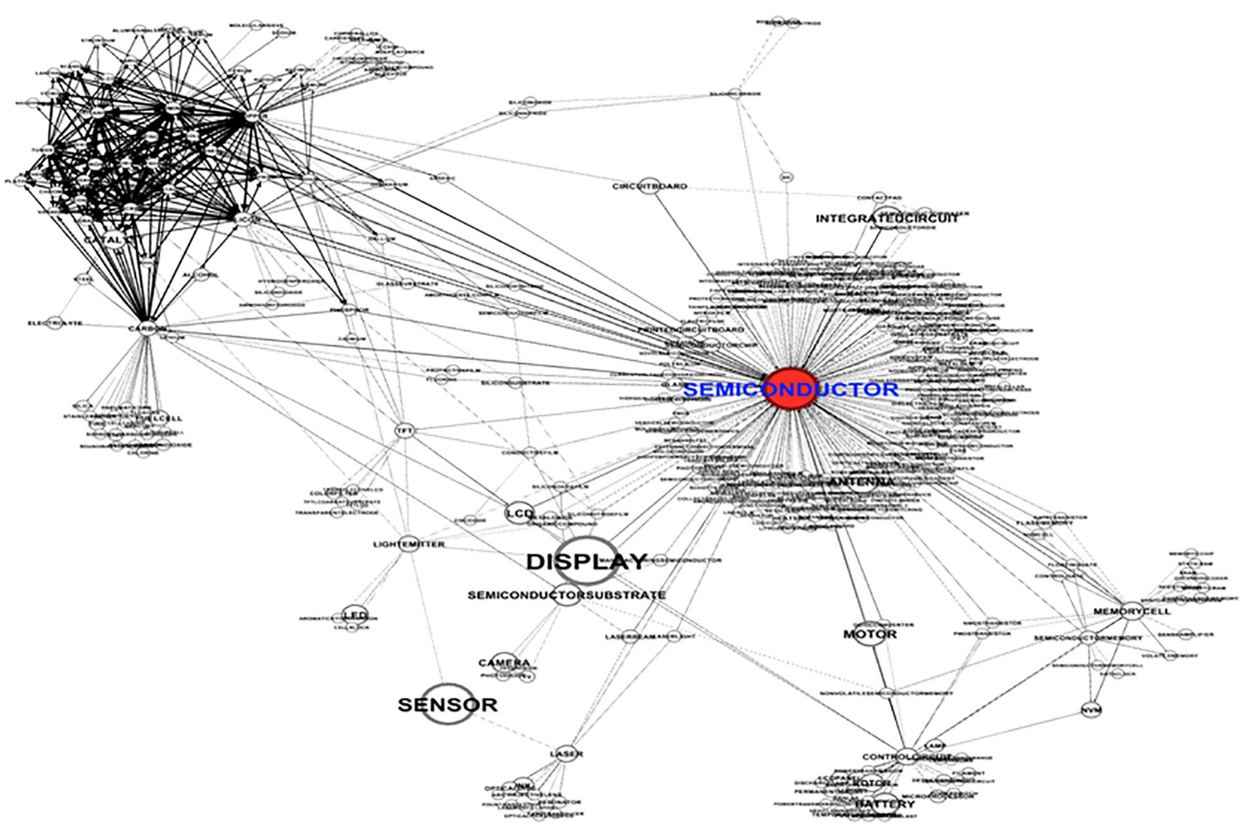

Fig. 4. Product ecology network constructed by link prediction analysis.

Please cite this article as: Kim, H., et al., Concentric diversification based on technological capabilities: Link analysis of products and technologies, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.025

indicators for all the possible links between products in the product ecology network were computed, as stated in Section 3.2.2. The value of 0.005 was used as a damping factor, considering that it should not exceed the reciprocal of the largest eigenvalue of the product relation matrix. The elements in the resulting matrix were normalised to have values between 0 and 1, after the diagonal element of the matrix was excluded. As a result, a total of 45,044,232 pairs of product were assigned with a link prediction score and classified into two categories (i.e. *potential areas for concentric diversification* and *undiversifiable areas*) according to the cut-off value of 0.5. Table 4 reports parts of the results of link prediction and classification analysis.

