Invention property-function network analysis of patents: a case of silicon-based thin film solar cells

Janghyeok Yoon · Sungchul Choi · Kwangsoo Kim

Abstract Technology analysis is a process which uses textual analysis to detect trends in technological innovation. Co-word analysis (CWA), a popular method for technology analysis, encompasses (1) defining a set of keyword or key phrase patterns which are represented in technology-dependent terms, (2) generating a network that codifies the relations between occurrences of keywords or key phrases, and (3) identifying specific trends from the network. However, defining the set of keyword or key phrase patterns heavily relies on effort of experts, who may be expensive or unavailable. Furthermore defining keyword or key phrase patterns of new or emerging technology areas may be a difficult task even for experts. To solve the limitation in CWA, this research adopts a property-function based approach. The property is a specific characteristic of a product, and is usually described using adjectives; the function is a useful action of a product, and is usually described using verbs. Properties and functions represent the innovation concepts of a system, so they show innovation directions in a given technology. The proposed methodology automatically extracts properties and functions from patents using natural language processing. Using properties and functions as nodes, and co-occurrences as links, an invention property-function network (IPFN) can be generated. Using social network analysis, the methodology analyzes technological implications of indicators in the IPFN. Therefore, without predefining keyword or key phrase patterns, the methodology assists experts to more concentrate on their knowledge services that identify trends in technological innovation from patents. The methodology is illustrated using a case study of patents related to silicon-based thin film solar cells.

J. Yoon · S. Choi · K. Kim (✉)
Industrial and Management Engineering, Pohang University of Science and Technology,
San 31, Hyoja-dong, Nam-gu, Pohang, Kyungbuk 790-784, Republic of Korea
e-mail: kskim@postech.ac.kr

J. Yoon
e-mail: janghyoon@postech.ac.kr

S. Choi
e-mail: blissray@postech.ac.kr

🖄 Springer

Introduction

Technology analysis is a process which uses textual analysis to detect trends in technological innovation. Methods for technology analysis can be categorized into qualitative versus quantitative. Qualitative methods perform technology analysis using knowledge and skills of experts. However, advances in information and communication technology have resulted in an exponential increase in the amount of technical information, and this increase has made relying only on experts' knowledge to analyze technical information almost impossible. Especially, qualitative methods are time-consuming, costly and inconsistent (Kostoff 1998). To remedy these limitations, quantitative methods such as bibliometrics have been studied.

Co-word analysis (CWA) (Callon et al. 1983) is one of the most widely used methods for quantitative technology analysis. CWA generates a network using co-occurrences of keywords or key phrases related to a technology, and applies social network analysis (SNA) to identify trends from word co-occurrences (Lee and Jeong 2008; Callon et al. 1991). However, to identify occurrences of keywords, CWA should define in advance a set of keyword or key phrase patterns which are represented in technology-dependent terms. This task relies heavily on effort of experts. When analyzing new or emerging technology areas, defining keyword or key phrase patterns of the technology areas may be a difficult task even for experts.

To solve the difficulty in CWA, this research adopts a property-function based approach. The property addresses 'what a system is or has', and expresses a specific characteristic of a system or its sub-systems; the function addresses 'what a system does or undergoes', and expresses a useful action of a system or its sub-systems (Dewulf 2006). The property is described mainly using adjectives, and the function is described mainly using verbs (Dewulf 2006). Therefore information concerning properties and functions can be automatically identified from technical information using natural language processing (NLP) because they can be extracted by grammatical analysis. Properties and functions constitute the abstraction of a system, so they show innovation directions (Dewulf 2006). This means that properties can be related to methods or materials of a system, and that functions can be related to uses or objectives of a system. Because these properties and functions represent key concepts of inventions, they allow analysis of inter-relationships among methods (or materials) and uses (or objectives) of inventions. These properties and functions are the basic unit of the invention property-function network (IPFN) suggested in this research. Organizing and analyzing IPFN, the proposed methodology assists experts to more concentrate on their knowledge services that identify technological trends. Especially, the methodology will facilitate experts to identify trends in technological innovation of new or emerging technology areas because the methodology does not need to predefine keywords or key phrases.

The methodology proposed in this research first presents a procedure for generating IPFN. To identify properties and functions from patents, the procedure extracts the binary relations of 'adjective + noun' or 'verb + noun' forms which are concise but concrete two-

word representations. The procedure introduces NLP to extract these binary relations, taking grammatical relations into considerations. Consider a simple sentence "the transparent window provides visibility". In the sentence, 'transparent window' is a property, and 'provides visibility' is a function. The property 'transparent window' is a method or a material to achieve the function 'provides visibility' that appears in the same sentence, and conversely the function 'provides visibility' is related to a use or an objective of the property 'transparent window'. Each sentence in patents contains properties, functions and their co-occurrences. If we consider properties and functions as nodes, and co-occurrences as links, a property-function network of inventions related to a given technology can be organized.

Second, this paper presents how to analyze the organized IPFN using SNA, which maps and measures relationships and interactions among people, groups, organizations, computers and other connected information/knowledge entities (Hanneman and Riddle 2005). The proposed methodology identifies technological implications of indicators obtained from IPFN such as degree, centrality and density with respect to the property and the function.

Chapter 2 presents an overview of related work. Chapter 3 proposes a procedure to generate the IPFN and analyzes indicators obtained from the IPFN in a technological perspective. Chapter 4 illustrates the proposed methodology using a case study of silicon-based thin film solar cell. Chapter 5 presents conclusions and future work.

Related works

Bibliometrics uses statistical and mathematical methods to identify specific patterns or trends from technical documents (Pritchard 1969). Bibliometrics for technology analysis in general can be categorized into citation analysis or content analysis. Citation analysis uses bibliographic information to track patent citations and identify technology diffusion (Alcacer and Gittelman 2006; Hendrix 2009; Thompson and Fox-Kean 2005; Zhang et al. 2010; Zhang et al. 2009). Content analysis detects technological trends by using text mining which refers to the process of deriving high quality patterns and trends from textual information (Choi et al. 2010). One method for content analysis is CWA (Callon et al. 1991), which identifies co-occurrences of keywords or key phrases, and represents them as a co-occurrence matrix. The co-occurrence matrix represents a network graph, so analysis indicators of the network such as degree, centrality and density are useful for identifying trends and keywords of a technology (Lee and Jeong 2008; Neff and Corley 2009; Yoon and Park 2004; Choi et al. 2010; Chang et al. 2010). However, CWA requires that a set of technology-dependent patterns such as keywords or key phrases be provided in advance by effort of experts, who may be expensive or unavailable.

