Research on Core Technology Chain Identification Methods Based on Input-Output SAO Networks: A Case Study in Quantum Computing Postprint

Authors: Haiying Ren, Li Zhen

Date: 2023-04-01T16:02:56+00:00

Abstract

[Purpose/Significance] Identifying core technology chains in industrial technology fields based on patent literature holds positive significance for clarifying industrial core technology architecture, analyzing weak links in China's core industrial chains, determining directions for technological breakthroughs, and improving both technology chains and industrial chains. [Method/Process] This study improves the classic SAO structure and proposes a core technology chain identification method based on an input-output SAO network. Input-output SAO structures are extracted from domain patent texts as technical elements, and a domain knowledge network of technologies is constructed based on the input-output relationships between these technical elements; the weighted k-Core method is employed to obtain a core knowledge subnetwork containing major technical relationships, and the decomposition of the core knowledge subnetwork is achieved by identifying strongly connected components; main path analysis and other methods are utilized to identify external and internal core technology chains within the core knowledge subnetwork, and the core technology chains are interpreted with reference to the patents involved. [Results/Conclusion] The proposed method is applied to the quantum computing field, identifying the embedded external and internal core technology chains, and the accuracy of the results is verified through expert validation and comparison with related scholars' research. This identification method offers advantages such as coherent technology chain relationships, high degree of automation, and strong flexibility.

Full Text

Preamble

ChinaXiv Partner Journal, Vol. 65, No. 19, October 2021

Research on Core Technology Chain Identification Method Based on Input-Output SAO Network: A Case Study of Quantum Computing Technologies

Ren Haiying, Li Zhen

School of Economics and Management, Beijing University of Technology, Beijing 100124

Abstract: [Purpose/Significance] Identifying the core technology chain in industrial technology fields based on patent literature is crucial for mapping industrial core technology architectures, analyzing weak links in China's core industrial chains, determining technology research directions, and improving both technology and industrial chains. [Method/Process] This study improves upon the classic SAO structure and proposes a core technology chain identification method based on an input-output SAO network. Input-output SAO structures are extracted from domain patent texts as technical elements, and a domain knowledge network is constructed according to the input-output relationships between these elements. The weighted k-Core method is applied to obtain a core knowledge subnetwork containing the main technical relationships, and strong connected components are identified to decompose this subnetwork. Main path analysis and other methods are then used to identify external and internal core technology chains within the core knowledge subnetwork, with interpretation of these chains referencing the patents involved. [Result/Conclusion] The proposed method is applied to the quantum computing domain, identifying both external and internal core technology chains. The accuracy of the results is verified through expert validation and comparison with related research. The identification method offers advantages including coherent technology chain relationships, high automation, and strong flexibility.

Keywords: core technology chain; SAO structure; knowledge network; main path analysis; weighted k-Core

Classification Number: G255.53

DOI: 10.13266/j.issn.0252-3116.2021.19.012

1. Introduction

An industrial chain is a chain formed by related links from upstream to downstream in an industry [1]. The Fifth Plenary Session of the 19th CPC Central Committee proposed to "promote integrated innovation among upstream, midstream, and downstream enterprises of all sizes along industrial chains," while also "addressing weaknesses in industrial and supply chains, implementing industrial foundation reconstruction projects, intensifying efforts to tackle critical products and core technologies, developing advanced and applicable technologies, and diversifying industrial and supply chains." The industrial technology chain (referred to simply as "technology chain" in this paper) is the technical chain that runs through all links of the entire industrial chain and constitutes a necessary condition for forming the industrial chain [2]. The core technology chain is a series of linked key technical segments that support industrial activities and serves as the main guarantee for the operation and development of core industrial chains [3-4]. Particularly in the current complex international economic and trade environment, building a complete core technology chain has important strategic significance for ensuring China's industrial security and promoting integrated innovation across industrial chains. Consequently, research on technology chains and core technology chains has received increasing attention.

Existing research on technology chains primarily covers concepts, identification, structure, competitive landscape analysis, and evolution mechanisms of technology chains or core technology chains. Among these, identifying core technology chains in industrial technology fields (referred to as "technology fields" in this paper) holds significant practical value for analyzing weak links in China's core industrial chains, planning directions for key core technology research, and improving both technology and industrial chains.

This paper first summarizes relevant concepts of core technology chains and identification methods in the intelligence field, analyzes the limitations of existing methods, and then proposes a new method for identifying core technology chains based on SAO networks. This method systematically identifies core technology chains in a given technology field based on patent literature, providing a new methodological foundation for in-depth analysis of core technology and industrial chains.