Fig. 4 shows the product ecology network for semiconductors constructed by link prediction analysis. This product ecology network is different from the product ecology network derived via association rule

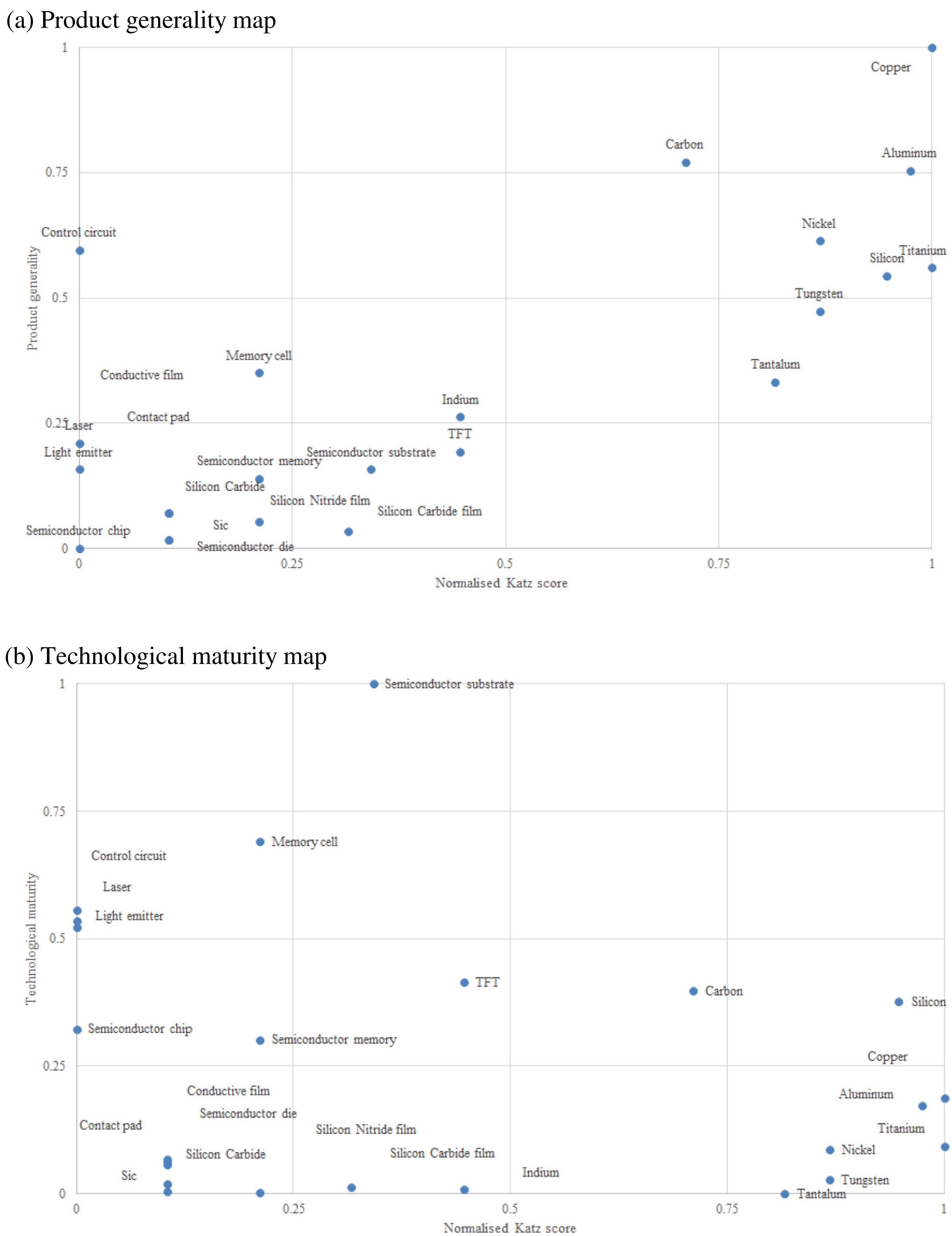

Fig. 5. Product maps. (a) Product generality map. (b) Technological maturity map. (c) Competition status map.

