

ORIGINAL ARTICLE

Open Access



Data-driven Product Functional Configuration: Patent Data and Hypergraph

Wenguang Lin¹, Xiaodong Liu¹ and Renbin Xiao^{2*}

Abstract

The product functional configuration (PFC) is typically used by firms to satisfy the individual requirements of customers and is realized based on market analysis. This study aims to help firms analyze functions and realize functional configurations using patent data. This study first proposes a patent-data-driven PFC method based on a hypergraph network. It then constructs a weighted network model to optimize the combination of product function quantity and object from the perspective of big data, as follows: (1) The functional knowledge contained in the patent is extracted. (2) The functional hypergraph is constructed based on the co-occurrence relationship between patents and applicants. (3) The function and patent weight are calculated from the patent applicant's perspective and patent value. (4) A weight calculation model of the PFC is developed. (5) The weighted frequent subgraph algorithm is used to obtain the optimal function combination list. This method is applied to an innovative design process of a bathroom shower. The results indicate that this method can help firms detach optimal function candidates and develop a multifunctional product.

Keywords: Hypergraph, Functional configuration, Patent analysis, Weighted frequent subgraph algorithm

1 Introduction

To remain relevant in a competitive market, manufacturers typically strengthen product innovation and development capability to satisfy customer demands. Customers are typically attracted to multifunctional products instead of products with a single function [1, 2]. Therefore, firms tend to develop multifunction products (MFPs) that can fulfil the various demands of customers. Effectively designing new products and combining many functions remain important tasks for firms. In this context, the MFP design process is a challenging task as it requires engineering designers to solve the following three problems of how-how-what (2H-W):

- How can functional information be obtained?
- How is the function evaluated?

- Which functions are suitable to be integrated into one product?

A functional configuration is typically realized by experts, where suitable functions are combined to satisfy customer demands. Although it is a frequently used method, the subjectivity inherent in this method relies on the designer's experience and skills. In addition, some methods analyze function demands through questionnaires or surveys, which are time consuming, and the data obtained are limited.

In recent years, owing to the development of big data technology, data-driven methods have been applied in product design, including functional analysis and decision support. The quality and reliability of data-driven design results are affected by the data source. Data sources have been expanded from a single type to various data sources, such as website reviews, machine data, physiological data, and patent data. In fact, the functional configuration must not only solve the data sources, but also be supplemented by reasonable and effective

*Correspondence: rbxiao@hust.edu.cn

² School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 400073, China
Full list of author information is available at the end of the article

methods. Compared to other data sources, patent data are a vital knowledge source for function-oriented product design as it allows a significant amount of data to be accessed, as well as a close relationship with product development [3]. In fact, it has become a vital knowledge source for function-oriented product design.

To achieve greater market share and margins, enterprises must continually develop new products with multiple functions. Moreover, firms must apply for patents to protect their intellectual property rights and avoid plagiarism. Many factors, such as product type, enterprises' business model, and production lead time, form a complicated relationship, as reflected in patents—this is problematic in design practice. Owing to the advantages of network theory in multi-entity relationship analysis, network-based patent analysis has recently garnered increasing attention from scholars [4]. Luo et al. investigated the product design space using a patent network [5].

Based on these observations, a patent-driven method for product function deployment based on a hypergraph is proposed herein. In addition, this study is performed to combine data mining technology with networks to solve multifunctional configuration problems during product development. The innovations of this study are reflected as follows: ① Functional knowledge is extracted from patents as nodes and then combined with product patents and patent applicants as edges to construct a multiedge hypergraph network; ② the function node's weight is calculated using the applicant edge, and the patent edge's weight is derived by combining the number of patent citations and the number of patent families; ③ a multifunctional combination weight calculation model is proposed and combined with the network subgraph algorithm to obtain the optimal function combination.

The remainder of this paper is organized as follows: The theoretical background and related literature review are presented in Section 2. In Section 3, a research framework for a patent-driven product functional configuration (PFC) is proposed. A weighted function hypergraph (FH) is described in Section 4, followed by an empirical analysis that verifies the validity of the proposed method. Finally, the conclusions are presented.

2 Literature Review

As a portfolio innovation, the PFC refers to the convergence of different technical features and solutions that can be distinguished from existing products [6]. This convergence is realized by combining elements with the original product [7]. However, the PFC is not an arbitrary superposition of elements; as such, an accurate and

comprehensive analysis is necessitated [8]. Therefore, current research pertaining to technology combination primarily focuses on three aspects: ① patent data analysis, ② technology opportunity analysis, ③ technology convergence analysis.

2.1 Patent Data Analysis

In patent mining, valuable data are extracted from both structured and unstructured patents. Structured data primarily include the application date, citation relationship, and classification number. Some methods attempt to extract technological information from data. In a previous study, Daim et al. [9] analyzed a technology development trend based on patent numbers and then identified core technologies for a firm's R&D. The results showed that the number of patent applications for emergency technologies increased faster than those for other technologies. Meanwhile, other scholars accumulated patent citation data and created a potential technology reorganization list [10]. A patent citation is a document cited by applicants or patent office examiners, and its content is associated with other patent applications. In general, more citations the higher quality and intrinsic technical value of the patent. Meanwhile, after a patent has been applied, the number of citations associated with the patent will continue to increase, and its value will increase as well. In addition to citations, patent classification numbers are typically used for technical analyses [11]. The patent classification scheme is a coding system that classifies inventions in a technical field [12]. The typically used classification numbers are the International Patent Classification (IPC) and Cooperative Patent Classification (CPC). Compared to the IPC, which comprises five levels, the CPC comprises six or seven levels and can provide more detailed technological information. Therefore, more scholars tend to apply CPC numbers instead of IPC numbers to identify technological opportunities [13].

Unstructured patent information is primarily composed of textual data, which are an important data source to current studies that provide abundant and detailed information. For instance, Blake and Ayyagari [14] obtained market hotspot information via the trend analysis of text themes in patents. Zhang and Yu [15] investigated technical topic extensions using a keyword analysis algorithm. In addition to keywords, some scholars used the relationships between words to identify design opportunities. Choi et al. [16] integrated dependency syntax and part-of-speech filtering methods to obtain subject, action, and object vocabulary, and then obtained word phrases related to technological opportunities.

Kwon et al. [17] identified the unintended consequences of emerging technologies by mining underlying semantics from patent texts.

2.2 Technology Opportunity Analysis (TOA)

TOA is a method of innovation monitoring based on bibliometric analysis and data mining. Arguably, TOA has become more important owing to the increase in the uncertainty and risk of product development. By monitoring the technological development of enterprises, Hou and Yang [18] identified valuable patents that were overlooked for a significant period as data sources for identifying design ideas. In addition, scholars have investigated the formation of patent jungle communities as a technological opportunity [19]. For instance, Jin et al. [20] used a technical efficiency matrix to identify vacuum technology, which has not yet been considered as a technology expansion objective. Li et al. [21] observed that goal-keeper patents are vital to the transfer of scientific theory to industrial applications.

Other researchers have attempted to identify technological opportunities in patents through text mining. Wang et al. [22] analyzed text topic development trends to identify topics associated with technological convergence. Yun and Geum [23] used the latent Dirichlet allocation algorithm to extract technical topics from patents. Kim et al. [24] monitored the development path through patent semantic similarity at different application times and provided a technology prediction reference. Li et al. [25] combined TRIZ theory and natural language processing technology to evaluate patent creativity and identified high-impact patents. Sheu and Yen [26] extracted information regarding harmful resources from patents to reduce risks associated with R&D.

2.3 Patent Data Visualization

Data visualization is crucial for understanding the results of patent analysis. Currently, patent data visualization methods are primarily classified into three categories: two-dimensional maps, incidence matrices, and network graphs.

A two-dimensional map is used to segment multidimensional information into two dimensions to facilitate visualization. Lee et al. [27] attempted to reduce the amount of patent data through principal component analysis to construct a technology map and then identify technology from blank areas. Lee et al. [28] constructed a landscape map from patent information as a vector space model to present the configuration of technological components. Seo et al. [29] proposed a portfolio map method

using two patent values for novelty indices as axes to investigate the patents of competing enterprises and then identify technological opportunities.

The incidence matrix is a logical matrix that shows the relationship between two classes of objects, including the morphology, design structure, and vector space matrices. Arciszewski [30] generated new schemes through literature mining, first mined technical keywords through patent text data, and then combined keywords through a morphological matrix to facilitate designers in conceiving new ideas. Feng et al. [31] calculated the correlation coefficient between technology and product using a correlation matrix and then identified technological development opportunities that are suitable for the current product. In addition, the design structure matrix is a typically used tool for analyzing the relationships between different objects. Zheng et al. [32] constructed a pairwise relationship matrix between themes, in which the matrix element is the number of co-occurring patents. The vector space model (VSM) is one of the most robust information-analysis methods developed hitherto. Jun et al. [33] introduced a matrix mapping and K-medoids clustering method based on a vector matrix to predict missing technology more accurately. Lei et al. [34] proposed a patent analytics method based on a VSM to solve semantics and curse-of-dimensionality loss.