2. Related Work

2.1 Concepts of Technology Chain and Core Technology Chain

In this paper, the technology chain refers to the industrial technology chain, which primarily describes relationships among technologies within an industry. Yuan Deyu [5] was among the first to propose the concept of industrial technology chain, viewing it as a synthesis of multiple production technologies and a systematization of production technology. Liu Qi and Ding Yunlong [6] defined the industrial technology chain as a technical chain that is interdependent and interconnected both within and between industries. Gao Ruxi et al. [7] proposed two different descriptions of technology chains: first, the succession relationship of technologies themselves, where "the acquisition and use of one technology must be premised on the acquisition and use of another technology, thus forming a linking relationship between related technologies"; second, a technical linking relationship derived from the upstream-downstream relationship between products, where "various technologies embodied in upstream and downstream products form a technology chain based on the linking relationship between products." Huang Liye et al. [8] proposed that the industrial technology chain is a technical chain connecting upstream, midstream, and downstream links of an industrial chain.

Regarding core technology chains, Hong Yong and Su Jingqin [3] defined the core technology chain as technical links containing key technical segments in industrial activities. Zhu Ruibo [9] viewed the core technology chain as the innovation and R&D activities for the core value chain of an industry, dividing it into three segments: key manufacturing technologies, core component technologies, and product architecture technologies.

Although these definitions of technology chains differ, they all emphasize the succession relationship between various technical segments in a technology chain, with the core technology chain being the technically central chain within the technology chain.

2.2 Research on Core Technology Chain Identification Methods

Traditional methods for identifying core technology chains rely on expert experience. For example, Fang Si and Li Guoqiu [10] constructed a model for the new energy vehicle industrial chain and technology chain after reviewing analyses by several scholars, identifying the industry's core technology chain by determining key technologies used at different industrial layers. The advantage of the expert experience method lies in its strong professionalism and reference value. However, its disadvantages include poor automation and limitations due to experts' own knowledge domains, leading to partiality and subjectivity. Even with multiple experts, inconsistencies in the granularity or scope of core technology chain descriptions may result in overlapping, missing, or even contradictory identifications. Therefore, scholars have recently begun to adopt objective quantitative methods such as patent analysis or combine qualitative and quantitative approaches to identify core technology chains.

2.2.1 Qualitative and Quantitative Combined Methods Qualitative and quantitative combined methods for core technology chain identification typically begin with expert assistance to identify main technical segments of an industry, followed by in-depth statistical and quantitative analysis of patents or other scientific and technical intelligence to find core technologies within each segment. Fang Si and Li Guoqiu [10] identified leading enterprises, core technologies, and technology branches at each link of the technology chain through statistical analysis of patents in the driverless vehicle industry. Wu Feifei et al. [11] conducted a systematic analysis of China's power battery industry from the perspective of industrial technology chains, analyzing technological weak links through core technology analysis of the industrial technology chain. Yuan Xiaodong and Bao Yewen [12] obtained an electronics and communication equipment manufacturing technology chain centered on ZTE Corporation by analyzing its upstream and downstream partners, then identified core IPC classification numbers in the IPC co-occurrence network at each link of the technology chain to derive core technologies.

2.2.2 Patent Citation-Based Core Technology Chain Identification Patent citation-based core patent chain identification methods construct patent citation networks based on citation relationships between patents, identifying core patent chains containing core technology diffusion paths by analyzing important paths in the citation network. Y.M. Wei et al. [13] constructed a patent citation network in the shale gas technology field and used FCFP main path analysis to identify key technology development paths, obtaining a core patent chain constructed from key technologies. Zhao Rongying et al. [14] argued that patent citation relationships can reflect technology inheritance and development, thus they built a multi-level citation network between patents to reflect technology aggregation and diffusion, using citation frequency as a measure of patent importance to obtain important technology development path chains. Qi Yun et al. [15] constructed a patent citation network in the graphene field, applying multiple main path analysis algorithms to identify core patent chains.