(c) Competition status map

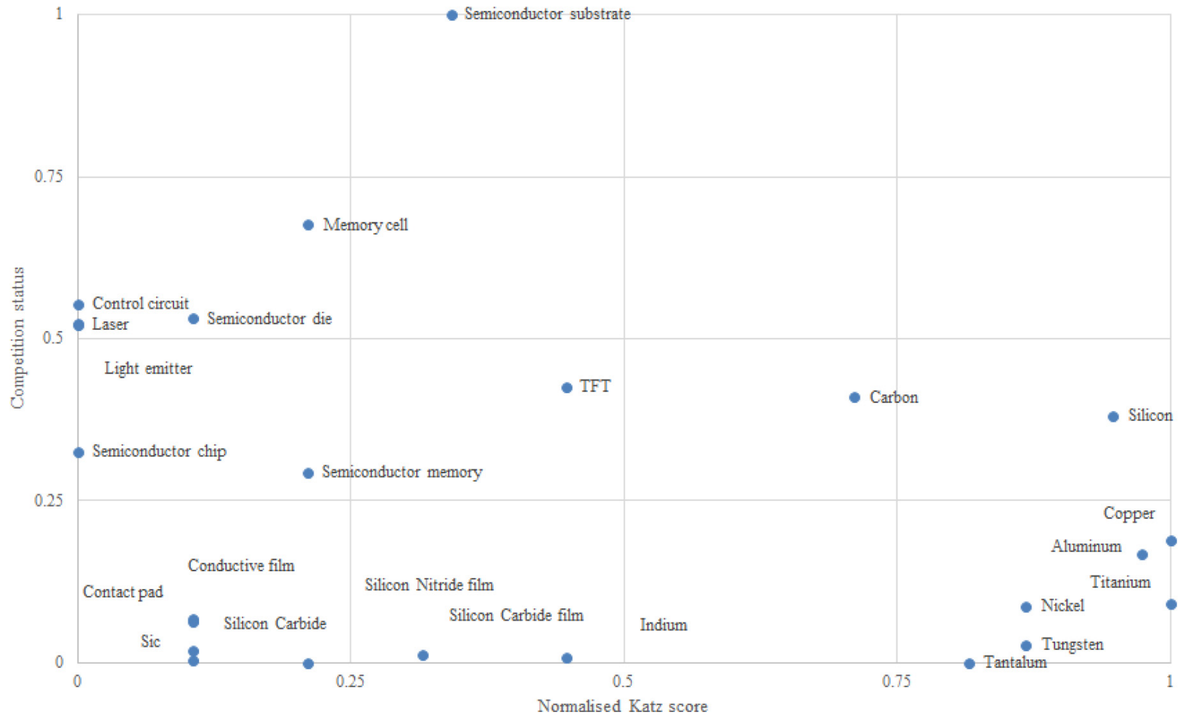

Fig. 5 (*continued*).

mining with conviction indicators (Fig. 3). Such products as silicon carbide film and silicon nitride film are newly added in the network, and the magnitudes of relationships between semiconductor and other products are changed. "Display", near "semiconductor" in the network, also has new relationships with such products as camera and TV. In the case of semiconductor, a total of 24 products – that have strong linkages from semiconductor in the network – were identified as *candidates for concentric diversification*. Along the value chain for semiconductor, such products as nickel, aluminium, and silicon are considered as the areas for upstream diversification, whereas memory cells and semiconductor chips are considered as the areas for downstream diversification. Note that although such products as LCD and displays have linkages to semiconductor, a company manufacturing semiconductors cannot easily diversify to those areas because semiconductor is at the foundation of complex electronic products. Given the technological complexities and similarities, it is easier and more reasonable for the company to diversify into product areas such as semiconductor substrates, semiconductor chips, and thin film transistor, that have linkages from semiconductor in the network.

4.2.3. Assessment of the characteristics of potential areas for concentric diversification

The three indicators – that are, product generality, technological maturity, and competition status – were calculated, as stated in Section 3.2.3, to assess the characteristics of potential areas for concentric diversification. Based on these indicators, we also developed three types of product maps – a product generality map, a technological maturity map, and a competition status map – to provide guidelines for the effective search for business areas for concentric diversification. Fig. 5 shows three product maps for 24 product areas that are likely to be diversified from semiconductor.

First, the product generality map (Fig. 5(a)) utilises the values of the normalised Katz indicators and product generality indicators as the horizontal and vertical axes. In this respect, the product areas can be classified into four categories in the map: *strong chance* & *general areas*, *low chance* & *general areas*, *low chance & specific areas*, and *strong chance & specific areas*. As for products – such as control circuits, copper, and aluminium – identified as *general areas* where further diversification can be conducted to various product areas, companies need to investigate whether the potential areas for further diversification fit with their long-term strategies as well as future prospects and economic profit/loss that the corresponding diversification would bring.

Second, the technological maturity map (Fig. 5(b)) utilises the values of the normalised Katz indicators and technological maturity indicators as the horizontal and vertical axes. In this respect, the product areas can be classified into four categories in the map: *strong chance* & *mature areas*, *low chance* & *mature areas*, *low chance & infancy areas*, and *strong chance & infancy areas*. Although the managerial guidelines may differ across the organisational contexts, the product areas should generally be prioritised in the direction shifting from *strong chance & infancy areas* to *low chance* & *mature areas*. As for products in the *infancy areas*, companies need to investigate if they can have a competitive edge in the area by leading in technology development. In the case of products – such as semiconductor substrate and memory cells – identified as *mature areas*, companies need to consider whether such strategies as M&A and cost reduction are possible.