In recent years, an increasing number of scholars have adopted graphs to perform patent analysis. Compared to two-dimensional maps and the incidence matrix, network graphs provide a better visualization through nodes and edges, and they are applied to technology weights and clusters via degree measurement algorithms [35]. Kim et al. [36] identified core technologies from the perspective of technological cross impacts using network graphs and association rule algorithms. Sung et al. [37] used expanding cell structure networks to analyze core technologies. Song et al. [38] demonstrated patent keywords through a core-peripheral network, as well as important technical keywords through gravity algorithms. Some studies were conducted using subgraphs formed from a subset of vertices of a graph and all the edges connecting pairs of vertices in the subsets. Lee et al. [39] used a subgraph unit based on the existing node analysis and applied a quadratic assignment problem algorithm to calculate the correlation between different subgraphs to analyze the technological integration. Lee et al. [40] adopted a frequent subgraph algorithm to analyze the correlation among network nodes and obtained the best technology combination by calculating the confidence and support. Sun et al. [41] formed different patent clusters using text mining technology and then

weighted the overlap between different clusters to analyze the technological integration.

Many scholars have performed TOA using patent-data-driven methods. The deficiencies of the current study are as follows:

- The objectives of previous studies focused primarily on technical opportunities and rarely involved the excavation of functional requirements. As such, good suggestions for functional market expansion are difficult to provide.
- Two necessary procedures are overlooked in the current research: calculating the weights of convergent objects in the functional configuration and organizing the clusters formed after convergence.
- The analysis tools used in current investigations typically assumes that objects exhibit a single relationship. However, multiple relationships exist in terms of the patent co-occurrence between patent functions and the applicant. These relationships can affect the identification and integration of functional opportunities.

In this context, a novel and efficient method must be developed to facilitate firms in detaching from market

function opportunities and creating optical functional configurations based on patent data by addressing the 2H-W.

3 Research Framework

Firms often apply patents for the design schemes of multifunction products, particularly consumer products, to expand the patent protection scope and reduce patent fees [42]. Thus, the functional configuration of existing products can be analyzed using patent data. In this study, a new patent-data-driven PFC method based on a hypergraph is proposed, and a framework based on this method is developed, as shown in Figure 1. This framework comprises four steps: patent data acquisition and mining, function hypergraph construction, functional configuration scheduling, and configuration analysis.

Step 1: Patent data acquisition and mining. R&D terms are used to retrieve industry patents based on customer demands and the industry life cycle. Patents are downloaded from the website to construct a database pertaining to local computers. In these patents, structured data, such as the number of citations, applicants, and application dates, are obtained via paragraph cutting. Unstructured data, such as text data, must be cleaned by removing noisy information such as numbers, symbols, and auxiliary vocabularies.

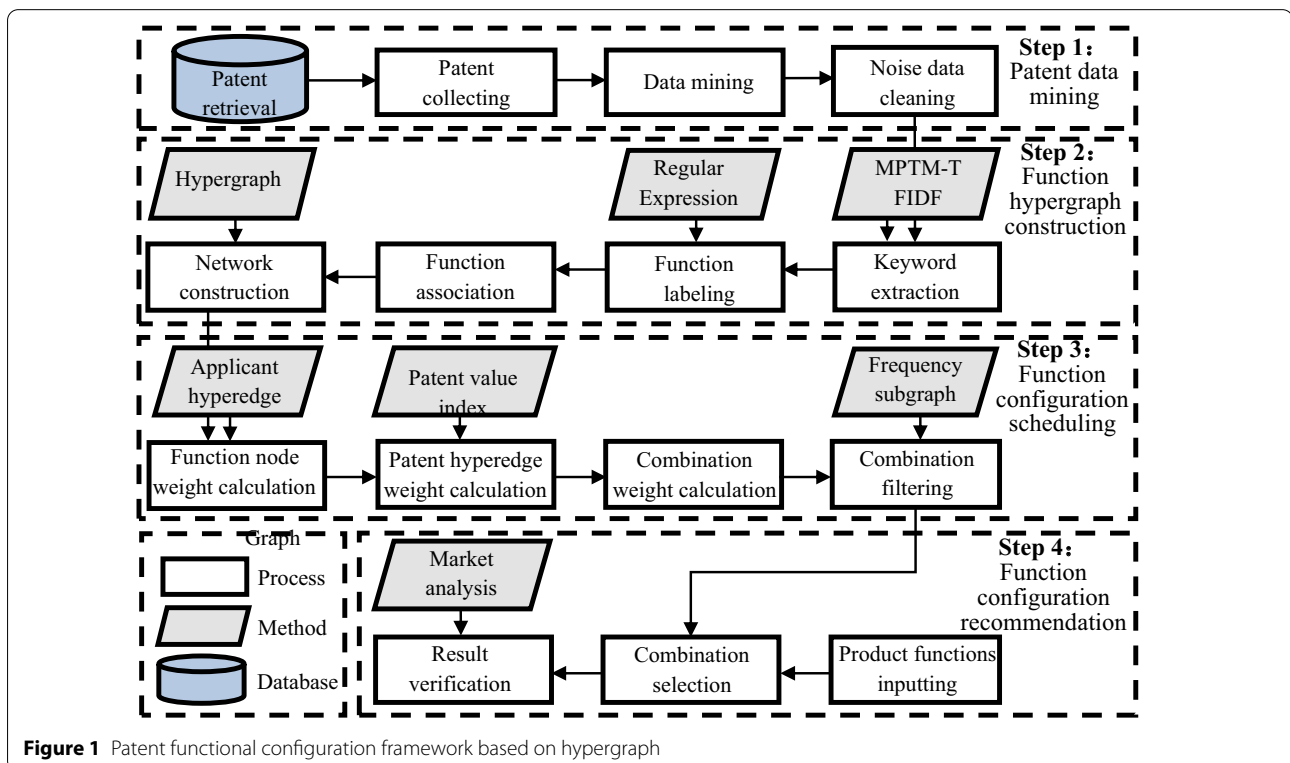


Figure 1 Patent functional configuration framework based on hypergraph

Step 2: Function hypergraph construction. First, multipart text mining based on the term frequency-inverse document frequency (TF-IDF) algorithm (MPTM-TFIDF) is used to weigh the words. Subsequently, keyword phrases are extracted from the patent text as patent function labels based on regular expressions. Words that compose a phrase should appear in the same sentence simultaneously, such as the phrase “cold water,” which is composed of both “cold” and “water” in the same sentence. Subsequently, a label set with different functions is formed. In addition, an adjacency matrix is applied to describe the relationship between the function and the patent or applicant. Finally, the patent function hypergraph model is constructed based on the matrix.

Step 3: Functional configuration scheduling. The applicant edge is used to weigh the function node. The citation number and patent family size are integrated to the weigh patent edges. A comprehensive calculation model is constructed for the weight function community, and an improved frequent subgraph algorithm (IFSA) is proposed to identify optical function combinations in the hypergraph network.

Step 4: Functional configuration recommendation. Based on the existing product functions or customer requirements, the target functions are obtained in Step 3. Finally, the accuracy of the configuration results is verified through market analysis.

4 Research Methodology

4.1 Keyword Extraction Based on MPTM-TFIDF

TF-IDF is a statistical measure algorithm that evaluates the importance of a word to a document in a collection [43]. TF-IDF is expressed mathematically in Eq. (1).

$$w_D^T = TF(T, D) \times IDF(T), \quad (1)$$

$$TF(T, D) = \frac{f(T, D)}{s(D)}, \quad (2)$$

$$IDF(T) = \log_2 \frac{s(N)}{1 + c(T, N)}, \quad (3)$$

w_D^T denotes the weight of term T in document D ; $TF(T, D)$ denotes the percentage of term T in document D ; $IDF(T)$ measures the rareness of term T that occurs across document D ; $f(T, D)$ denotes the frequency of term T in document D ; $s(D)$ denotes the number of terms in document D ; $s(N)$ denotes the number of all documents; $c(T, N)$ denotes the number of documents that contain the term T .