2.2.3 SAO-Based Technology Tree or Roadmap Identification In recent years, SAO structures have been widely used to analyze semantic relationships between technical concepts in patent texts. The classic SAO structure includes Subject (S), Action (A), and Object (O). Some scholars extract SAO structures from patent texts to construct technology trees, technology morphology matrices, or technology roadmaps based on relationships between technical concepts in SAO structures, providing new analytical tools for core technology chain identification. S. Choi et al. [16] calculated similarity between SAO structures, clustering them into four categories—product, technology, function, and technical attribute—and divided AO into three types: inclusion, effect, and attribute, to build a technology tree showing relationships among products, technologies, and functions. Zhai Dongsheng et al. [17] added an "object complement" component to the SAO structure and annotated SAO semantic features as product, function, technical attribute, scientific effect, and efficacy to construct technology trees for technical fields. Guo Junfang et al. [18] distinguished technical categories of key concepts in technical fields by part of speech, constructing SAO structure chains based on semantic and hierarchical relationships among key concepts to build morphological structures of technical fields. C. Yang et al. [19] treated subjects (S) and objects (O) in SAO structures as network nodes, using the quantity and importance of action relationships (A) between nodes as edge weights to construct SAO networks for analyzing technology development trends. Lai Chao'an et al. [20] divided SAO structures into four categories—theme, solution, structure, and effect—based on semantic features, and considered technical hierarchies between SAO structures to construct technology trees in the form of SAO structure chains, linking them based on similarity to build patent networks. Li Xiaoman et al. [21] constructed a streamlined technical element identification framework, classifying technical elements into five categories—material, product, method, function, and use—based on positional, syntactic, verb, and semantic features to clarify the structure of technical fields.

2.3 Review of Existing Methods

The above core technology chain identification methods actually identify different forms of core technology chains, as shown in Figure 1 [Figure 1: see original paper]. We compare these methods in combination with Figure 1:

2.3.1 Qualitative and Quantitative Combined Methods In Figure 1, (a) represents the core technology chain identified by this method, where X1-X5, Y1-Y4, and Z1-Z5 represent technologies related to upstream, midstream, and downstream products, respectively. This method can construct an overall industrial core technology chain, but it essentially identifies technical linking relationships derived from upstream-downstream relationships between products [7], with each core technology segment appearing as a technology or patent cluster centered on products. The connections between core technology segments are not sufficiently tight or direct. Additionally, this method relies heavily on expert experience.

2.3.2 Patent Citation-Based Methods

(b) represents the core technology chain identified by this method, where P1-P5 represent different patents, with arrows pointing from cited patents to citing patents. Using a specific partial citation network as an example, consider patent “US2011142242-A1” (P1), which primarily describes quantum public key generation technology. Among its citing and cited patents, patent “US9736147-B1” (P4) is a “method for authenticating a user of a client device on a server,” involving a specific application of key generation technology and belonging to downstream technology of P1. Patents “WO2008142816-A1” (P2), “JP2002055606-A” (P3), and “WO2016164325-A1” (P4) describe technologies for generating quantum keys through inverse calculation of quantum gate matrices using predetermined states in qubit space, generating public keys using algebraic equations, and generating keys through quantum-resilient random number generation, respectively—all belonging to the same type of technology as P1. Thus, while core patent chain identification based on citation features can effectively identify core patents in a technology field and knowledge flow and technology evolution between them, this knowledge flow relationship is not necessarily the succession relationship of technologies themselves [7], but more likely relationships between similar types of technologies. Therefore, this type of core technology chain struggles to reflect upstream-downstream succession relationships between technologies.

2.3.3 SAO-Based Methods

(c) represents the technology tree based on SAO structures [16], where A-G are technical elements and C1-G2 are branches under each technical element. Based on the semantic information contained in SAO elements, technical elements can be divided into various types such as technology nodes, product nodes, and function nodes. While technology trees built using SAO can classify technical elements into specific types, they struggle to comprehensively display upstream-downstream succession relationships between technologies and construct technology chains at the overall domain level. This is because subdividing technical elements leads to rigid role assignment, ignoring the common relationship where outputs from upstream in a technology chain (such as products) can serve as inputs for downstream (such as technologies), making it difficult for relationships between technical elements to 贯通 the entire technology field.

2.3.4 Our Method

(d) represents the target core technology chain of the identification method proposed in this paper, where A-N represent different technical elements and arrows indicate input-output relationships between technical elements. The chain "A→D→E→F→J→K" in the figure represents one core technology chain that can run through upstream, midstream, and downstream technologies (and products) of a technology chain, clearly displaying succession relationships between technologies. This paper adopts this form as the identification target for core technology chains.