A property-function based approach can be used to solve the difficulty in CWA. Dewulf (2006) proposed a property-function based approach to identify connection among products, processes and systems in different domains. With analysis of about 16,000 patents from the US Patent and Trademark Office (USPTO), the research concluded that properties of a product is described mainly using adjectives and functions of a product is described mainly using verbs. Dewulf introduced the metaphor "product DNA" to describe the set of properties of a system. Based on the concept of product DNA, related domains or products that act as a source for knowledge transfer could be identified for directed innovation. Verhaegen et al. (2009) extended this concept to extract properties and functions automatically using grammatical analysis, related them to evolutionary trends of technology

(Mann 2002), and predicted directions of further improvements of a product. The research showed that properties and functions can be used for technology analysis.

However, property-function based research does not identify relationships among properties and functions; nor does it determine priorities of properties and functions used in a specific technology area. On the contrary, CWA allows identification of the priorities and relationships by applying SNA indicators such as degree, centrality and density. The methodology proposed in this paper incorporates the advantages of the property-function based approach and CWA to generate and analyze the IPFN, which is a network graph obtained from patents related to a given technology; the IPFN represents relationships among properties and functions.

Invention property-function network (IPFN) analysis methodology

The proposed methodology extracts properties in the form of 'adjective + noun' pairs, and functions in the form of 'verb + noun' pairs from sentences of patents. The extracted properties and functions can be considered as nodes and their co-occurrences in the same sentence can be considered as links. Therefore a network graph can be organized using these pairs (Fig. 1).

In this chapter, Procedure for generating IPFN subsection describes a procedure to generate an IPFN using NLP, and Analysis of IPFN subsection demonstrates how to use SNA to interpret the indicators from the IPFN in the technological perspective.

Procedure for generating IPFN

The suggested procedure for generating an IPFN consists of collecting patents, extracting properties and functions, organizing a co-occurrence matrix for each patent, and generating an IPFN of the whole patent set (Fig. 2).
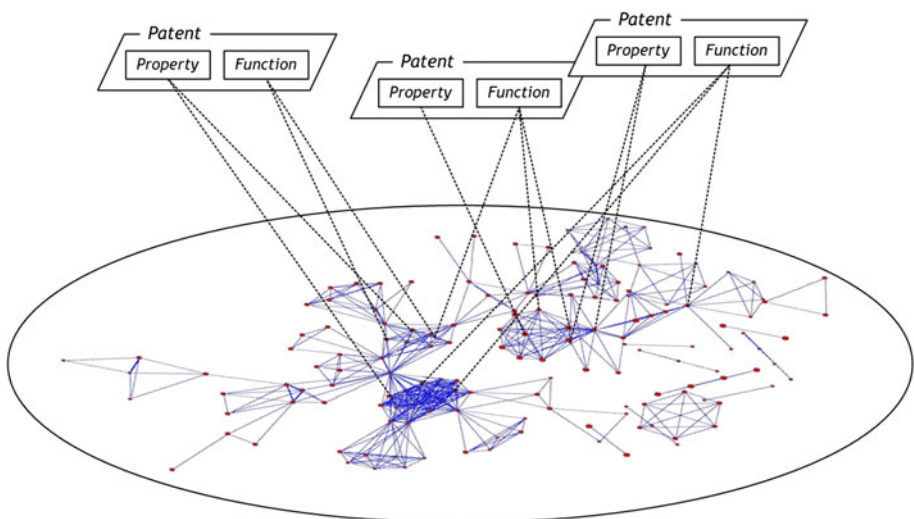

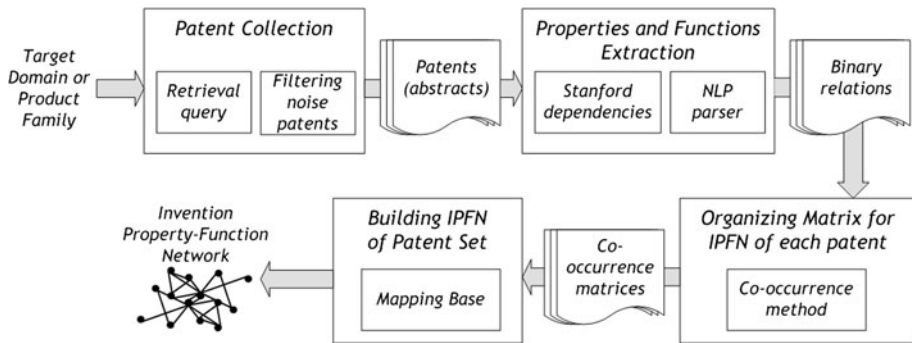**Fig. 1** Invention property-function network

Fig. 2 A procedure for generating an IPFN

Patent collection

The first step in generating an IPFN is to collect patents. Although various types of technical documents exist, including technical reports and papers in science and engineering, only patents are used in this paper because they contain up-to-date and reliable information about inventions. Patents can be easily collected from patent databases. To collect patents, a patent retrieval query is used; it is composed of textual information related to a target technology, and bibliographic information such as international patent code, applicants and application date. Next, a final patent set for analysis is prepared by eliminating irrelevant patents.

Each patent document in patent database is composed of several parts: some provide numerical information including patent number and application date; some contain specific pieces of text information such as inventors, country and applicants; and some are narratives under five headings: title, abstract, background summary, detailed description, and claims. Among these various sections, the narrative sections can be used to extract binary relations related to properties and functions. Tong et al. (2006) and Verhaegen et al. (2009) indicated the importance of including titles and abstract in the automatic analysis of patents. Chen et al. (2003) regarded the human-generated abstracts of patent documents as the most important part. On the basis of these studies, this research uses only abstracts to extract binary relations. Another reason for using abstracts is that they are available in English in most patents (Verhaegen et al. 2009).