In the keyword extraction process, the TF-IDF algorithm is typically used to detach value words that are

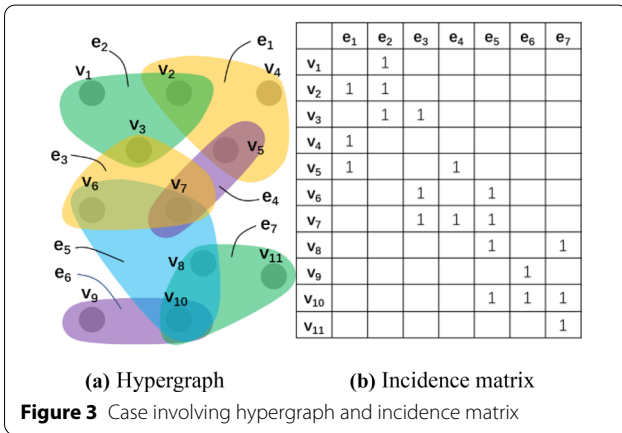
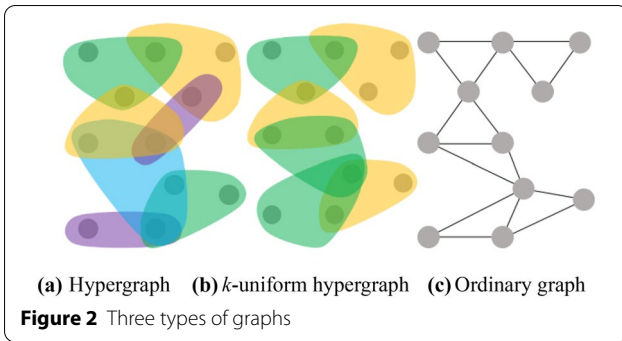
rarely shown in documents but are essential [44]. However, the effect of the algorithm is determined by the text data volume and synonyms [45]. In practice, product function words are distributed in different sections of the patent, such as the title, abstract, claim, and technical background, and the amount of text data in different sections varies significantly. Moreover, many synonyms exist for the function keywords in the patents. All of the abovementioned factors affect the accuracy of the algorithm.

Hence, the MPTM-TFIDF method is proposed herein. First, to ensure high keyword extraction accuracy, the critical information extracted from all patent titles as text is significantly less than that from other sections. Subsequently, keywords with higher TF-IDF weights in different patents are obtained, and synonyms with the same meanings and high similarity are merged. Notably, the similarity is calculated using WordNet, which is a large English lexical database comprising 155287 words and 117659 synonyms (it can be downloaded from the website <https://wordnet.princeton.edu>). The semantic distance information of words is recorded in the database and can be extracted using natural language toolkit to calculate the similarity between words and then used to mine synonyms based on an empirical threshold. This method has been described in many papers [46–48] and thus will not be further explained herein. Finally, through the set elements, similar words in the abstract and the technical background of all patents are searched to determine the patent’s functions via regular expressions.

4.2 Hypergraph Model Construction

In an ordinary graph, one edge precisely connects two vertices that denote a one-to-one relationship [49]. The structure is concise but limited in expressing the relationships between multiple vertices [50]. By contrast, the hyperedge in the hypergraph links the number of nodes, and hyperedges in the same networks can exist simultaneously. Therefore, a hypergraph was selected for this study.

Definition 1. A hypergraph is expressed as $H=(V,E)$, where $V=\{v_1, v_2, \dots, v_n\}$ is a finite set of nodes known as vertices, and $E=\{e_1, e_2, \dots, e_m\}$ is an indexed family of sets known as hyperedges, in which $e_i \in V$. The degree of a vertex is the number of hyperedges to which it belongs, i.e., $d(v)=|\{e:v \in e\}|$, and the size of a hyperedge is its cardinality node, i.e., $|e_i|=k(1 \leq k \leq n)$. A hypergraph with hyperedges of size k is known as a k -uniform hypergraph, whereas a 2-uniform hypergraph is known as an ordinary graph [51]. Figure 2 shows an example of three types of graphs.



The hypergraph can be illustrated as an incidence matrix $|V| \times |E|$ with element $h(v,e)$, whose value is defined as shown in Eq. (4).

$$h(v,e) = \begin{cases} 1 & v \in e, \\ 0 & \text{others.} \end{cases} \quad (4)$$

In addition, Figure 3 illustrates the relationship between the incidence matrix and hypergraph.

4.3 PFC Model Construction

The PFC involves not only products and firms, but also complex multi-entity and multilateral relationships. When the patent-driven functional configuration method is adopted, these two relationships are transformed into patent and applicant relationships to form a hypergraph model of product functions.

Definition 2. A patent FH can be expressed as $FH=(F,E)$, where $F=\{f_1,f_2,\dots,f_m\}$ denotes the set of function nodes, $E=\{ep,ec\}$ the set of hyperedges, $ep=\{ep_1,ep_2,\dots,ep_p\}$ the set of patent hyperedges, and $ec=\{ec_1,ec_2,\dots,ec_c\}$ the set of applicant hyperedges.

Because the FH has more than one hyperedge, it can be illustrated using incidence matrices with hyperedges and nodes. The value of the matrix elements can be calculated using Eq. (4).

4.4 Hypergraph Weight Calculation

The calculation for the hypergraph weight includes those for the function node weight and patent hyper-edge weight.

4.4.1 Function Node Weight Calculation Based on Patent Applicant Hyperedge

For a product to be a leader in the market, it must satisfy customer requirements continuously. In this regard, important functions must be integrated to increase market attractiveness. Currently, the definitions of essential functionality are scarce. Based on the definition of technological opportunities [52], an important function can be defined as follows:

Definition 3. Important functions refer to those that are widely and promptly accepted by the market.

From a market perspective, ensuring that a function is generally accepted by consumers is important. From the perspective of patents, important functions are widely used by respondents. The higher the involvement of enterprises in product development, the more critical the function becomes [53]. The later the feature appears, the greater is the probability of it becoming popular.

Therefore, the weight of the function node w_{f_i} in the hypergraph is calculated based on the time index of the applicant's hyperedge and function, as shown in Eq. (5).

$$w_{f_i} = wt_i \cdot \sum_{ec_j^i \in ec} h(f_i, ec_j^i), \quad (5)$$

wt_i denotes the time index of function f_i , and ec_j^i denotes the hyperedges covering the node of function f_i . The longer the function is available, the less popular it will be in the market. By contrast, new features are more likely to become popular in the market. Therefore, the inter-function index was calculated as shown in Eq. (6).

$$wt_i = \frac{1}{(T_n - T_{if} + 1)}, \quad (6)$$

T_n denotes the current year, and T_{if} denotes the year when the product of function f_i is first applied as a patent. Based on the formulas above, the following equation is derived to calculate the weight of function f_i :

$$w_{f_i} = \frac{\sum_{ec_j^i \in ec} h(f_i, ec_j^i)}{(T_n - T_{if} + 1)}. \tag{7}$$

Owing to the significant deviation in the number of patent applicants' hyperedges for different functional nodes, data normalization is necessary to adjust values from large to minor scales. Many types of statistical normalization methods exist, including standardized moment, coefficient of variation, standard score, and max–min normalization. Based on the patent data characteristics, max–min normalization is adopted such that all values are within the range [0,1], as shown in Eq. (8).

$$w'_{f_i} = \frac{w_{f_i} - \min(w_{f_i})}{\max(w_{f_i}) - \min(w_{f_i})}, \tag{8}$$

w_{f_i} and $\max(w_{f_i})$ denote the lowest and highest values of the weight range of all functional nodes, respectively.

4.4.2 Calculation of Patent Hyperedge Weight Based on Patent Quality

The functional distribution of the corresponding products can be extracted through product patent analysis. The patent hyperedge weight is calculated from the perspective of patent quality. According to the World Intellectual Property Organization [54], citation number and patent family size are two core indexes for calculating patent quality. The more frequently a patent is cited, the greater is its impact [55]. When the size of the patent family increases, the number of countries filed for the patent as well as the economic value of the patent increase [56]. Therefore, these two indexes are incorporated into the patent hyperedge weight calculation, as shown in Eq. (9).

$$w_{ep_i} = \varphi \cdot fep_i + (1 - \varphi) \cdot cep_i^t, \tag{9}$$

w_{ep_i} denotes the weight of the patent hyperedge ep_i ; fep_i denotes the number of patent families of patent ep_i ; cep_i^t denotes the number of citations per year of patent ep_i and reflects the degree to which the patent is valued by peers; φ denotes the weight ratio of fep_i and cep_i^t .

It is noteworthy that the company should apply a patent for the product as soon as it is developed. The earlier a patent is applied, the greater is the possibility for it to be cited. Consequently, the number of citations will become higher than that of subsequent patents [57]. To eliminate the effect of patent application time, the number of patent citations per year in the entire life cycle is set as the weight calculation index. Hence, cep_i^t is calculated as follows:

$$cep_i^t = \frac{cep_i}{T_n - T_{ep}^i}, \tag{10}$$

cep_i represents the total number of citations of patent ep_i , and T_{ep}^i represents the application year of patent ep_i .