3. Research Design

3.1 Overall Approach

Addressing the limitations of existing core technology chain identification methods, this paper proposes a core technology chain identification method based on an input-output SAO network, built upon SAO structures and network analysis methods. The overall approach is as follows:

3.1.1 Design Input-Output SAO Structures to Describe Succession Relationships Between Technical Elements In classic SAO structures, the roles of S and O elements and their relationships are not explicitly defined. In recently proposed SAO structures, S and O roles are fixed as product, technology, attribute, effect, etc. Neither type of SAO structure facilitates clear and 贯通 identification of relationships between upstream and downstream technical elements in a technology field. This paper designs a new SAO structure called the "input-output SAO structure." The input-output SAO structure broadly categorizes technical elements (S and O) as "input elements" and "output elements," and relationships between technical elements (A) as "input-output relationships," using "input-output relationships" to represent upstream-downstream succession relationships in technology chains. Specifically, this study extracts inclusion, causal, conditional, directional, and processing relationships between S and O based on the part of speech (where A is not limited to verbs but can also be prepositions, etc.) and semantics of A in extracted SAO structures. This allows the original "S→A→O" relationship to be transformed into "input element→processing element→output element," or "Input (I)→Process (P)→Output (O)," yielding the input-output SAO structure (referred to as "IPO structure" in this paper), where I and O are upstream and downstream technical elements, respectively, and P represents the technical succession relationship.

3.1.2 Construct Domain Knowledge Networks to 贯通 Upstream and Downstream of Technology Fields In IPO structures, I and O are not fixed roles; the I element of one IPO structure may be the O element of an upstream technology. Therefore, when a technology field develops to a certain stage, all patent IPOs can form an interconnected knowledge network. Drawing on SAO network construction methods [19], this paper uses I and O as nodes, uses the frequency of IO relationships in the technology field as importance weights, and constructs a directed, weighted input-output SAO network called the "domain knowledge network," which serves as the "matrix" of core technology chains.

3.1.3 Simplify and Decompose Domain Knowledge Networks to Obtain Most Important Technical Elements and Relationships The domain knowledge network obtained through the above method is large-scale and may contain interdependent states where technical elements serve as each other's inputs and outputs (manifested as "loops" in the network), making it difficult to present clear chains from upstream to downstream. This prevents the application of main path analysis, commonly used in patent citation networks to identify core technology flows. To address this, this paper uses the weighted k-Core method to simplify the domain knowledge network, discarding most secondary technical elements and loops to extract the connected subnetwork with the largest edge weights. This subnetwork contains the most frequently occurring technical elements and their technical relationships in all patents and is referred to as the "core knowledge subnetwork." However, even this simplified subnetwork may still contain loops, making core technology chain identification difficult. Each loop in the core knowledge subnetwork corresponds to a strongly connected component, which should represent the interdependent relationships (non-upstream-downstream relationships) among knowledge elements within it. By identifying strongly connected components, the core knowledge subnetwork can be decomposed into a loop-free "external core knowledge subnetwork" and one or more loop-containing "internal core knowledge subnetworks," and the upstream-downstream relationships of the external core knowledge network can be determined based on the "input-output" relationships between strongly connected components.

3.1.4 Use Main Path Analysis and Other Methods to Analyze Core Knowledge Subnetworks and Obtain Core Technology Chains Main path analysis is an important analytical tool for citation networks, excelling at analyzing the dissemination and diffusion of mainstream knowledge and recently applied to identifying key nodes in information diffusion networks [22]. The limitation of main path analysis is its unsuitability for networks with loops. Therefore, this paper extends the application of main path analysis to networks with loops through preprocessing using weighted k-Core and strongly connected component extraction. Specifically, after obtaining the external core knowledge subnetwork (which no longer contains loops, with each loop collapsed into a representative node) and ensuring upstream-downstream 贯通 of the entire technology field, main path analysis is used to identify the most important "external core technology chain" in the external core knowledge subnetwork. Finally, internal core knowledge subnetworks (loops) may reflect more complex interdependencies among technologies in certain segments, and the above simplification, decomposition, and main path analysis methods are applied again to identify "internal core technology chains."

In summary, this paper constructs an IPO structure-based domain knowledge network for a specific industrial technology, identifies and interprets the technology's core technology chains from the domain knowledge network using weighted k-Core and main path analysis methods. The main technical process includes data acquisition and text preprocessing, domain knowledge network construction, and core technology chain identification and interpretation, as shown in Figure 2 [Figure 2: see original paper].