Properties and functions extraction

The second step is to extract binary relations related to properties and functions from abstracts. Dewulf (2006) and Verhaegen et al. (2009) identified properties from patents using adjectives. Hirtz et al. (2001) used 'verb + noun' pairs to defined functions of mechanical systems. In this research, 'adjective + noun' pairs identify properties and 'verb + noun' pairs identify functions.

To automate gathering of properties and functions, the Stanford typed dependencies representation and the Stanford parser are used. The 'Stanford typed dependencies representation for English' was designed to provide a simple description of the grammatical relationships in a sentence that can be easily understood and effectively used by people who have no linguistic expertise but who want to extract textual relations (de Marneffe and Manning 2008b). Currently Stanford NLP Group defines 55 grammatical relations which

An example sentence:
"Bell, based in Los Angeles, makes and distributes electronic, computer and building Products."

nsubj(makes-8, Bell-1)
nsubj(distributes-10, Bell-1)
partmod(Bell-1, based-3)
nn(Angeles-6, Los-5)
prep_in(based-3, Angeles-6)
conj_and(makes-8, distributes-10)
amod(products-16, electronic-11)
conj_and(electronic-11, computer-13)
amod(products-16, computer-13)
conj_and(electronic-11, building-15)
amod(products-16, building-15)
dobj(makes-8, products-16)
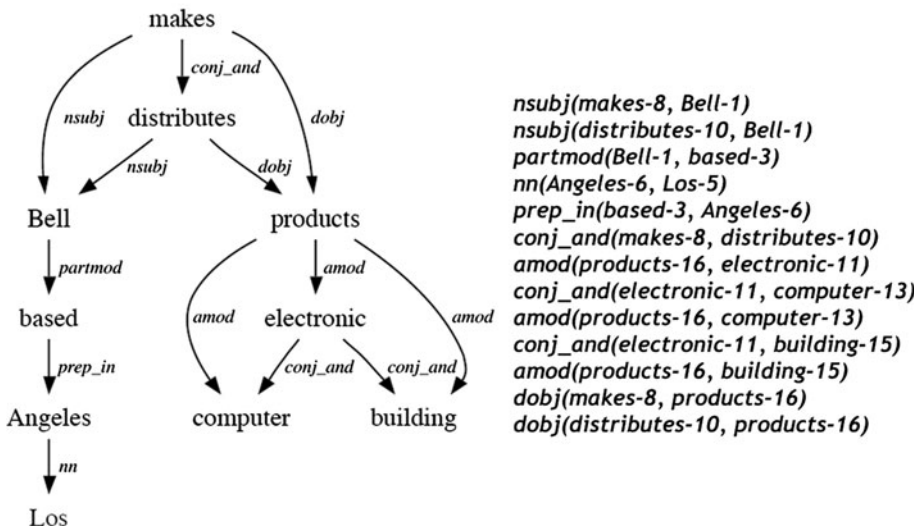dobj(distributes-10, products-16)

Fig. 3 An example of Stanford typed dependencies representation

are all binary: in each, a grammatical relation holds between a governor and a dependent (de Marneffe and Manning 2008a). Any grammatically-correct sentence can be represented using the Stanford typed dependencies (Fig. 3); thus, specific binary relations can be extracted from the sentence.

Among the 55 Stanford typed dependencies, five Stanford typed dependencies grammatically related to the forms 'adjective + noun' or 'verb + noun' are adjectival modifiers (amod), direct objects (dobj), infinitival modifier (infmod), participial modifiers (partmod) and relative clause modifiers (rcmod) (Table 1). The Stanford parser (Stanford 2010), an NLP parser based on the research of Klein and Manning (2003), can be used to automatically extract the binary relations while considering grammatical relationships.

After extracting binary relations that appear together in each sentence of patents using the Stanford parser, unintended or too-general binary relations are filtered out using English stopwords (STOPWORDS 2010). For example, the binary relations having 'comprise', 'invention', apparatus', 'have' and 'make' are removed.

Organizing co-occurrence matrix for each patent

The third step is to organize co-occurrence matrices using binary relations extracted from sentences in each patent. The co-occurrence method is used because the binary relations that appear in the same sentence are related to each other. In the view of network theory, binary relations can be considered as nodes and their co-occurrences can be considered as links. Because the co-occurrences do not imply directions and their values are affected by co-occurrence frequency, the co-occurrence matrix of a patent is an undirected valued network graph. Consider patent P: sentence A has a set of binary relations ⟨v, w, x, y⟩, sentence B has ⟨x, y, z⟩, and sentence C has ⟨t, u, w⟩. Frequency of co-occurrence between

Table 1 The five stanford typed dependencies used in this research

| Stanford dependency(SD)	Description and example: SD name(dependent, governor)
Amod (adjectival modifier)	An adjectival modifier of a noun phrase (NP) is any adjectival phrase that serves to modify the meaning of the NP.
	"Sam eats red meat" amod(meat, red)
Dobj (direct object)	The direct object of a verb phrase (VP) is the noun phrase which is the (accusative) object of the verb; the direct object of a clause is the direct of the VP which is the predicate of that clause.
	"She gave me a raise" dobj(gave, raise)
Infmod (infinitival modifier)	An infinitival modifier of an NP is an infinitive that serves to modify the meaning of the NP.
	"I don't have anything to say" infmod(anything, say)
Partmod (participial modifier)	A participial modifier of an NP or VP is a participial verb form that serves to modify the meaning of the NP or VP.
	"Truffles picked during the spring are tasty" partmod(truffles, picked)
Rcmod (relative clause modifier)	A relative clause modifier of an NP is a relative clause modifying the NP. The relation points from the head noun of the NP to the head of the relative clause, normally a verb.
	"I say the man you love" rcmod(man, love)
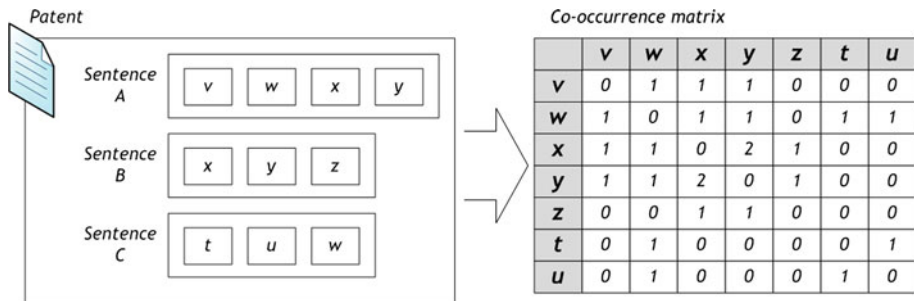
Fig. 4 Generating a co-occurrence matrix from a patent (*left* example text, *Right* matrix: each entry represents the number of times in the example text that the values in the row and column occurred in the same sentence.)

x and y is 2, and frequency of other co-occurrences is 0 or 1, so a co-occurrence matrix for the patent P can be generated (Fig. 4).