Through the maximum and minimum standardization processes, the weight w_{ep_i} of ep_i can be calculated as follows:

$$w_{ep_i} = \varphi \cdot \frac{fep_i - \min(fep)}{\max(fep) - \min(fep)} + (1 - \varphi) \cdot \frac{cep_i^t - \min(cep^t)}{\max(cep^t) - \min(cep^t)}. \tag{11}$$

4.5 Functional Configuration Based on Hypergraph

The PFC comprises two aspects: evaluation and acquisition of the functional community.

4.5.1 Evaluation of Function Community

To ensure the versatility of a product, a functional community is formed and reflected in an FH. Before the PFC is formed, the functional community must be evaluated comprehensively in advance and filtered through a hypergraph. In the PFC model, nodes representing functions are connected through patented hyperedges. The strength of the connection depended on the weight of the hyperedge. The higher the weight, the closer is the connection between nodes, such that some nodes form clusters or communities. The hyperedge weight is an indicator for evaluating the community. The importance of nodes is another indicator for evaluating communities. Node weight is positively correlated with the importance of the community. For example, the handheld function and rain function are closely related, and the weights of the two functions are high; therefore, the two functions can be easily integrated into the same product.

Suppose that function nodes and hyperedges form the same community subgraph $FH^o(F^o, ep^o)$, where $F^o = \{f_1^o, f_2^o, \dots, f_n^o\}$ and $ep^o = \{ep_1^o, ep_2^o, \dots, ep_m^o\}$. The weight of the functional community w_{FH^o} is calculated using Eq. (12).

$$w_{FH^o} = \sqrt{\frac{\left(\sum_{ep_i^o \in ep^o} w_{ep_i^o} + 1\right) \cdot \left(\sum_{f_i^o \in F^o} w_{f_i^o} + 1\right)}{N_{F^o}}}, \tag{12}$$

$w_{ep_i^o}$ denotes the weight of the patent hyperedge ep_i^o ; $w_{f_i^o}$ denotes the weight of function node f_i^o ; N_{F^o} denotes the number of function communities F^o .

Based on Eq. (12), the weight of one community is higher when it contains more functions and patents. This is because a product can satisfy various individual requirements when it contains many functions, which

is welcomed by customers. Meanwhile, the more products with similar functional combinations, the more critical the functional community becomes.

4.5.2 Function Community Acquisition Based on Frequent Subgraphs

An FH contains many subgraphs, each representing a functioning community. To obtain the optimal combination of functions, a frequent subgraph algorithm is introduced to select the optimal function community. Currently, two frequent subgraph mining algorithms are typically used: Apriori and FP-Growth. Compared to the FP-Growth algorithm, the Apriori algorithm is more mature and widely used [58]. Therefore, Apriori was adopted in this study for FH subgraph mining.

The Apriori algorithm is generally used to screen subgraphs. However, existing studies that obtain the optimal subgraph based on the Apriori algorithm disregards the weight of the subgraph, and the results are inaccurate. Therefore, an IFSA is proposed herein. This algorithm uses the weights of functional communities as the basis for subgraph screening.

Definition 4. IFSA. For a hypergraph H and a minimum comprehensive weight τ , $S_{up}(H, FH^p)$ represents the weight of the subgraph FH^p in H ; when $S_{up}(H, FH^p) \geq \tau$, FH^p is a frequent subgraph of H . $S_{up}(H, FH^p)$ is calculated as follows:

$$s_{up}(H, FH^o) = \frac{w_{FH^o}}{w_H}, \tag{13}$$

w_H is the sum of weights covering all patent communities, and it is calculated as follows:

$$w_H = \sum_{FH^o \in H} w_{FH^o}. \tag{14}$$

The algorithm is described as follows:

```
// IFSA
Function apriori (FHk, D, τ)
//D – Graph Database
//FHk – frequent subgraphs of size “k”
//τ – minimum weight of subgraph
// Lk+1 - list of graphs whose size is “k+1”
FHk+1 ← NULL;
for all gi ∈ FHk:
    Lk+1 ← form as many as graphs as possible by
    combining gi and gj: the size of the resultant graph is
    “k+1.”
    for all gu ∈ Lk+1:
        if [ (Sup(H, gu) ≥ τ) AND (gu ∉ FHk+1) ]:
            FHk+1 ← FHk+1 ∪ gu
        else if [FHk+1 is NULL]:
            k is optimal size
    k ← k+1
end
end function
```

Through the RFSa, the optimal number of function combinations k is obtained, and groups with better weights from the same number of function combinations are explored. The algorithm provides a reference for designers to determine the function quantity and optimal function combinations.

5 Case Study

Fulfilling the demands of every individual customer for bathroom furniture and accessories is a challenging task, particularly for showerheads. Many enterprises aim to develop fashionable and attractive products. This section presents a case study of the proposed method. The algorithms were encoded and executed using Python

Table 1 Information of patents

No	Patent number	Applicant	Title of patent	TF-IDF value
1	US14/360881	DURANT	Concealed top cover-type shower head	"concealed": 0.506, "cover": 0.481, "head": 0.205, "shower": 0.134, "top": 0.481, "type": 0.471
2	US15/334907	DURANT	Combination showerhead with swing button switching	"button": 0.422, "combination": 0.431, "showerhead": 0.275, "swing": 0.603, "switching": 0.376, "with": 0.234
3	US15/871091	JOHANNES	Massage shower head	"head": 0.404, "massage": 0.877, "shower": 0.2613
...
1358	US17/040644	KOHLER	Washing unit	"unit": 0.651, "washing": 0.759

Table 2 List of functions

Function code	Description of functions	Synonyms of function
f01	Adapted for seats	Seat* or chair *
f02	Adapted for wash-basins or bath	Wash-basins or bath
f03	Washing specific parts of the body	Special part
f04	Jet particular shape water	Particular shape
f05	Scanner water	Scanner
f06	Massage function	Massage
f07	Jet annular, tubular or hollow conical form water	Annular, tubular or hollow conical form
f08	Moving outlet	Moving outlet
f09	Recovery of heat	Recovery of heat
f10	Diverting cold water	Cold water
f11	Concealed shower	Conceal*
f12	Void slippage or dropping	Slippage or dropping
f13	Combined shower	Combined shower
f14	Fixed shower	Fixed shower
f15	Handle shower	Handle shower
f16	Void leakage	Leakage
f17	Water saving	Water saving
f18	Rainfall water	Rain
f19	Mist water	Mist water
f20	Pulsating water	Pulsating water
f21	SPA	SPA
f22	Void blockages	Blockages
f23	Self-clean function	Self-clean
f24	Lighting	Light* or led
f25	Anion Energy	Anion*
f26	Stop flow	Stop
f27	Kill bacteria	Bacteria
f28	Wand	Wand
f29	Extendable outlet	Extendable outlet
f30	Selective outlet	Selective outlet
f31	Suspending or supporting shower	Suspending or supporting
f32	Diverter valves	Diverter valves
f33	Control by button	Button
f34	Adjustable head	Adjustable
f35	Electrical control	Electrical control
f36	Multi-outlets	Outlets
f37	Detach function	Detach
f38	Strainers	Strainers
f39	Oriented jet	Oriented jet
f40	Mixing ratio	Mix ratio
f41	Restrict flow	Restrict flow
f42	Temperature response	Temperature response
f43	Lifting valve	Lift valve
f44	Control pressure	Control pressure
f45	Mounted faucet	Mounted faucet
f46	Control volume	Control volume
f47	Heating water	Heating
f48	Filter stream	Filters located upstream
f49	Control temperature	Control temperature

Table 2 (continued)

Function code	Description of functions	Synonyms of function
f50	Generate electricity by motor	Motor
f51	Mixture with other material	Mixture or aromatherapy
f52	Jet particular shape	Particular shape
f53	Scald	Scald
f54	Touchless control	Touchless
f55	Adjusting position of head	Adjusting the position
f56	Jetting hot air	Hot air
f57	Collecting water	Collect*
f58	Making radio	Radio
f59	Making alarm	Alarm
f60	Making music	Music
f61	Communication of internet	Internet
f62	Communication of Bluetooth	Bluetooth
f63	cleaning hair	Hair
f64	Cleaning pets	Pet* or cat* or dog*
f65	Using for child	Child* or baby

*Multiple arbitrary characters

Table 3 Function labels of patents

No	Patent number	Function label
1	US14/360881	f11, f14, f30,
2	US15/334907	f15, f30, f37,
3	US15/871091	f06, f14, f18, f46
	...	
1358	US17/040644	f15, f17, f57

software. Patent data were retrieved and downloaded from Patsnap (<https://www.patsnap.com/>), which is a well-known commercial patent database.