Finally, the competition status map (Fig. 5(c)) utilises the values of the normalised Katz indicators and competition status indicators as the horizontal and vertical axes. In this respect, the product areas can be classified into four categories in the map: *strong chance* & *turbulent areas*, *low chance* & *turbulent areas*, *low chance & calm areas*, and *strong chance & calm areas*. Similar to the technological maturity map, the product areas should generally be considered in the direction shifting from *strong chance & calm areas* to *low chance* & *turbulent areas*. Products – such as contact pad and conductive film – in the *calm areas*, are generally more preferable to those in *turbulent areas*. As for the products – such as semiconductor substrate and memory cells – identified as *turbulent areas*, companies need to secure differentiated factors from

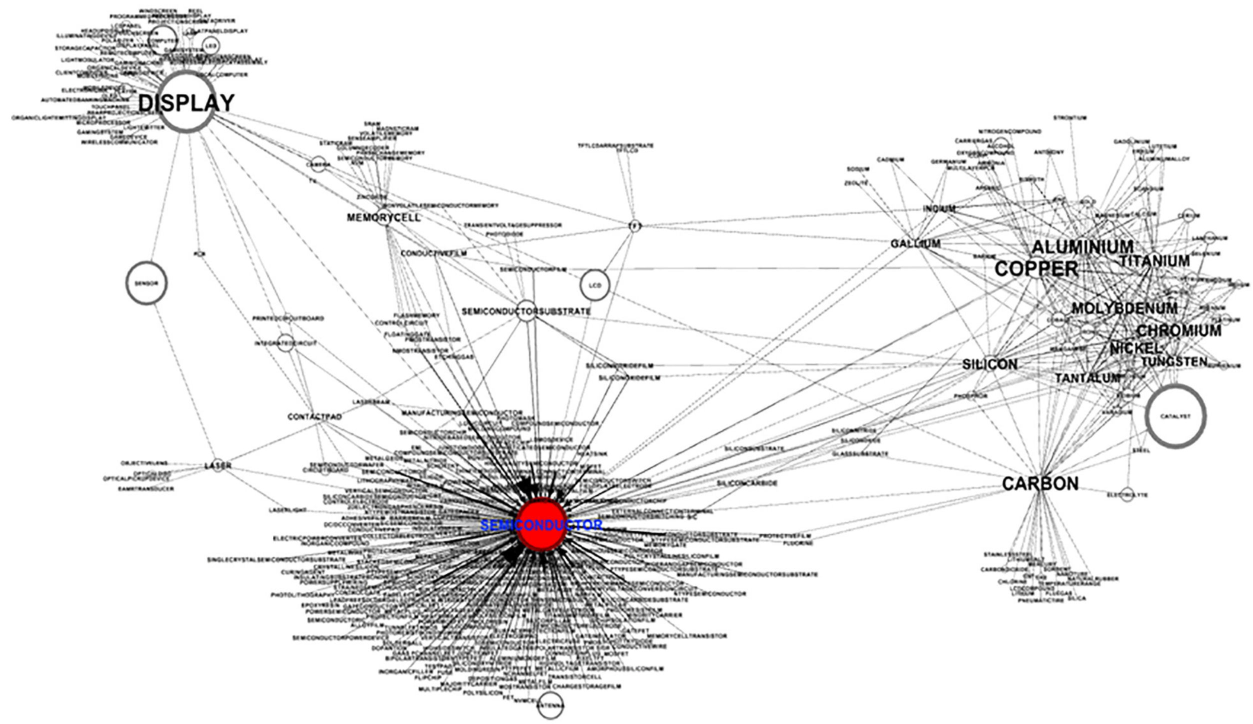

Fig. 6. Product ecology network (2010–2014).

competitors' existing businesses. Also, such strategies as the establishment of strategic alliances with major firms and the discovery of niche markets should be considered.

4.3. Verification

Newly developed methods should be carefully deployed in practice, since there is no confirmation regarding the validity of results. On the one hand, the *candidates for concentric diversification* were deduced indirectly from the integrated patent-product database, so the results could be different from that of experts. For this reason, a panel of experts assessed the informative value of the results. We provide them the list of *candidates for concentric diversification* for semiconductor with an explanation about the analytic process, product ecology networks, link prediction scores, results of classification analysis, and three product maps. The experts verified that most of the areas provided, fit their understanding of technological applicability and that our method is a useful complementary tool for supporting experts' decision making since it enables diverse areas to be analysed swiftly. Moreover, the process of data analysis was found to be more objective and reliable than that of previous methods because the proposed approach uses association rule mining to set-up the cut-off value. Furthermore, the analysis results can support more tangible decision making, as the areas for concentric diversification are provided at a product level rather than a technology level. However, they commented that the performance of the proposed methods can be improved further if other market-related factors are incorporated into the model.

On the other hand, the distinct characteristics of the proposed approach stem from association rule mining with conviction indicators and link prediction analysis, which is interpreted as a binary classification problem. In other words, identifying concentric diversification areas in the product ecology network is equivalent to classifying all the pairs of products into two groups: *potential areas for concentric diversification* and *undiversifiable areas*. For this reason, the results of the link prediction on the product ecology network (2010 − 2012) were compared to the product ecology network (2010–2014) describing the cumulative relationships between products, as shown in Fig. 6. Here, the identical parameters such as the cut-off value for conviction indicators were used for the construction of the product ecology network (2010–2014).