Building an IPFN of patent set

The fourth step is to build a co-occurrence matrix for the whole patent set; this is accomplished by merging the co-occurrence matrices of all patents. If the semantically identical nodes are merged into a representative node, the identical links can be merged automatically and their value is the number of identical links. To merge the co-occurrence matrices, this research defines a mapping base to recognize identical nodes by grouping the extracted binary relations. Finally, IPFN for the whole set of patents can be organized by an algorithm that uses a mapping base and co-occurrence matrices (Fig. 5).

```
Merge_Matrices(co-occurrences) RETURNS a solution
    IPFN_matrix

    FOR each co-occurrence DO
        FOR each node of co-occurrence DO
            node ← hasNode(node, IPFN_matrix)
            IF node = null THEN
                remove_link(node)
            ENDIF
        ENDFOR
        FOR each link of co-occurrence DO
            addWeight(link, IPFN_matrix)
        ENDFOR
    ENDFOR

    return IPFN_matrix

hasNode(node, IPFN_matrix) RETURNS a solution or failure
    //if node exists, return the existing one.
    FOR each element of IPFN_matrix DO
        //pattern check
        IF isExist(node, element) THEN
            return element
        ENDIF
    ENDFOR
    //if there is not the existing node, add it after pattern checking
    IF pattern_check(node) THEN
        addNode(node, IPFN_matrix)
        return agent
    ENDIF

    return null

addWeight(link1, IPFN_matrix)
    //add 1 to the weight of link in IPFN_matrix
    FOR each link of IPFN_matrix DO
        IF link = link1 THEN
            weight of link ← weight of link + 1
            return;
        ENDIF
    ENDFOR
```

Fig. 5 An algorithm for building IPFN of the whole set of patents

Analysis of IPFN

The IPFN generated in Procedure for generating IPFN subsection is a network graph. SNA is a useful analysis tool to obtain meaningful information from network graphs. Using SNA techniques, this subsection discusses how the indicators obtained from IPFN such as degree, centrality and density can be interpreted in the technological context.

Table 2 Degree analysis of IPFN

Degree	Property	Function
High	Dominant, cost-effective, useful, technologically verified method or material	Dominant and common usage or application

Degree analysis

In graph theory, the degree of a node is defined as the number of links or the sum of values of links incident to the node (Diestel 2005). In the IPFN, the degree of a property or a function is the sum of values of links incident to the property or the function. In a network, degree can be used as a concept to show the activity or influence of a specific actor in a network. An actor with high degree can mobilize people and is important in information diffusion. Because an actor with high degree influences the cohesion of a network, the network may collapse if the actor is eliminated.

Applying this concept of degree to the IPFN, technological implications of the degree of property and function can be formulated (Table 2). If the degree of a property is high, the property is very likely to be a method or a material that can be used for various purposes or applications, or that can be used with various methods or materials. The property has a strong possibility to be technologically verified, so it can be a widely-used method or a dominant material. If a function has high degree, it is very likely to be a function that can be achieved using various methods or materials, or that should be used to achieve other functions.

Centrality analysis

In analysis of a friendship network, an actor who is most popular is located nearest the center in the friendship network. This means that the centrality of a node determines its relative importance (Freeman 1979). This research adopts three centrality measures that are widely-used in network analysis: degree, betweenness and closeness. Nodes that have a large number of incident links have high degree centrality. Nodes that occur on many shortest paths between other nodes have high betweenness centrality. Nodes that tend to have short geodesic distances to other nodes within a network have high closeness.

Applying this centrality measures to IPFN, technological implications of centrality of property and function can be formulated (Table 3). A property with high degree centrality

Table 3 Centrality analysis of IPFN

Centrality	Property	Function
High degree	Widely used or leading methods or materials in a sub-domain related to a product	Usually required or necessary uses or applications in a sub-domain related to a product
High betweenness	Methods or materials applicable to various purposes. methods or materials used with other methods or materials	Uses or applications using various methods or materials. Uses or applications used with other uses or applications
High closeness	Widely-used or leading methods or materials over a product. Technologically verified key methods or materials	Widely-required or necessary uses over a product. Key objectives of the patents related to a product

has a strong possibility of being a widely-used or leading method (or material) in a specific technology to which the property belongs. A function with high degree centrality is strongly related to the prerequisite or necessary uses or objectives in a specific technology to which the function belongs. A property with high betweenness centrality has a strong possibility of being a fresh or novel method that can be used for various aims or that can be used together with various methods. A function with high betweenness centrality is related to a common objective that can be achieved by various methods or materials. A property with high closeness centrality has a strong possibility of being a widely-used or leading method (or material) of a product, and a function with high closeness centrality can be strongly related to a key or necessary use (or objective) of a product.

Density analysis

Degree and centrality are the measures for nodes, but density is a measure for networks. The density of a network is the proportion of links in a network related to the number of total possible links in the network. Density indicates how closely all nodes in the network are related to each other. The completeness of a network can be identified using density. If the density of a network is 0, he network does not have links; if the density of a network is 1, all nodes of the network are connected all others.