Currently, showerheads with rain functions are primarily manufactured by one firm. This product lacks competitiveness as its functions are scarce. Therefore, the firm intends to develop a new multifunction shower and has commissioned us to aid in patent analysis to detach new

function opportunities from the market and then arrange the functional configuration. Initially, we used the keywords and CPC numbers to search for patents applied in the USPTO. The search formula used was as follows: Title or Abstract:(showerhead* or shower head* or sprayer*) AND CPC:(B05B1/18) AND Time:(from 19140101 to 20200101). A total of 1358 patents were obtained from the USPTO database (as listed in Table 1).

The TF-IDF algorithm is used to calculate the weight of the vocabulary in patent titles, and the results are shown in Table 1. Words with high values are often associated with product functions. Word similarity is calculated using the WordNet database, and synonyms with a threshold exceeding 0.1 are merged into function keywords, as shown in Table 2.

The functions in Table 2 were used to label patents with regular expressions, and the results are listed in Table 3. Table 4 indicates that patents can have multiple

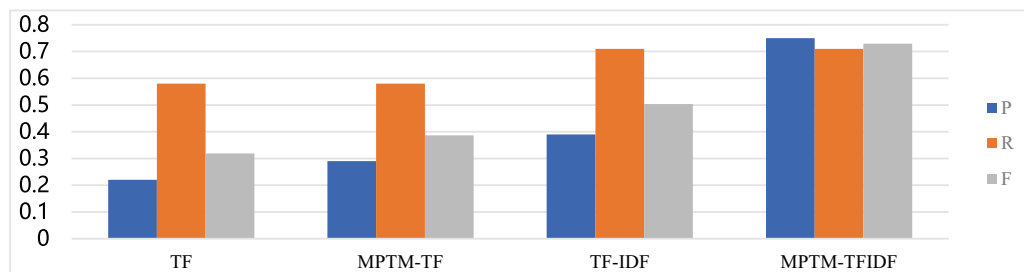


Figure 4 Comparison of four algorithms

Table 4 Weight of function nodes in hypergraph

Function code	T_f (year)	wt	Number of applicants	w_f	w'_f
f01	2016	0.167	12	2.000	0.135
f02	1972	0.020	32	0.640	0.039
f03	2001	0.048	12	0.571	0.035
f04	1991	0.032	15	0.484	0.029
f05	2011	0.091	7	0.636	0.039
f06	1988	0.029	46	1.353	0.090
f07	2013	0.111	13	1.444	0.096
f08	1972	0.020	55	1.100	0.072
f09	2006	0.063	6	0.375	0.021
f10	2008	0.071	15	1.071	0.070
f11	1945	0.013	53	0.688	0.043
f12	1970	0.019	23	0.442	0.026
f13	1942	0.013	24	0.300	0.016
f14	1981	0.024	220	5.366	0.371
f15	1914	0.009	679	6.287	0.436
f16	1924	0.010	164	1.673	0.112
f17	1989	0.030	71	2.152	0.146
f18	1920	0.010	395	3.873	0.266
f19	1959	0.016	106	1.683	0.113
f20	1981	0.024	31	0.756	0.048
f21	2009	0.077	1	0.077	0.000
f22	1989	0.030	21	0.636	0.039
f23	1959	0.016	19	0.302	0.016
f24	1984	0.026	31	0.816	0.052
f25	1984	0.026	7	0.184	0.008
f26	1982	0.025	35	0.875	0.056
f27	2004	0.056	5	0.278	0.014
f28	1992	0.033	9	0.300	0.016
f29	1990	0.031	81	2.531	0.172
f30	1962	0.017	188	3.133	0.214
f31	1914	0.009	93	0.861	0.055
f32	1914	0.009	75	0.694	0.043
f33	1939	0.012	27	0.325	0.017
f34	1971	0.020	50	0.980	0.063
f35	1983	0.026	36	0.923	0.059
f36	1988	0.029	40	1.176	0.077
f37	1974	0.021	66	1.375	0.091
f38	1961	0.016	67	1.098	0.072
f39	2007	0.067	34	2.267	0.154
f40	1963	0.017	35	0.593	0.036
f41	1983	0.026	39	1.000	0.065
f42	1990	0.031	42	1.313	0.087
f43	2001	0.048	32	1.524	0.101
f44	2013	0.111	129	14.333	1.000
f45	1996	0.038	27	1.038	0.067
f46	1959	0.016	102	1.619	0.108
f47	2013	0.111	6	0.667	0.041

Table 4 (continued)

Function code	T_f (year)	wt	Number of applicants	w_f	w'_f
f48	2003	0.053	23	1.211	0.080
f49	1959	0.016	22	0.349	0.019
f50	1978	0.023	20	0.455	0.026
f51	1916	0.009	181	1.708	0.114
f52	1991	0.032	24	0.774	0.049
f53	1942	0.013	78	0.975	0.063
f54	2011	0.091	17	1.545	0.103
f55	1982	0.025	54	1.350	0.089
f56	1994	0.036	16	0.571	0.035
f57	1997	0.040	29	1.160	0.076
f58	1988	0.029	47	1.382	0.092
f59	1994	0.036	27	0.964	0.062
f60	1988	0.029	12	0.353	0.019
f61	2015	0.143	5	0.714	0.045
f62	2010	0.083	21	1.750	0.117
f63	1934	0.011	141	1.602	0.107
f64	1999	0.043	22	0.957	0.062
f65	2013	0.111	3	0.333	0.018

functions. Additionally, the number of features of different patents can vary significantly.

To further verify the effectiveness of this method, we compared our results with three typical keyword extraction algorithms, i.e., TE, MPTM-TE, and TF-IDE, based on the precision (P), recall (R), and F-value (F), as shown in Eqs. (13)–(15).

$$P = \frac{TP}{TP + FP}, \tag{15}$$

$$R = \frac{TP}{TP + FN}, \tag{16}$$

$$F = \frac{2 \times P \times R}{P + R}, \tag{17}$$

TP , FP , and FN donate the numbers of true positive, false positive, and false negative instances, respectively. Based on these counts, 10 patents containing more than 600 words were randomly selected as test objects, and experts were recruited to verify the effect; the results are shown in Figure 4. Compared to other algorithms, MPTM-TFIDF yielded significantly better P , R , and F values.

Applicant hyperedges were used to calculate the weight of the node. First, the current year was set at 2021. The number of patent applicants for all functions was calculated, and the min–max normalization