3.2 Specific Steps

3.2.1 Data Acquisition and Text Preprocessing This process consists of two sub-processes:

(1) **Data acquisition and cleaning sub-process.** This paper uses the Derwent Innovations Index patent database (DII database) as the data source. Search expressions are formulated based on knowledge of the technology field to retrieve and download relevant patent texts from the DII database. Patent documents are parsed and irrelevant information is removed.

(2) **Text preprocessing sub-process.** Patent texts (abstracts in this paper) are extracted to build a domain patent text corpus. Patent texts are then normalized through steps such as anaphora resolution and sentence segmentation to complete text preprocessing.

3.2.2 Domain Knowledge Network Construction This process consists of two sub-processes: IPO structure extraction and domain knowledge network construction.

(1) **IPO structure extraction.** This sub-process consists of three steps: syntactic analysis, IPO structure triplet extraction, and IPO structure

direction annotation and normalization.

Step 1: Syntactic analysis. Extract phrase structure trees from single sentences to represent constituent structures and describe dependency relationships between words or phrases. Using the sentence “The quantum computing state detection system comprises a shelving laser to transfer any ytterbium ion from an initial state to a shelved state.” as an example, Figure 3 [Figure 3: see original paper] shows the phrase structure tree for this sentence.

Step 2: IPO structure triplet extraction. First, based on the phrase structure tree obtained in the previous step, regular expressions are used to extract verbs, prepositions, and conjunctions as processing elements (P) (adjacent P elements should be merged). Second, starting from the root node of the phrase structure tree, the nearest parent and child nodes corresponding to NP chunks are searched based on P elements and recorded as NP1 and NP2 chunks, serving as input elements (I) or output elements (O). The [NP1, P, NP2] obtained at this point is the IPO structure triplet. Table 1 shows the IPO structure triplets extracted from Figure 3.

Step 3: IPO structure direction annotation and normalization. Based on the semantics of P elements, an “IPO structure direction annotation rule table” is constructed (see Table 2), and regular expressions are used to annotate the direction of IPO structure triplets extracted in the previous step. The relationship between I and O elements in triplets is often modifying or limiting (such as relationships containing prepositions like in, of, etc.), so these are removed. After direction annotation, IPO structures are normalized, including lemmatization, converting initial letters to lowercase, and removing meaningless stopwords unrelated to technical content such as “a” and “the.” The final IPO structure for the example sentence is shown in Table 3 .

(2) **Domain knowledge network construction.** This sub-process uses I and O elements obtained above as nodes in the knowledge network, P elements as edge attributes, and I pointing to O as edge direction to construct a directed, weighted domain knowledge network. Specific steps are as follows:

Step 1: Obtain edge list for a single patent. For a single patent, IPO structures for each sentence are obtained through the IPO extraction step above to generate the patent’s IPO edge list, after which duplicate input-output technical relationships (duplicate edges) within the single patent are removed.

Step 2: Calculate edge weights for the IPO network. Edge lists for all patents in the technology field are aggregated to obtain the domain edge list. Edges with identical I and O elements are merged, with P elements aggregated as edge attributes and the main accession numbers of patents where IPO structures are located also recorded as edge attributes to facilitate later interpretation of core technology chains in combination with patent information. The number of P elements contained in each edge in the IPO network is used to measure the importance of the edge, serving as the edge weight.

Step 3: Prune bidirectional edges. After weighting, the IPO network may contain "bidirectional edges," where two nodes serve as each other's input and output elements, which is detrimental to technology chain derivation and analysis. Therefore, this step prunes these bidirectional edges by removing the direction with smaller edge weight based on the relative importance of input-output relationships between the two nodes. If both directions of a bidirectional edge have the same weight, both directions can be retained. After these three steps, domain knowledge network construction is completed.

3.2.3 Core Technology Chain Identification and Interpretation This process consists of four sub-processes: core knowledge subnetwork construction, core knowledge subnetwork decomposition, external core technology chain identification and interpretation, and internal core technology chain identification and interpretation.

(1) **Core knowledge subnetwork construction.** The goal of this sub-process is to simplify the domain knowledge network and extract core technical elements and relationships in the technology field. Specific steps are as follows:

Step 1: Export the maximum connected subnetwork, removing other less important small connected subnetworks to serve as the main knowledge elements for core technology chains.