Applying this density to an IPFN, the density of sub-networks of the IPFN can be used to identify technological implications (Table 4). Using methods such as bi-component (Hopcroft and Tarjan 1973) and K-core (Bollobas 1983), an IPFN can be partitioned into one or more sub-networks. If a sub-network has large size and high density, the properties and functions in the sub-network are strongly coupled to each other and are strongly related to the frequently-used or dominant methods (or materials) or uses (or objectives). Moreover they have a strong possibility of having been technologically verified. If a sub-network has large size and low density, various methods (or materials) are being developed to achieve various uses (or objectives) but they may not be dominant or technologically verified, but the technology area related to that sub-network has a strong possibility of being a prospective area, in which many inventions actively appear because no dominant designs exist. If a sub-network has small size and high density, properties and functions in the sub-network are strongly related to the methods (or materials) or uses (or objectives) of a specific technology. Therefore they may deliver novel and particular methods (or materials) or uses (or objectives) with respect to a specific technology.

Table 4 Density analysis of sub-networks of IPFN

Sub-network	Property	Function
Large size; high density	Highly coupled to other properties or functions. Frequently used, dominant, technologically verified methods or materials	Highly connected to other properties or functions. Frequently used, dominant, technologically verified uses or applications
Large size; low density	Prospective methods or materials, but not yet dominant and technologically verified	Prospective uses or applications, but not yet dominant
Small size; high density	Novel and fresh methods or materials for a specific sub-domain	Novel and fresh uses or applications for a specific sub-domain

Case study

This section illustrates the proposed procedure using a case study of silicon-based thin film solar cells. Solar energy is believed to be the most promising renewable energy source because it is unlimited and unpolluting. A solar cell is a device that exploits the photovoltaic effect to convert sunlight directly into electricity. Silicon-based thin film solar cells are among the most advanced types of solar cells. This research collects US and non-US inventors' patents from USPTO using a patent retrieval query composed of keywords related to silicon-based thin film solar cells and date conditions referring to patents issued or applied since 2006. In the query, patent classification codes are not used to search patents regardless of patent categories. After eliminating irrelevant patents, 50 patents are used to build and analyze an IPFN of silicon-based thin film solar cells (Table 5).

According to the procedure for generating an IPFN (Procedure for generating IPFN subsection), the Stanford parser was used to extract binary relations that appear in the same sentence with respect to patents, then binary relations that contain the English stopwords were eliminated, and co-occurrence matrices of all patents were constructed. To merge the co-occurrence matrices, we defined a mapping base file to recognize identical nodes by

Table 5 Patent retrieval query

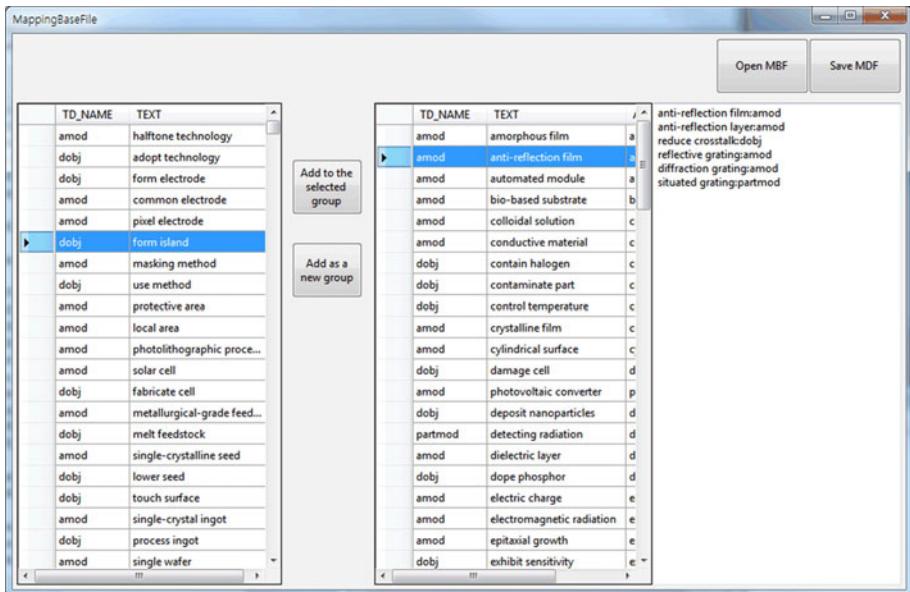| Patent retrieval query	Collected patents
((photovol* solarcell* ((solar* photo*) NEAR2 (vol* cell* device* semiconductor*))) and (silicon* thin* amorphous* microcrystal* nanocrystal* tandem*) and (hwcvd* not hot wire cvd* not ((hot* hw* plasma* vacuum*) NEAR3 (cvd* deposit*)))) and (@RD>=20060101 @AD>=20060101)	Total 50 patents (after filtering irrelevant patents)

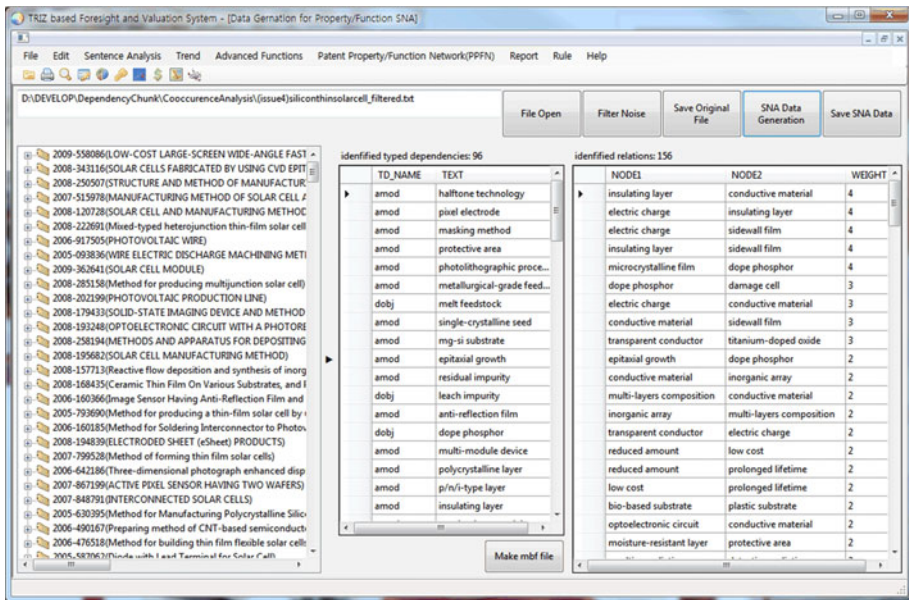**Fig. 6** Defining a mapping base file

Fig. 7 Nodes and links identified from IPFN

grouping the extracted binary relations (Fig. 6). Nodes and links were merged using the mapping base file; a co-occurrence matrix of the 50 patents was generated; this matrix consists of 96 nodes and 156 links (Fig. 7). An SNA tool, NetMiner 3[TM], was used to generate an IPFN related to the silicon-based thin film solar cell and to perform degree, centrality and density of the IPFN (Fig. 8).