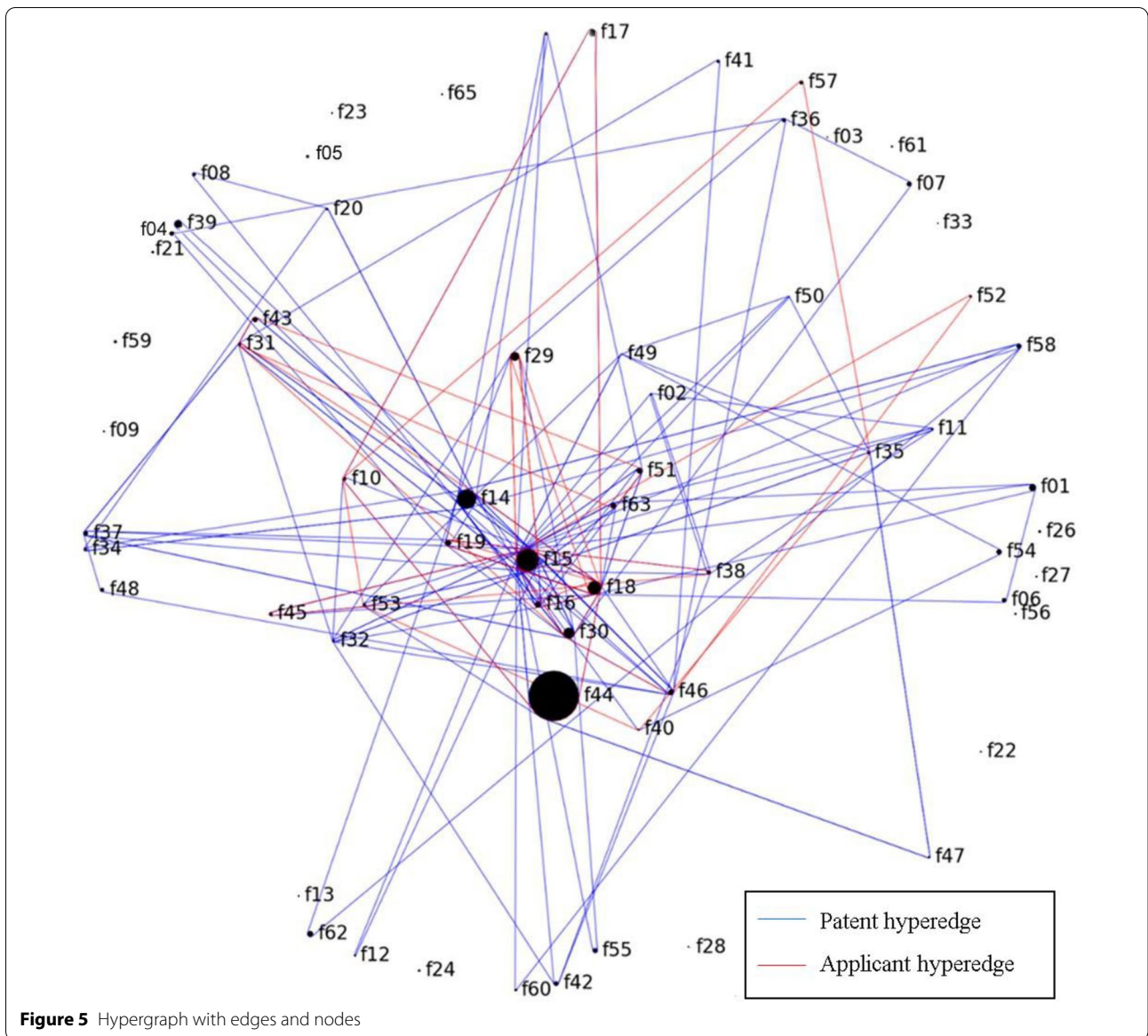


Figure 5 Hypergraph with edges and nodes

Table 5 Weights of patent edges in hypergraph

No	Application number	T_{ep}	fep	cep	w_{ep}
1	US14/360881	2012	3	0	0.009
2	US15/334907	2016	4	0	0.014
3	US15/871091	2018	2	0	0.005
...
1354	US17/040644	2019	1	0	0

Note: To save space, only part of the evaluation data is provided

Table 6 Weight of function community

Functions community	w_{FH^o}
f51, f40, f47, f35, f44, f49, f15	1.608
f51, f40, f47, f35, f44, f49, f16, f53, f15	1.241
f51, f40, f47, f35, f50, f44, f49, f15	1.066
f15, f18	0.894
...	...
f52, f20	0.069

Note: To save space, only part of the evaluation data is provided

algorithm was applied to obtain the weight in the range (0,1). The calculation results are listed in Table 4.

For a more concise visualization of the graph, the hypergraph is shown using the Python hypergraph tool (see Figure 5), where the patent hyperedge is labelled as “ep,” and the applicant is labelled as “ec.” The functions in Table 3 are used as the nodes, and both the patents and applicants in Table 1 are used as the hyperedges. To distinguish between different functions, functions with higher weights are represented by nodes with a larger radius.

Because the importance of the patent citation number ep and patent family size fep is equal, the weight ratio of the two indicators φ is set to 0.5 after a discussion among the experts. Based on the patent data, the weight w_{ep} of the patent hyperedges is calculated using cep and fep (as listed in Table 5). The value range of the patent family is (1, 115) and the patent citation is (0, 245), as shown in Table 5. The max and min values are counted in the max–min normalization.

Although the total number of patents is 1354, the number of communities is 753 when patents with the same function are merged into one community. The weights of the functional communities are identified using Eq. (12) and are listed in Table 6. Clearly, many MFPs are more popular than the products with fewer functions or a single function in the market. This indicates that the MFP is well-received by the market.

Because the firm’s existing product functions are the handling function (f15) and rainfall water function (f18), the RFSA introduced in Section 4.5.2 is implemented to aid the firm in completing the PFC. The minimum weight τ is set to 0.2, and the subgraphs include (f35, f60, f18, f58, f15), (f18, f63, f15), (f18, f63, f15, f01), (f18, f15, f17, f41), (f42, f17, f38, f18, f15), and (f18, f19, f15), whose weights that exceed τ are mined and filled in yellow (as shown in Figure 6). Meanwhile, nodes with high weights, such as f44 and f14, do not appear in these optimal communities. This implies that the functional configuration process depends on both the element weight and degree of correlation between different elements, which suggests a complicated process.

To verify the results obtained using our method, the number of products on an e-commerce website was counted. Currently, more than 7000 shower products are listed on Amazon (<https://www.amazon.com>). Because 753 function communities exist, based on an analysis of shower patents, each community has fewer than 130

functions. However, the functional communities are primarily identified in more than 200 products (Table 7). To further verify the effectiveness of the method, six functional communities with lower weights are listed. Table 7 shows that the quantities of these products are significantly lower than the average. This implies that MFP designs are more popular in the market.

In addition, one case involving an MFP placed on the shelf with functions f35, f60, f18, f58, and f15 is shown in Figure 7 (https://www.amazon.com/LINLIM-Handheld-Bluetooth-Holiday-Creative/dp/B08ZSLRXGX/ref=sr_1_59?dchild=1&keywords=shower+music&qid=1617437581&s=kitchen-bath&sr=1-59). The case study shows that multifunctional products are welcomed by customers.

6 Conclusions

A patent-data-driven method based on a hypergraph network was proposed herein to solve the 2H-2 problem in the MFP design process. In addition, NLP and association-rule algorithms were applied. The contributions of this study are summarized as follows:

(1) In this study, the MPTM-TFIDF algorithm was used to extract functional keywords from patent title text; subsequently, these keywords were used to retrieve patent full-text data to label each patent with function keywords. This method can accurately mine most functional data.

(2) An FH was constructed, in which patents or applicants represent the function and edges represent nodes. The applicants calculated the weight of a node, and the weight of the patent edge was calculated based on the number of citations and families. In addition, a community weight calculation model for the function nodes was proposed.

(3) Based on the improved Apriori algorithm, an IFSA algorithm suitable for a weighted hypergraph network was proposed. By calculating and comparing the weights of functional communities to determine the optimal functional combinations, this algorithm can promptly provide market opportunities for product design.

Finally, the method proposed herein was applied to the design of shower products and then verified using e-commerce data. In fact, a PFC must consider many factors, such as fashion, regulations, policies, and incentives. Therefore, patent data alone are insufficient for product design, and other types of data are required.

(See figure on next page.)

Figure 6 Frequent subgraphs of hypergraph network whose weight exceeds the threshold