Step 2: Obtain the core knowledge subnetwork through the weighted k-Core algorithm. The k-Core algorithm is a method for finding tightly connected subnetwork structures that meet specified coreness degrees in networks. In classic k-Core, each vertex has at least degree k and all vertices are connected to at least k other nodes in the subnetwork, emphasizing node importance. However, the core technology chain is the technology segment with the greatest value added and is indispensable in the industrial chain. From a global perspective, technologies involved in core technology chains inevitably face the most intense R&D competition, and patents are typically most concentrated. The more common technical connections revealed by these patents, the greater the weight in the knowledge network, and the more likely they are to become edges in core technology chains. Therefore, this paper uses the weighted k-Core algorithm that emphasizes edge importance, i.e., by setting a maximum edge weight threshold e, each vertex in the k-Core has at least one adjacent edge with weight greater than e, thereby obtaining a domain core knowledge subnetwork that retains the most important technical relationships.

(2) **Core knowledge subnetwork decomposition.** The goal of this sub-process is to decompose the core knowledge subnetwork into external and internal core technology subnetworks to prepare for core technology chain identification and interpretation. Specific steps are:

Step 1: Obtain the set of strongly connected components in the core knowledge subnetwork. Strongly connected components are exported to obtain the set {T} of strongly connected components in the core knowledge subnetwork for later specific identification and interpretation of technical content in each component. Assume a core knowledge subnetwork of a certain domain as shown in Figure 4 [Figure 4: see original paper]. Since nodes B, C, D, and E have strong connectivity relationships, the strongly connected component set {A, {B, C, D, E}, F, G} can be obtained.

Step 2: Decompose the core knowledge subnetwork. Each strongly connected component is collapsed into a node, with one representative node selected as a label. The collapsed domain core knowledge subnetwork is called the "external core knowledge subnetwork," describing technical knowledge at the overall domain level. Strongly connected components with multiple nodes are called "internal core knowledge subnetworks," describing interdependent internal technical processes in detail. For example, the strongly connected component {B, C, D, E} in Figure 4 is collapsed into a node (labeled #B), which together with A, F, and G forms the external core knowledge subnetwork (see Figure 5 [Figure 5: see original paper]), while {B, C, D, E} and its edges constitute the internal core knowledge subnetwork.

(3) **External core technology chain identification and interpretation.** This sub-process analyzes the structure of the external core knowledge subnetwork, identifies external core technology chains, and then interprets the specific content of each chain. It consists of the following steps:

Step 1: Analyze the structure of the external core knowledge subnetwork. After removing strongly connected components, the structure of the external core knowledge subnetwork becomes clearer, generally presenting one or more chain or tree structures. As shown in Figure 5, the external core knowledge subnetwork can be divided into three parts based on the input-output situation of each node: input section, transformation section, and output section (possibly more). The input section contains the source node set of the subnetwork, mainly describing principles and components of the technology field; the transformation section is typically composed of several strongly connected components, located in the middle or core position of the technology field, mainly describing operations and processes of internal core technologies; the output section contains the sink node set of the subnetwork, mainly describing functions and results of the technology field.

Step 2: Identify external core technology chains. Based on the structure of the external core knowledge subnetwork, the relationships between principles and components, operations and processes, and functions and results in the input and output sections are analyzed, and main path analysis methods are used to identify external core technology chains. First, global main path analysis is applied to obtain the global main path of the external core knowledge subnetwork as one core technology chain. This chain has the highest total edge weight among all external core technology chains, thus describing relatively important technical relationships and representing the mainstream technical process in the technology field. We call it the "mainstream external core technology chain."

Second, key main path analysis methods [23] are used to treat edges with higher weights in the external core knowledge subnetwork as key edges to search for and identify other major external core technology chains, which we call “major external core technology chains.”

Step 3: Interpret external core technology chains. Mainstream and major external core technology chains are interpreted in combination with source patents for important edges in each chain. First, IPO structures with larger edge weights in each external core technology chain are located, and the correspondence between I and O is determined based on P elements in IPO structures. Then, each external core technology chain is interpreted in depth based on the original abstracts of source patents for IPO structures.

(4) **Internal core technology chain identification and interpretation.** This sub-process applies the above analytical methods nestedly to internal core knowledge subnetwork S to identify and interpret internal core technology chains. It consists of four steps:

Step 1: Referring to Section 3.2.3 (1), increase the e value to extract the weighted k-Core of S to obtain a more core technical subnetwork, denoted as S .

Step 2: Referring to Section 3.2.3 (2), obtain the strongly connected component set of S and collapse S .