First, the top 20 nodes with high degree were identified (Table 6). High-ranked properties and functions were strongly related to the dominant methods or objectives of silicon-based thin film solar cells. Actually, the reason that the technology trend of solar cells changed from bulk silicon solar cells to silicon-based thin film solar cells was the increase in the price of silicon. Therefore, many inventions in the field of silicon-based thin film solar cells have presented methods of forming thin films such as n-layer, i-layer, p-layer, substrate and transparent conductive oxide, methods of constructing multi-layered structure such as tandem cells and triple cells to increase electricity conversion efficiency, methods of increasing light penetration and light trapping, and methods of assembling solar cell modules. These trends are strongly related to the top ranked properties and functions such as 'form film', 'multi-layers composition', 'p-i-n layer', 'anti-reflection film', 'transparent conductor', 'refractive layer' and 'multi-module device'. In this way, degree analysis of nodes allows identification of the key concepts related to a target technology.

Second, closeness (Fig. 9) and betweenness (Fig. 10) of nodes were graphically displayed using NetMiner 3[TM]. Properties or functions with high centrality in IPFN were 'form film', 'microcrystalline film', and 'multi-layered composition'. They actually represent the most dominant and necessary methods or objectives to develop thin film solar cells. In association with the property 'microcrystalline film', microcrystalline silicon manufacture has been invented to overcome amorphous silicon's limitation that its photoconductivity can be significantly reduced by prolonged illumination with intense
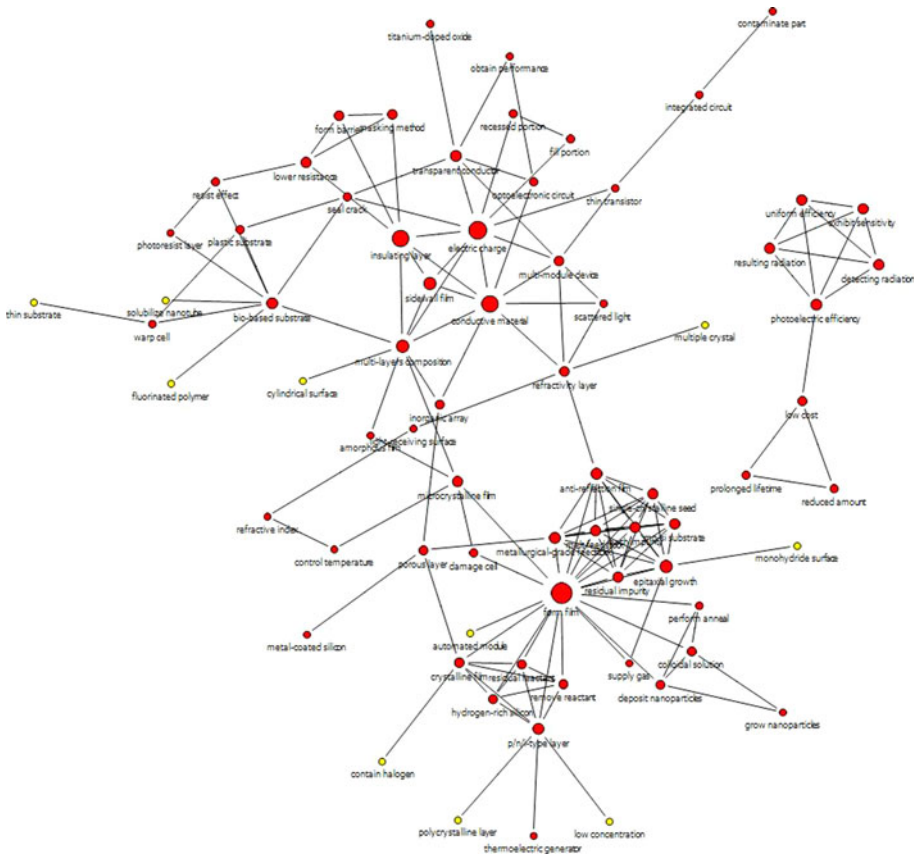
Fig. 8 An IPFN of patents related to silicon-based thin film solar cells

Table 6 Node degrees identified from IPFN

Rank	Property-function	Degree	Rank	Property-function	Degree
1	Form film	20	8	Residual impurity	8
2	Electric charge	10	8	Mg-si substrate	8
2	Epitaxial growth	10	8	Single-crystalline seed	8
4	Multi-layers composition	9	8	Melt feedstock	8
4	Conductive material	9	15	Microcrystalline film	7
4	Anti-reflection film	9	15	Insulating layer	7
4	Metallurgical-grade feedstock	9	17	Multi-module device	6
8	Bio-based substrate	8	17	Refractive layer	6
8	p-i-n-type layer	8	17	Transparent conductor	6
8	Leach impurity	8	20	Photoelectric efficiency	5