Specifically, we verified the performance of the proposed approach in three ways. Firstly, given that the performance of methods is severely affected by its ability to distinguish between *potential areas for concentric diversification* and *undiversifiable areas*, we conducted the *t*-test to statistically compare the mean values of normalised Katz indicators in two different sets. The first set consists of the pairs of products that exist in the product ecology network (2010–2014), while the second set consists of the pairs of products that do not. Specifically, a two-tailed *t*-test for unequal sample size and unequal variance was carried out; the null hypothesis is $X1 = X2$ while the alternative hypothesis is $X1 \neq X2$, where $\times 1$ and $\times 2$ denote the mean values of normalised Katz indicators for the two sets. As shown in Table 5, it is validated that the means of two sets are not equal, as the *p*-value is below the significance level of 0.05. This shows that our approach can make highly reliable and accurate predictions on potential areas for concentric diversification.

Secondly, given that identifying candidates for concentric diversification corresponds to a binary classification problem, several performance metrics were examined to assess the accuracy of our results after a confusion matrix was constructed, as reported in Table 6(a). Regarding the performance indicators, although the accuracy indicator is a basic metric, but not reliable in our case study since it yields misleading results when the data set is unbalanced. We employed the *F*-measure (Van Rijsbergen, 1979) and Youden's *J* statistic (Youden, 1950) to assess the performance of the prediction results. First, the *F*-measure considers both the precision (positive predictive value) and recall (true positive rate or sensitivity) to measure the accuracy of a test. Here, precision is

Table 5
Summary of *t*-test results.

	t-Statistic	Degree of freedom	Mean	Variance	*p*-Value
Set 1	282.21	10,994	0.7249	0.073	2.2e − 16
Set 2			0.0000929	0.00005719	

Set 1: Pairs of products that exist in the product ecology network (2010–2014).
Set 2: Pairs of products that do not exist in the product ecology network (2010–2014).

Please cite this article as: Kim, H., et al., Concentric diversification based on technological capabilities: Link analysis of products and technologies, Technol. Forecast. Soc. Change (2017), http://dx.doi.org/10.1016/j.techfore.2017.02.025

Table 6
Summary of performance analysis.

(a) Confusion matrix

		Product ecology network (2010–2014)	
		PA	UA
Classification analysis	PA	9651	3844
	UA	1343	45,029,394

(b) *F*-measure and Youden's *J* statistic

Precision	Sensitivity	Specificity	*F*-score	Youden's *J* statistic
0.878	0.715	1.000	0.788	0.878

the number of true positive results divided by the number of all positive results, while recall is the number of true positive results divided by the number of positive results that should have been returned. This measure is defined mathematically as:

$$F\text{—measure} = 2 \times \frac{1}{\frac{1}{recall} + \frac{1}{precision}} = 2 \times \frac{precision \times recall}{precision + recall}. \quad (4)$$

The *F*-measure can be interpreted as a weighted average of the precision and recall, and ranges from 0 to 1. It reaches its best value at 1 and worst at 0. Second, Youden's *J* statistic is a measure that captures the performance of a binary test, and is defined as:

$$J = sensitivity + specificity - 1. \quad (5)$$

Here, specificity is the number of true negative results divided by the number of all negative results. This indicator ranges from −1 to 1, and has a zero value when a test gives the same proportion of positive results for groups regardless of true conditions. A value of 1 indicates that there are no false positives or false negatives, so the test is perfect. The proposed approach shows reliable and significant performance for different cut-off values, as reported in Table 6(b).

Finally, a lift chart was developed to assess the performance of the proposed approach in terms of identifying the most likely potential areas for concentric diversification and comparing the performance to that of a naïve rule, as shown in Fig. 7. For this, we accumulate the correctly classified *potential areas for concentric diversification* records (Y axis) and compare them to the number of total records (X axis). This is done after the link prediction scores are sorted from most likely to least likely potential areas for concentric diversification. The figure shows that the proposed approach outperforms the naïve rule. Taking these results together, the proposed approach is proved to be superior in terms of both accuracy and significance, supporting our contention that it identifies potential areas for concentric diversification.