Step 3: Referring to Section 3.2.3 (3), identify the “external” core technology chains of S through main path analysis and interpret them based on source patents for each core technology chain.

Step 4: Further analyze and interpret the strongly connected component set of S to identify and interpret important “internal” core technology chains.

4. Case Study

Quantum computing is a new type of computing that follows quantum mechanics principles for high-speed computation, storage, and information processing. It has strong storage capabilities and fast computing speeds, bringing a qualitative leap to existing computing power [24]. China also attaches great importance to quantum computing technology development, including it in national development plans. Today, quantum computing is developing rapidly, with increasing attention from countries, enterprises, and R&D institutions. Currently, quantum computing is still in a critical stage of technology verification and prototype development [25]. Therefore, identifying core technology chains in the quantum computing field can help technology management departments construct the core technology architecture of quantum computing and help research institutions identify technological weak links and determine research directions, which has practical significance.

4.1 Data Acquisition and Text Preprocessing

This paper selects patent data from the DII database as the data source to identify core technology chains in the quantum computing field. The Derwent Manual Code (DMC) is assigned by Derwent index experts based on patent abstracts and full texts to reveal and express technology categories, which is more specific than International Patent Classification (IPC) codes. Therefore, this paper uses DMC as the search field. After reviewing relevant literature, the DMC for quantum computing is determined to be T01-E05Q, and the search expression "MAN=T01-E05Q" is formulated. Data retrieval was conducted on January 2, 2021, yielding 2,303 patents in the quantum computing field since 1998. Figure 6 [Figure 6: see original paper] shows the annual patent application trends in this field. According to statistics on annual patent application changes, the number of patent applications has shown an increasing trend since 2002, with slow growth before 2015 and rapid development after 2015, indicating that quantum computing technology is receiving increasing attention. Statistics on patent application countries/regions show that China ranks second only to the United States in patent applications, indicating good R&D strength and market prospects for China in the quantum computing field.

Python and its natural language toolkit were then used to write code to extract patent abstract texts, remove irrelevant patent information, and clean data. Sentence segmentation, anaphora resolution, and other steps were performed to normalize patent texts and obtain the domain patent abstract corpus.

4.2 Domain Knowledge Network Construction

After obtaining the quantum computing domain patent abstract corpus, IPO structures in this technology field were extracted following the IPO structure extraction steps described above, with the assistance of the Stanford NLP Parser tool, NLTK toolkit, and regular expressions.

Next, the edge list for a single patent, calculation of IPO network edge weights, and pruning of bidirectional edges were performed to construct the quantum computing domain knowledge network. The network includes 22,782 nodes and 29,504 edges.

4.3 Core Technology Chain Identification and Interpretation

4.3.1 Core Knowledge Subnetwork Construction and Decomposition

Complex network analysis software (such as Pajek) was used to analyze the quantum computing IPO knowledge network, revealing 1,954 weakly connected components. Except for the largest connected subnetwork, other weakly connected components are small and contain few knowledge elements. Therefore, the largest connected subnetwork was selected to identify core technology chains. This largest connected subnetwork contains 18,080 nodes and 26,490 edges. The weighted k-Core of this network was then extracted (edge weight threshold e set to 5) to obtain the quantum computing domain core knowledge subnetwork (see Figure 7 [Figure 7: see original paper]), which has 69 nodes and 189 edges, with each node having at least one adjacent edge with weight greater than 5.

Second, strongly connected component analysis was performed on the core knowledge subnetwork, yielding 37 strongly connected component sets (see Figure 7 [Figure 7: see original paper]). Among them, 36 components are single nodes, with only the 33 nodes in the dashed box forming the largest strongly connected component. The technologies they contain are interdependent from an input-output perspective, all related to quantum computing device principles and structures. Therefore, “#device” is used as the label for this strongly connected component (internal core knowledge subnetwork). The collapsed external core knowledge subnetwork is shown in Figure 8 [Figure 8: see original paper] (numbers on edges represent edge weights).

4.3.2 External Core Technology Chain Identification and Interpretation The external core knowledge subnetwork in Figure 8 [Figure 8: see original paper] can be divided into three parts based on the input-output situation of each node: input section, transformation section, and output section. External core technology chains containing IPO relationships are obtained through main path analysis, which are then interpreted in detail in combination with patents involved in each chain.