light. Properties 'refractivity layer' and 'anti-reflection film' with high centrality indicate that light trapping is a widely-used method in silicon-based thin film solar cells. In fact, light tapping technology for transparent conductive oxides and back reflectors is strongly
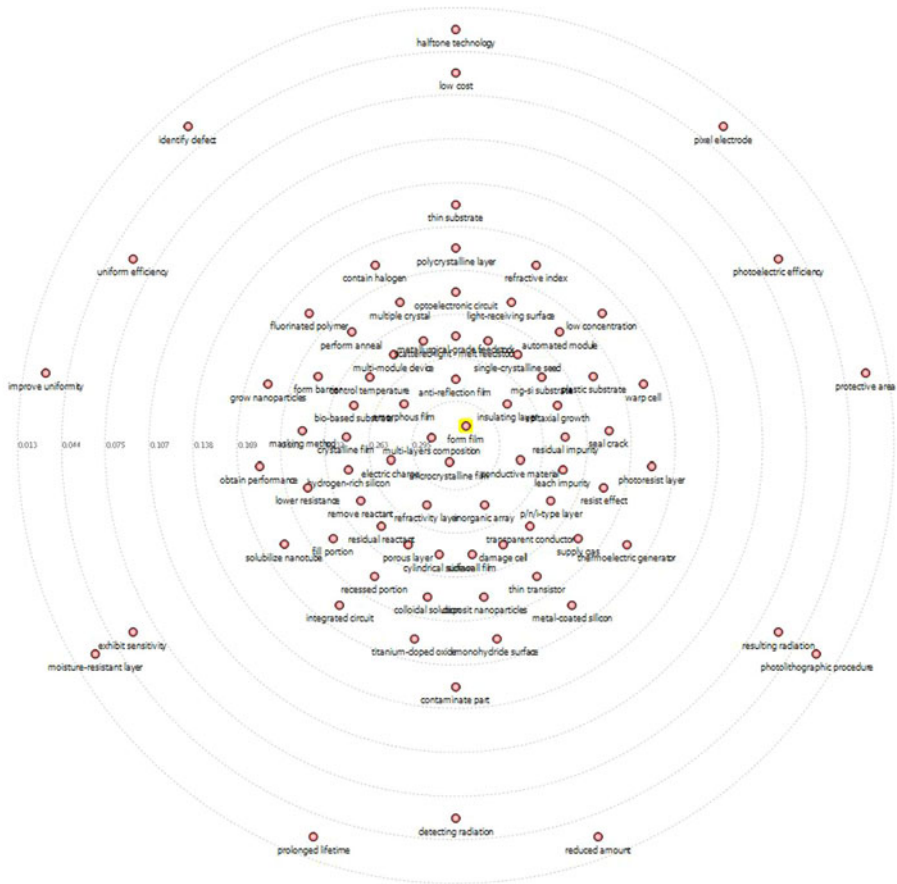
Fig. 9 Closeness centrality of IPFN

related to improvement of electricity conversion efficiency because it minimizes reflection of incident rays and increases the light path within thin-film silicon.

Third, five sub-networks of the IPFN were identified using bi-component method, and their sizes and densities were calculated (Table 7). Technological implications could be identified from sub-networks G2 and G3. Sub-network G2 was composed of 'perform anneal', 'deposit nanoparticles', 'form film', 'colloidal solution', and 'grow nanoparticles'; it has several nodes and high density, so it shows that annealing, colloidal solution, epitaxial growth, and growth of nanoparticles are strongly related to each other in the manufacture of thin film solar cells. The density of sub-network G3 was relatively very low because it had too many nodes. This low density make identifying implications difficult, but major properties with high degree and high centrality including 'crystalline film', 'multi-layers composition' and 'refractive layer' have many links to many properties or functions. Because sub-network G3 has large size and low density, a strong possibility exists that prospective methods (or materials) and their uses are being invented actively in silicon-based thin film solar cells. Actually, even now, many companies and research institutes have been applying for patents about sputtering methods, tandem or triple solar cells, and light trapping.
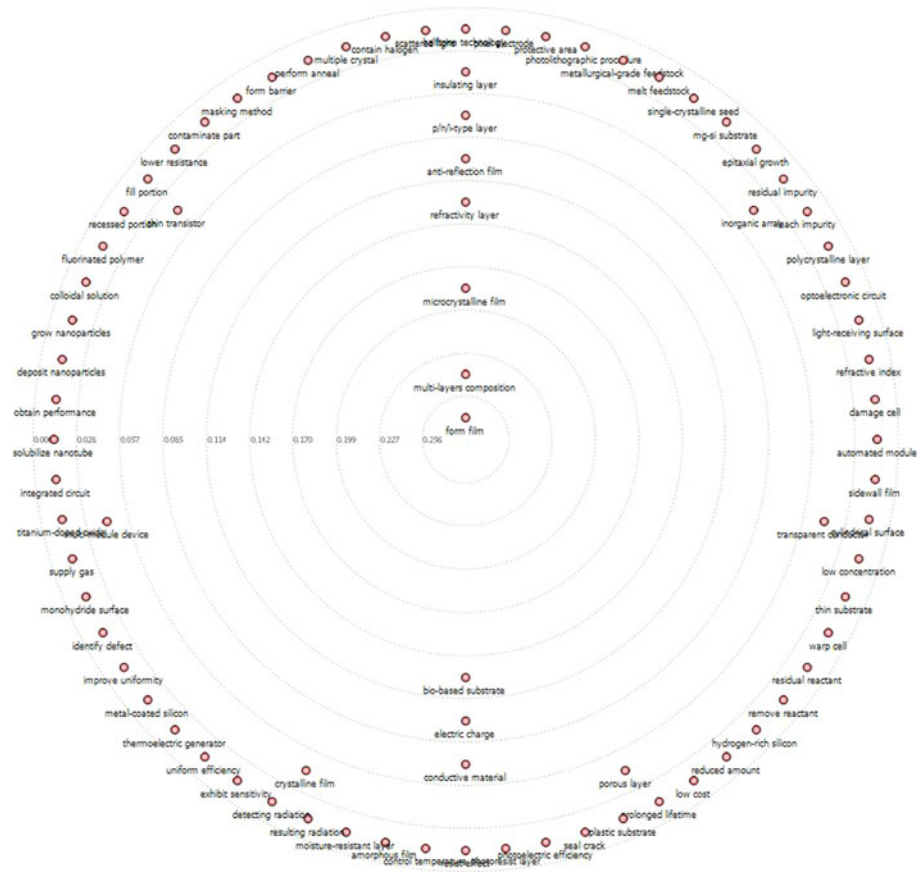
Fig. 10 Betweenness centrality of IPFN

Concluding remarks

CWA requires a set of keyword or key phrase patterns for technology analysis which are represented in technology-dependent terms. However, predefining the set of patterns relies heavily on effort of experts, who may be expensive or unavailable. Furthermore, defining keyword or key phrase patterns of new or emerging technology areas may be a difficult task even for experts. To solve this problem in CWA, this research adopted a property-function based approach. Without predefining keyword or key phrase patterns, the proposed methodology generates an IPFN by using NLP to extract properties and functions from patents, and help experts such as researchers and R&D policy makers to identify technological implications of properties and functions in the IPFN using SNA. In the case study, we generated an IPFN of patents related to silicon-based thin film solar cells and analyzed the technological implications of the IPFN's degree, centrality and density. As an expert support tool, the methodology will help experts to more concentrate on their knowledge services that identify technological trends in a given technology. Furthermore, we expect that the methodology will help identify invention key concepts in new or emerging technology areas during the early stages of research.