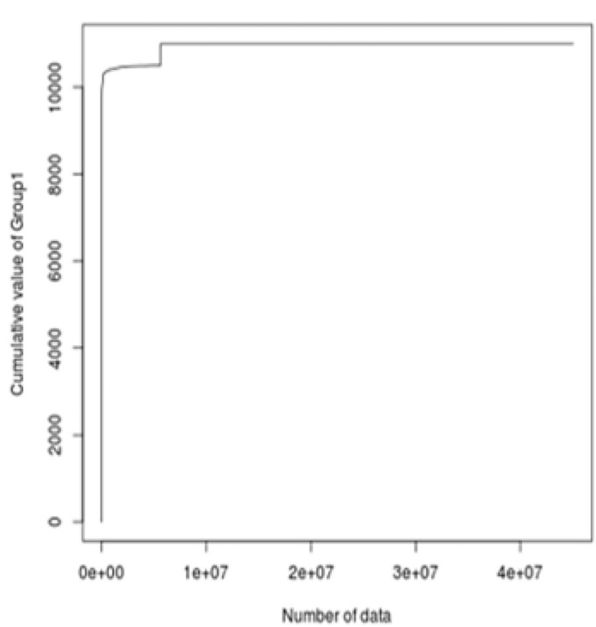

Fig. 7. Lift chart.

5. Conclusions

This study has proposed a systematic approach to identifying potential areas for concentric diversification. The proposed approach is based on the premise that significant technological relationships between products extracted from large-scale quantitative databases can provide valuable information on the feasibility of concentric diversification from one area to another. Our case study shows that the proposed approach enables a wide ranging search for potential areas for concentric diversification and the quick assessment of their characteristics, with analysis results that are statistically significant.

The primary contributions of this research are two-fold. First, from an academic perspective, this study contributes to concentric diversification research by extending previous *ex post* technology-level evaluation to a predictive product-level approach using an integrated patent-product database. To this end, we proposed the systematic approach integrating various methods in an effective way. Specifically, association rule mining with conviction indicators was used to model the directions of diversification and a similarity-based link prediction analysis was used to identify potential areas for concentric diversification in the future. Three quantitative indicators were developed to assess the characteristics of potential areas. Our attempt to adopt diverse methods will provide a basis for future studies in the field of diversification research. Second, from a practical standpoint, the proposed approach enables a wide-ranging search for potential areas for concentric diversification and the quick assessment of their characteristics, with statistically significant results. Furthermore, the proposed product-level approach is of more practical use than previous technology-level approaches, in that it can support more tangible decision making in concentric diversification. We expect the proposed approach and software system could be useful as a complementary tool to support expert decision making, particularly for small and medium sized high-tech companies that are considering entering new technology areas, but which have little domain knowledge.

However, this study is subject to certain limitations, which should be complemented by future research. First, the proposed approach cannot consider non-technological factors affecting diversification strategies. Such non-technological factors as market size, sales networks, human resources, and manufacturing equipment also need to be considered, although this study focused on concentric diversification based on technological capabilities. Second, this study only uses the USPTO database to construct an integrated patent-product database. However, patents are oftentimes purposefully written in the form that comprises enough details for patenting, but does not expose too much for competitors how to circumvent the inventions. For this reason, other types of databases such as Securities and Exchange Commission (SEC) 10-k databases could be employed to improve the accuracy and reliability of the proposed approach. Third, other types of indicators such as product life cycles and technological originality should be developed to diversify the scope of analysis and enhance the richness of potential implications. Finally, the expert assessment has been limited to semiconductor data in our case study, and so the results cannot easily be generalised in other types of products where patenting behaviours may differ, although the results of statistical tests (i.e. *t*-test, *F*-measure, and Youden's *J* statistic) show that our approach can make highly reliable and accurate predictions on potential areas for concentric diversification. The types of products in which the suggested method can best operate should be investigated to confirm the validity of the proposed approach. Nevertheless, the systematic processes and quantitative outcomes of the proposed approach offer a substantial contribution to both current research and future practice.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIP) (No. 2014R1A1A1005931) and supported by the Future Strategic Fund (No. 1.140010.01) of Ulsan National Institute of Science and Technology (UNIST).