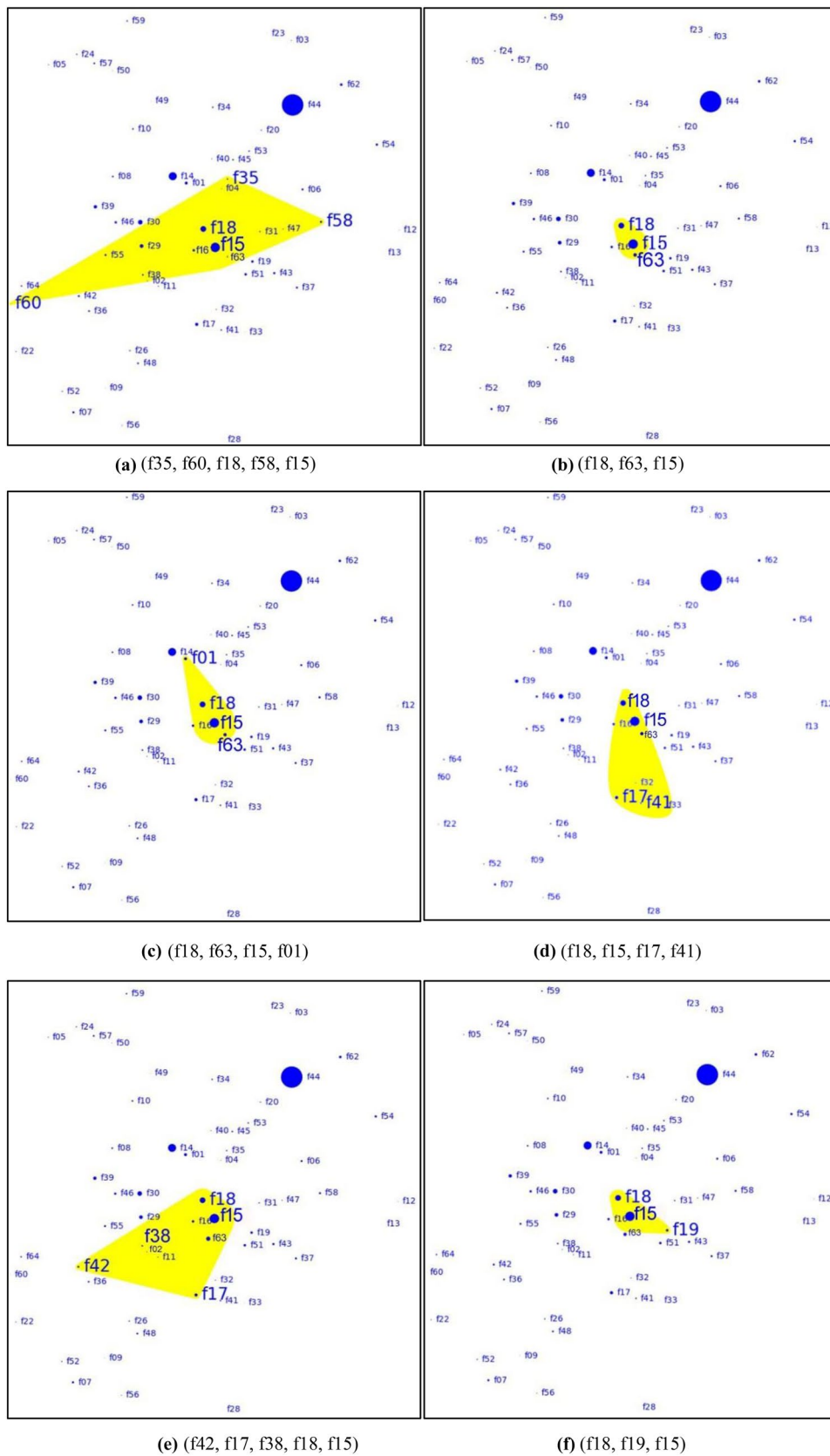
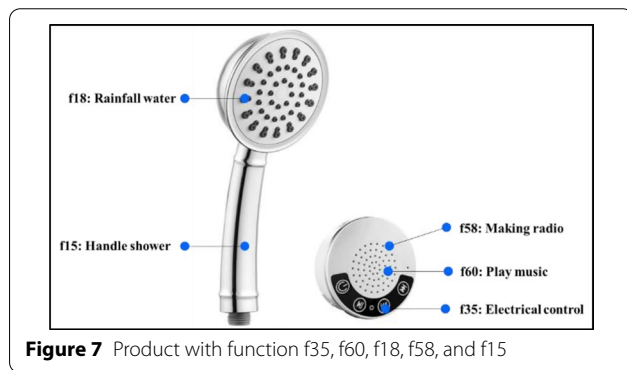


Figure 6 (See legend on previous page.)

Table 7 Number of some multifunctional products on Amazon

Product functions communities	Weight	Number of products	Web of Amazon
(f35, f60, f18, f58, f15)	0.627	252	https://www.amazon.com/s?k=shower+music&rh=n%3A680352011&ref=nb_sb_noss
(f18, f63, f15)	0.586	381	https://www.amazon.com/s?k=shower+hand+hair&i=tools&rh=n%3A680352011%2Cp_n_feature_seventeen_browse-bin%3A16966680011&dc&qid=1617192575&rnid=16966675011&ref=sr_nr_p_n_feature_seventeen_browse-bin_5
(f18, f63, f15, f01)	0.475	381	https://www.amazon.com/s?k=shower+hand+hair&i=tools&rh=n%3A680352011%2Cp_n_feature_seventeen_browse-bin%3A16966680011&dc&qid=1617192575&rnid=16966675011&ref=sr_nr_p_n_feature_seventeen_browse-bin_5
(f18, f15, f17, f41)	0.469	221	https://www.amazon.com/s?k=shower+music+flow&rh=n%3A680352011&ref=nb_sb_noss
(f42, f17, f38, f18, f15)	0.466	195	https://www.amazon.com/s?k=shower+filter+temperature&rh=n%3A680352011&ref=nb_sb_noss
(f18, f19, f15)	0.402	356	https://www.amazon.com/s?k=shower+mist&i=tools&rh=n%3A680352011%2Cp_n_feature_seventeen_browse-bin%3A16966680011&dc&qid=1617192903&rnid=16966675011&ref=sr_nr_p_n_feature_seventeen_browse-bin_5
(f18, f15, f55, f57)	< 0.01	62	https://www.amazon.com/s?k=shower+rainfall+hand+adjust+collect&crd=3DPK04QBR3V6&prefix=shower+rainfall+hand+adjust+collec%2Caps%2C664&ref=nb_sb_noss
(f18, f15, f50)	< 0.01	58	https://www.amazon.com/s?k=shower+rainfall+hand+electric+generate&crd=RG6W52A24K0&prefix=shower+rainfall+hand+electric+generat%2Caps%2C427&ref=nb_sb_noss
(f19, f18, f15, f55)	< 0.01	123	https://www.amazon.com/s?k=shower+rainfall+hand+mist+water+adjust+position+head&crd=1BSKRER16LLPE&prefix=shower+rainfall+hand+mist+water+adjust+posit ion+hea%2Caps%2C396&ref=nb_sb_noss
(f18, f63, f15, f57)	< 0.01	37	https://www.amazon.com/s?k=shower+rainfall+hand+collect+hair&crd=4JPC20250Y6J&prefix=shower+rainfall+hand+collect+h%2Caps%2C447&ref=nb_sb_noss
(f18, f64, f37, f24, f15, f65)	< 0.01	62	https://www.amazon.com/s?k=shower+rainfall+hand+pet+light+child+detach&crd=29EAXZA5E34VT&prefix=shower+rainfall+hand+pet+light+child+deta%2Caps%2C387&ref=nb_sb_noss



Acknowledgements

Not applicable.

Author Contributions

WL and RX were in charge of the whole trail; XL assisted with patent text mining. All authors read and approved the final manuscript.

Authors' Information

Wenguang Lin, born in 1985, is currently a lecturer at Key Laboratory of Intelligent Manufacturing Equipment, Xiamen University of Technology, China. He received his doctor degree from Xiamen University, China, in 2014. His research

interests include innovation design theory and data driven product design method.

Xiaodong Liu, born in 1995, is currently a master candidate at Key Laboratory of Intelligent Manufacturing Equipment, Xiamen University of Technology, China.

Renbin Xiao, born in 1965, is currently a professor and PhD candidate supervisor at School of artificial intelligence and automation, Huazhong University of Science and Technology, China. His main research interests include data driven product design, design method, and swarm intelligence. E-mail: rbxiao@hust.edu.cn

Funding

Supported by National Natural Science Foundation of China (Grant No. 51875220), and China Fujian Province Social Science Foundation Research Project (Grant No. FJ2021B128).

Competing Interests

The authors declare no competing financial interests.

Author Details

¹Key Laboratory of Intelligent Manufacturing Equipment, Xiamen University of Technology, Xiamen 361005, China. ²School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 400073, China.