Table 7 Density analysis of sub-networks of IPFN

Sub-network	Members	Size	Density
G1	Fill portion, electric charge, recessed portion	3	1
G2	Perform anneal, deposit nanoparticles, form film, colloidal solution, grow nanoparticles	5	0.8
G3	Supply gas, epitaxial growth, form film, crystalline film, remove reactant, residual reactant, sidewall film, multi-layers composition, conductive material, insulating layer, electric charge, thin transistor, multi-module device, scattered light, refractivity layer, amorphous film, microcrystalline film, damage cell, control temperature, refractive index, light-receiving surface, anti-reflection film, transparent conductor, form barrier, lower resistance, masking method, resist effect, bio-based substrate, photoresist layer, plastic substrate, seal crack, warp cell, obtain performance, optoelectronic circuit, inorganic array, porous layer, metallurgical-grade feedstock, p/n/i-type layer, hydrogen-rich silicon, leach impurity, residual impurity, mg-si substrate, single-crystalline seed, melt feedstock	44	0.118
G4	Uniform efficiency, detecting radiation, resulting radiation, photoelectric efficiency, exhibit sensitivity	5	1
G5	Prolonged lifetime, reduced amount, low cost	3	1

This research focused on the analysis of several indicators including degree, centrality and density of the IPFN. Future research will further investigate technological implications of other analysis indicators such as structural holes and strength of ties in SNA. Another future topic will be identification of common or convergent key concepts from IPFNs of related domains.

References

Alcacer, J., & Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *Review of Economics and Statistics, 88*(4), 774–779.

Bollobas, B. (1983). *The evolution of sparse graphs. Graph theory and combinatorics* (pp. 35–57). London: Academic Press.

Callon, M., Courtial, J., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics, 22*(1), 155–205.

Callon, M., Courtial, J., Turner, W., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information, 22*(2), 191.

Chang, P., Wu, C., & Leu, H. (2010). Using patent analyses to monitor the technological trends in an emerging field of technology: A case of carbon nanotube field emission display. *Scientometrics, 82*(1), 5–19.

Chen, L., Tokuda, N., & Adachi, H. (2003). A patent document retrieval system addressing both semantic and syntactic properties. In: *Proceedings of the ACL-2003 workshop on patent corpus processing* (pp. 1–6). Japan: Sapporo.

Choi, S., Lim, J., Yoon, J., & Kim, K. (2010). Patent function network analysis: In: *A function based approach for analyzing patent information*. Egypt, Cairo: IAMOT.

de Marneffe, M., & Manning, C. (2008a). Stanford typed dependencies manual. Technical report.

de Marneffe, M., & Manning, C. (2008b). The stanford typed dependencies representation. In: *Association for computational linguistics* (pp. 1–8). Stanford: Stanford University.

Dewulf, S. (2006). Directed variation: Variation of properties for new or improved function product DNA, a base for 'connect and develop'. Kortrijk, Belgium: ETRIA TRIZ Futures.

Diestel, R. (2005). Graph theory. Graduate texts in mathematics (vol. 173). Berlin: Springer-Verlag.

Freeman, L. (1979). Centrality in social networks conceptual clarification. *Social Networks, 1*(3), 215–239.

Hanneman, R., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside.

Hendrix, D. (2009). Institutional self-citation rates: A three year study of universities in the united states. *Scientometrics, 81*(2), 321–331.

Hirtz, J., Stone, R., McAdams, D., Szykman, S., & Wood, K. (2001). A functional basis for engineering design: Reconciling and evolving previous efforts. *Research in Engineering Design, 13*(2), 65–82.

Hopcroft, J., & Tarjan, R. (1973). Algorithm 447: Efficient algorithms for graph manipulation. *Communications of the ACM, 16*(6), 372–378.

Klein, D., & Manning, C. (2003). Fast exact inference with a factored model for natural language parsing. *Advances in Neural Information Processing Systems, 15*, 3–10.

Kostoff, R. (1998). The use and misuse of citation analysis in research evaluation. *Scientometrics, 43*(1), 27–43.

Lee, B., & Jeong, Y. (2008). Mapping Korea's national r&d domain of robot technology by using the co-word analysis. *Scientometrics, 77*(1), 3–19.

Mann, D. (2002). *Hands-on systematic innovation*. Belgium: Creax press.

Neff, M., & Corley, E. (2009). 35 years and 160, 000 articles: A bibliometric exploration of the evolution of ecology. *Scientometrics, 80*(3), 657–682.

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation, 25*(4), 348–349.

Stanford. (2010). The stanford parser: A statistical parser. http://nlp.stanford.edu/software/lex-parser.shtml.

STOPWORDS. (2010). English stopwords. http://www.ranks.nl/resources/stopwords.html.

Thompson, P., & Fox-Kean, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review, 95*(1), 450–460.

Tong, L., Cong, H., & Lixiang, S. (2006). Automatic classification of patent documents for triz users. *World Patent Information, 28*(1), 6–13.

Verhaegen, P. A., D'Hondt, J., Vertommen, J., Dewulf, S., & Duflou, J. R. (2009). Relating properties and functions from patents to triz trends. *CIRP Journal of Manufacturing Science and Technology, 1*(3), 126–130.

Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research, 15*(1), 37–50.

Zhang, L., Janssens, F., Liang, L., & Glanzel, W. (2010). Journal cross-citation analysis for validation and improvement of journal-based subject classification in bibliometric research. *Scientometrics, 82*(3), 687–706.

Zhang, W., Qian, W., & Ho, Y. (2009). A bibliometric analysis of research related to ocean circulation. *Scientometrics, 80*(2), 305–316.