References

Agrawal, R., Imielinski, T., Swami, A., 1993. Mining association rules between sets of items in large databases. Sigmod Rec. 22, 207–216.
Ahuja, G., Lampert, C.M., 2001. Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions. Strateg. Manag. J. 22, 521–543.
Ansoff, H.I., 1965. Corporate Strategy: An Analytic Approach to Business Policy for Growth and Expansion. McGraw Hill, New York.
Archibugi, D., Pianta, M., 1992. Specialization and size of technological activities in industrial countries: the analysis of patent data. Res. Policy 21, 79–93.
Breschi, S., Lissoni, F., Malerba, F., 2003. Knowledge-relatedness in firm technological diversification. Res. Policy 32, 69–87.
Brin, S., Motwani, R., Ullman, J., Tsur, S., 1997. Dynamic itemset counting and implication rules for market basket data. Sigmod Rec. 26, 255–264.
Cantwell, J., Piscitello, L., 2000. Accumulating technological competence: its changing impact on corporate diversification and internationalization. Ind. Corp. Chang. 9, 21–51.
Chen, Y., Chang, K., 2012. Using the entropy-based patent measure to explore the influences of related and unrelated technological diversification upon technological competences and firm performance. Scientometrics 90, 825–841.
Chiu, Y.C., Lai, H.C., Lee, T.Y., Liaw, Y.C., 2008. Technological diversification, complementary assets, and performance. Technol. Forecast. Soc. 75, 875–892.
Clauset, A., Moore, C., Newman, M.E., 2008. Hierarchical structure and the prediction of missing links in networks. Nature 453, 98–101.
Collis, D.J., Montgomery, C.A., 1995. Competing on resources: strategy in the 1990s. Harv. Bus. Rev. 73, 118–129.
Creighton, C., Hahash, S., 2003. Mining gene expression database for association rules. Bioinformatics 19, 79–86.
Dutton, G., 1997. What business are we in? Manag. Rev. 86, 54.
Friedman, N., Getoor, L., Koller, D., Pfeffer, A., 1999. Learning probabilistic relational models. IJCAI 99, 1300–1309.
Garcia-Vega, M., 2006. Does technological diversification promote innovation? An empirical analysis for European firms. Res. Policy 35, 230–246.
Getoor, L., Diehl, C.P., 2005. Link mining: a survey. ACM SIGKDD Explor. Newsl. 7, 3–12.
Granstrand, O., 1998. Towards a theory of the technology-based firm. Res. Policy 27, 467–491.
Granstrand, O., Patel, P., Pavitt, K., 1997. Multi-technology corporations: why they have "distributed" rather than "distinctive core" competencies. Calif. Manag. Rev. 39, 8–25.
Grimaldi, M., Cricelli, L., Di Giovanni, M., Rogo, F., 2015. The patent portfolio value analysis: a new framework to leverage patent information for strategic technology planning. Technol. Forecast. Soc. 94, 286–302.
Hamel, G., Prahalad, C.K., 1994. Competing for the Future. Harvard Business Press.
He, C., Loh, H.T., 2010. Pattern-oriented associative rule-based patent classification. Expert Syst. Appl. 37, 2359–2404.
Jaffe, A., 1986. Spillover of R&D: evidence from firm's patents, profits and market value. Am. Econ. Rev. 76, 984–1001.
Katz, L., 1953. A new status index derived from sociometric analysis. Psychometrika 18, 39–43.
Kim, C., Lee, H., Seol, H., Lee, C., 2011. Identifying core technologies based on technological cross-impacts: an association rule mining(ARM) and analytic network process(ANP) approach. Expert Syst. Appl. 38, 12559–12564.
Korea Institute of Science and Technology Information (KISTI), 2012T. Design and Construction of a Knowledgebase for Technology Opportunity Discovery. KISTI.
Lee, S., Yoon, B., Lee, C., Park, J., 2009. Business planning based on technological capabilities: patent analysis for technology-driven roadmapping. Technol. Forecast. Soc. 76, 769–786.
Lee, C., Song, B., Park, Y., 2013. How to assess patent infringement risks: a semantic patent claim analysis using dependency relationships. Technol. Anal. Strateg. 25, 23–38.
Lee, C., Kang, B., Shin, J., 2015. Novelty-focused patent mapping for technology opportunity analysis. Technol. Forecast. Soc. 90, 355–365.
Lei, C., Ruan, J., 2013. A novel link prediction algorithm for reconstructing protein–protein interaction networks by topological similarity. Bioinformatics 29, 355–364.
Liao, S., Chen, Y., 2004. Mining customer knowledge for electronic catalog marketing. Expert Syst. Appl. 27, 521–532.
Lü, L., Zhou, T., 2011. Link prediction in complex networks: a survey. Physica A 390, 1150–1170.
Markides, C.C., 1997. To diversify or not to diversify. Harv. Bus. Rev. 75, 93–100.
Miller, D., 2004. Firms' technological resources and the performance effects of diversification: a longitudinal study. Strateg. Manag. J. 25, 1097–1119.
Miller, D., 2006. Technological diversity, related diversification, and firm performance. Strateg. Manag. J. 27, 601–619.
Narin, F., 1994. Patent bibliometrics. Scientometrics 30, 147–155.
Nasiriyar, M., Nesta, L., Dibiaggio, L., 2013. The moderating role of the complementary nature of technological resources in the diversification–performance relationship. Ind. Corp. Chang. http://dx.doi.org/10.1093/icc/dtt058.
Nesta, L., Saviotti, L.P., 2005. Coherence of the knowledge base and the firm's innovative performance: evidence from the US pharmaceutical industry. J. Ind. Econ. 53, 123–142.