Received: 3 September 2021 Revised: 20 April 2022 Accepted: 22 April

2022

Published: 26 May 2022

References

- [1] M Liu, D Trefler. What's the big idea? Multi-function products, firm scope and firm boundaries. *J Econ Behav Organ*, 2020, 180(5): 381-406.
- [2] Y Lin, D Cao, Y Han, et al. Guiding conceptual design modeling based on product systems analysis. *Kybernetes*, 2011, 40(5-6): 790-799.
- [3] Z J Li, C L Wu, X H Zhu, et al. Design by analogy: achieving more patentable ideas from one creative design. *Chinese Journal of Mechanical Engineering*, 2018, 31(2): 25-34.
- [4] S Jun, S J Lee. *A small world network for technological relationship in patent analysis*. New York: Springer, 2013.
- [5] J Luo, B Song, L Blessing, et al. Design opportunity conception using the total technology space map. *AI Edam*, 2018, 32(4): 449-461.
- [6] F Gallouj, O Weinstein. Innovation in Services. *Research Policy*, 1997, 26(4-5): 537-556.
- [7] M M Keupp, O Gassmann. Resource constraints as triggers of radical innovation: Longitudinal evidence from the manufacturing sector. *Research Policy*, 2013, 42(8): 1457-1468.
- [8] C J M Jeroen, Van den Bergh. Optimal diversity: Increasing returns versus recombinant innovation. *Journal of Economic Behavior & Organization*, 2008, 68(3-4): 565-580.
- [9] T Daim, I Iskin, X Li, et al. Patent analysis of wind energy technology using the patent alert system. *World Patent Information*, 2012, 34(1): 37-47.
- [10] Y Takano, C Mejia, Y Kajikawa. Unconnected component inclusion technique for patent network analysis: Case study of Internet of Things-related technologies. *Journal of Informetrics*, 2016, 10(4): 967-980.
- [11] S J Lee, S Jun. Key IPC codes extraction using classification and regression tree structure. *International Symposium Conference on Advanced Intelligent Systems*, Daejeon, Korea, November 13-16, 2013: 101-109.
- [12] S Oh, J Choi, N Ko, et al. Predicting product development directions for new product planning using patent classification-based link prediction. *Scientometrics*, 2020, 125(3): 1833-1876.
- [13] Y Geum, M Kim. How to identify promising chances for technological innovation: Keygraph-based patent analysis. *Advanced Engineering Informatics*, 2020, 46(2): 101155.
- [14] R Blake, R Ayyagari. Analyzing information systems security research to find key topics, trends, and opportunities. *Journal of Information Privacy and Security*, 2012, 8(3): 37-67.
- [15] J Zhang, W Yu. Early detection of technology opportunity based on analogy design and phrase semantic representation. *Scientometrics*, 2020, 125(1): 551-576.
- [16] S Choi, J Yoon, K Kim, et al. SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells. *Scientometrics*, 2011, 88(3): 863-883.
- [17] H Kwon, J Kim, Y Park. Applying LSA text mining technique in envisioning social impacts of emerging technologies: The case of drone technology. *Technovation*, 2017, 60-61: 15-28.
- [18] J Hou, X Yang. Patent sleeping beauties: evolutionary trajectories and identification methods. *Scientometrics*, 2019, 120(1): 187-215.
- [19] X Yuan, X Li. A network analytic method for measuring patent thickets: A case of FCEV technology. *Technological Forecasting and Social Change*, 2020, 156: 120038.
- [20] G Jin, Y Jeong, B Yoon. Technology-driven roadmaps for identifying new product/market opportunities: Use of text mining and quality function deployment. *Advanced Engineering Informatics*, 2015, 29(1): 126-138.
- [21] X Li, D Zhao, X Hu. Gatekeepers in knowledge transfer between science and technology: an exploratory study in the area of gene editing. *Scientometrics*, 2020, 124(2): 1261-1277.
- [22] Z Wang, A L Porter, X Wang, et al. An approach to identify emergent topics of technological convergence: A case study for 3D printing. *Technological Forecasting and Social Change*, 2019, 146(5): 723-732.
- [23] J Yun, Y Geum. Automated classification of patents: A topic modeling approach. *Computers & Industrial Engineering*, 2020, 147(Jan): 106636.
- [24] M Kim, Y Park, J Yoon. Generating patent development maps for technology monitoring using semantic patent-topic analysis. *Computers & Industrial Engineering*, 2016, 98: 289-299.
- [25] Z Li, D Tate, C Lane, et al. A framework for automatic TRIZ level of invention estimation of patents using natural language processing, knowledge-transfer and patent citation metrics. *Computer-Aided Design*, 2012, 44(10): 987-1010.
- [26] D D Sheu, M Yen. Systematic analysis and usage of harmful resources. *Computers & Industrial Engineering*, 2020, 145: 106459.
- [27] S Lee, B Yoon, Y Park. An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation*, 2009, 29(6-7): 481-497.
- [28] C Lee, G Lee. Technology opportunity analysis based on recombinant search: patent landscape analysis for idea generation. *Scientometrics*, 2019, 121(2): 603-632.
- [29] W Seo, J Yoon, H Park, et al. Product opportunity identification based on internal capabilities using text mining and association rule mining. *Technological Forecasting and Social Change*, 2016, 105: 94-104.
- [30] T Arciszewski. Morphological analysis in inventive engineering. *Technological Forecasting and Social Change*, 2018, 126: 92-101.
- [31] L Feng, Y Niu, J Wang. Development of morphology analysis-based technology roadmap considering layer expansion paths: application of TRIZ and text mining. *Applied Sciences*, 2020, 10(23): 84-98.
- [32] P Zheng, C Chen, S Shang. Towards an automatic engineering change management in smart product-service systems – A DSM-based learning approach. *Advanced Engineering Informatics*, 2019, 39: 203-213.
- [33] S Jun, S S Park, D S Jang. Technology forecasting using matrix map and patent clustering. *Industrial Management & Data Systems*, 2012, 122(5): 786-807.
- [34] L Lei, J Qi, K Zheng. Patent analytics based on feature vector space model: a case of IoT. *IEEE Access*, 2019, 7: 45705-45715.
- [35] C Sternitzke, A Bartkowski, R Schramm. Visualizing patent statistics by means of social network analysis tools. *World Patent Information*, 2008, 30(2): 115-131.
- [36] C Kim, H Lee, H Seol, et al. Identifying core technologies based on technological cross-impacts: An association rule mining (ARM) and analytic network process (ANP) approach. *Expert Systems with Applications*, 2011, 38(10): 12559-12564.
- [37] H Sung, H Yeh, J Lin, et al. A visualization tool of patent topic evolution using a growing cell structure neural network. *Scientometrics*, 2017, 111(3): 1267-1285.
- [38] B Song, J Luo, K Wood. Data-driven platform design: patent data and function network analysis. *Journal of Mechanical Design*, 2019, 141(2): 21101.
- [39] W J Lee, W K Lee, S Y Sohn. Patent network analysis and quadratic assignment procedures to identify the convergence of robot technologies. *PLoS One*, 2016, 11(10): e165091.
- [40] W S Lee, E J Han, S Y Sohn. Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. *Technological Forecasting and Social Change*, 2015, 100: 317-329.
- [41] H Sun, H Du, J Huang, et al. Detecting semantic-based communities in node-attributed graphs. *Computational Intelligence*, 2018, 34(4): 1199-1222.
- [42] V Bruno, D François. The cost factor in patent systems. *Journal of Industry, Competition and Trade*, 2009, 9(4): 329-355.
- [43] J Murphy, K Fu, K Otto, et al. Function based design by-analogy: a functional vector approach to analogical search. *Journal of Mechanical Design*, 2014, 136(10): 10110210.
- [44] N Preschitschek, H Niemann, J Leker, et al. Anticipating industry convergence: semantic analyses vs IPC co-classification analyses of patents. *Foresight (Cambridge)*, 2013, 15(6): 446-464.
- [45] G Al-Talib, H Hassan. A study on analysis of SMS classification using TF-IDF Weighting. *International Journal of Computer Networks and Communications Security*, 2013, 1(5): 189-194.
- [46] G V Georgiev, D D Georgiev. Enhancing user creativity: Semantic measures for idea generation. *Knowledge-Based Systems*, 2018, 151: 1-15.
- [47] S Sarica, J Luo, K L Wood. TechNet: Technology semantic network based on patent data. *Expert Systems with Applications*, 2020, 142: 112995.
- [48] Y Cai, Q Zhang, W Lu, et al. A hybrid approach for measuring semantic similarity based on IC-weighted path distance in WordNet. *Journal of intelligent information systems*, 2018, 51(1): 23-47.

- [49] N Zhang, Y Yang, J Su, et al. Modelling and analysis of complex products design based on supernetwork. *Kybernetes*, 2019, 48(5): 861-887.
- [50] J Zhu, J Zhu, S Ghosh, et al. Social Influence Maximization in Hypergraph in Social Networks. *IEEE Transactions on Network Science and Engineering*, 2019, 6(4): 801-811.
- [51] S G Aksoy, C Joslyn, M C Ortiz, et al. Hypernetwork science via high-order hypergraph walks. *EPJ Data Science*, 2020, 9(1): 16-49.
- [52] H Noh, Y Song, S Lee. Identifying emerging core technologies for the future: Case study of patents published by leading telecommunication organizations. *Telecommunications Policy*, 2016, 40(10-11): 956-970.
- [53] A Gambardella, M S Giarratana. General technological capabilities, product market fragmentation, and markets for technology. *Research Policy*, 2013, 42(2): 315-325.
- [54] W S Brüngrer, D Geiß, G Herlan, et al. Quality – Key factor for high value in professional patent, technical and scientific information. *World Patent Information*, 2011, 33(3): 230-234.
- [55] D Harhoff, S Wagner. The duration of patent examination at the european patent office. *Management Science*, 2009, 55(12): 1969-1984.
- [56] C A Cotropia, M A Lemley, B Sampat. Do applicant patent citations matter?. *Research Policy*, 2013, 42(4): 844-854.
- [57] D Chen, W C Lin, M Huang. Using essential patent index and essential technological strength to evaluate industrial technological innovation competitiveness. *Scientometrics*, 2007, 71(1): 101-116.
- [58] T Ramraj, R Prabhakar. Frequent subgraph mining algorithms – a survey. *Procedia Computer Science*, 2015, 47: 197-204.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