Oyama, T., Kitano, K., Satou, K., Ito, T., 2002. Extraction of knowledge on protein-protein interaction by association rule discovery. Bioinformatics 18, 705–714.
Palich, L., Cardinal, L., Miller, C., 2000. Curvilinearity in the diversification performance linkage: and examination of over three decades of research. Strateg. Manag. J. 21, 155–174.
Patel, P., Pavitt, K., 1997. The technological competencies of the world's largest firms: complex and path-dependent, but not much variety. Res. Policy 26, 141–156.
Quintana-Garcia, C., Benavides-Velasco, C., 2008. Innovative competence, exploration and exploitation: the influence of technological diversification. Res. Policy 37, 492–507.
Rong, J., Vu, H.Q., Law, R., 2012. A behavioral analysis of web sharers and browsers in Hong Kong using targeted association rule mining. Tour. Manag. 33, 731–740.
Rumelt, R., 1974. Strategy, Structure and Economic Performance. Harvard University Press.
Schafer, J.B., Konstan, J.A., Riedl, J., 2001. E-commerce recommendation applications. Data Min. Knowl. Disc. 5, 115–153.
Seol, H., Lee, S., Kim, C., 2011. Identifying new business areas using patent information: a DEA and text mining approach. Expert Syst. Appl. 38, 2933–2941.
Shih, M.J., Liu, D.R., Hsu, M.L., 2010. Discovering competitive intelligence by mining changes in patent trends. Expert Syst. Appl. 37, 2882–2890.
Shin, J., Coh, B.Y., Lee, C., 2013. Robust future-oriented technology portfolios: Black–Litterman approach. R&D Manag. 43, 409–419.
Silverman, B.S., 1999. Technological resources and the direction of corporate diversification: toward an integration of the resource-based view and transaction cost economics. Manag. Sci. 45, 1109–1124.
Stumpf, M.P., Thorne, T., de Silva, E., Stewart, R., An, H.J., Lappe, M., Wiuf, C., 2008. Estimating the size of the human interactome. Proc. Natl. Acad. Sci. U. S. A. 105, 6959–6964.
Valvano, S., Vannoni, D., 2003. Diversification strategies and corporate coherence evidence from Italian leading firms. Rev. Ind. Organ. 23, 25–41.
Van Rijsbergen, C.J., 1979. Information Retrieval. Butterworth, London.
Wang, M.Y., 2012. Exploring potential R&D collaborators with complementary technologies: the case of biosensors. Technol. Forecast. Soc. 79, 862–874.
Wang, P., Xu, B., Wu, Y., Zhou, X., 2015. Link prediction in social networks: the state-of-the-art. SCIENCE CHINA Inf. Sci. 58, 1–38.
Wright, A., Chen, E.S., Maloney, F.L., 2010. An automated technique for identifying associations between medications, laboratory results and problems. J. Biomed. Inform. 43, 891–901.
Yoon, B., Park, Y., 2004. A text-mining-based patent network: analytical tool for high-technology trend. J. High Technol. Manag. Res. 15, 37–50.
Yoon, B., Park, Y., 2005. A systematic approach for identifying technology opportunities: keyword-based morphology analysis. Technol. Forecast. Soc. 72, 145–160.
Yoon, B., Park, Y., 2007. Development of new technology forecasting algorithm: hybrid approach for morphology analysis and conjoint analysis of patent information. IEEE Trans. Eng. Manag. 54, 588–599.
Yoon, B., Park, I., Coh, B.Y., 2014. Exploring technological opportunities by linking technology and products: application of morphology analysis and text mining. Technol. Forecast. Soc. 86, 287–303.
Yoon, J., Park, H., Seo, W., Lee, J.M., Coh, B.Y., Kim, J., 2015. Technology opportunity discovery (TOD) from existing technologies and products: a function-based TOD framework. Technol. Forecast. Soc. 100, 153–167.
Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3, 32–35.
Zook, C., Allen, J., 2001. Profit From the Core: Growth Strategy in an Era of Turbulence. Harvard Business School Publishing, Boston, MA.

Hyunwoo Kim is a PhD student of School of Science and Technology Management Policy at University of Science and Technology. He holds a BS in business administration from Hanyang University and an MS in international hotel management from University of Surrey. His research interests include technology intelligence, technology forecasting, and intellectual property management.

Suckwon Hong is a master's student of School of Management Engineering at Ulsan National Institute of Science and Technology (UNIST). He holds a BS in business administration from UNIST. His research interests include applied data mining and machine learning, technology intelligence, and intellectual property management.

Ohjin Kwon is a director of Centre for Future Information R&D at Korea Institute of Science and Technology Information (KISTI). He obtained a BS and MS in computer science at Kwangwoon University, and a PhD in computer science at University of Seoul. His research areas are information systems, technology intelligence, and patent analysis.

Changyong Lee is an associate professor of the School of Management Engineering at UNIST. He holds a BS in computer science and industrial engineering from Korea Advanced Institute of Science and Technology (KAIST), and a PhD in industrial engineering from Seoul National University. His research interests lie in the areas of applied data mining and machine learning techniques, future-oriented technology analysis, robust technology planning, intellectual property management, and service science.