First, global main path analysis yields one “mainstream external core technology chain” (shown in bold in Figure 9 [Figure 9: see original paper]). Key main path analysis is then used to find key main paths containing edges with top 15 weights. Since these main paths share many overlapping edges, they are merged for convenience to form the “major external core technology chain,” shown in Figure 9 [Figure 9: see original paper].

Interpretation of external core technology chains is divided into two parts. First, the mainstream external core technology chain is interpreted. The IPO structures contained in this chain and their source patents are shown in Table 4
.

Based on this table, the mainstream external core technology chain can be analyzed and interpreted as follows:

(1) Edge “quantum well layer→quantum well stack” describes the relationship between quantum well layers and quantum well stacks. Combined with processing elements of this edge and source patents, this edge mainly describes the physical structure of quantum well stacks, i.e., quantum well stacks are composed of quantum well layers.

(2) Edge “quantum well stack→#device” describes the relationship between quantum well stacks and quantum computing devices. Combined with processing elements and source patents, this edge mainly describes the physical structure of quantum dot devices, which are composed of quantum well stacks, doping layers, and barrier layers. Therefore, quantum

well stacks are components of quantum dot devices.

(3) Edge "#device→problem" describes the relationship between quantum computing devices and problems. Processing elements in IPO structures containing this edge include words expressing "complete" or "solve" such as "to," "complete," "to solve," etc. Combined with source patents, this edge mainly describes problems that quantum computing devices can solve, including optimization problems of quantum computing device physical structures and mathematical optimization problems such as integer programming, combinatorial optimization, big data search optimization, and multivariate quadratic problems.

In summary, the upstream of the mainstream external core technology chain is the physical composition structure of quantum dot devices, including quantum well stacks composed of quantum well layers; the midstream mainly describes internal technologies of quantum computing devices; and the downstream mainly describes applications and functions of quantum computing devices.

Second, following the interpretation steps for the mainstream external core technology chain, these major external core technology chains are interpreted together. As shown in Figure 9 [Figure 9: see original paper], almost all main paths start from the input section of the external core technology subnetwork, pass through "#device" in the transformation section, and reach the output section. Due to space limitations, the technical content of major external core technology chains is summarized into two aspects:

(1) Sub-chains in the input section mainly describe the physical structure or working principles of quantum computing devices. Among them, the sub-chain "random quantum random number→source security key→binarization" describes the principle and physical structure of quantum computing devices; "strong spatial localization→#device" describes the spatial characteristics of quantum dot devices; "electrical connection→#device" describes the integration method of quantum dot devices; and "quantum well layer→quantum well stack→#device" (on the global main path) describes the physical structure of quantum dot devices.

(2) Sub-chains in the output section mainly describe the uses of quantum computing devices, including uses in internal optimization of quantum computing devices and functions they can achieve. Among them, sub-chain "#device→security" describes its function in improving network security; "#device→non-optimization calculation" describes its function in solving non-optimization calculation problems; "#device→problem" (on the global main path) describes quantum computing device internal physical structure optimization problems and various complex mathematical optimization problems it can solve; and "#device→impact" describes its uses in internal optimization of quantum computing devices.

It should be noted that some main paths obtained through the above main path analysis may not appear to be core technology chains. For example, edges from the "uses" node to "cellular phone," "handheld," "word processing," and "personal computer" in Figure 9 [Figure 9: see original paper] mainly describe uses of processors and other devices in personal computers, portable phones, etc., while "uses→#device" mainly describes functions of these devices in quantum computing devices. The former does not involve core quantum computing technologies, while the latter relates to physical composition structures in quantum computing devices.

4.3.3 Internal Core Technology Chain Identification and Interpretation After identifying and interpreting external core technology chains in the quantum computing field, the internal core knowledge subnetwork (the "#device" component) is analyzed to identify and interpret internal core technology chains. First, the weighted k-Core of the "#device" component is extracted again (edge weight threshold e increased to 9) to obtain a more core knowledge subnetwork. Then, strong connected component analysis is performed, and strongly connected components are collapsed to obtain the "external" core knowledge subnetwork S of the "#device" component and two "internal" core knowledge subnetworks "#processor" and "#quantum processing device," denoted as S1 and S2 , as shown in Figure 10 [Figure 10: see original paper].

Third, key main path analysis is used to identify main paths in S containing edges with top 3 weights, yielding three "major internal core technology chains" (shown in bold in Figure 10 [Figure 10: see original paper]).

These three major internal core technology chains are interpreted based on their source patents:

1. (1) Core technology chain "#processor→device→use" describes the physical structure of quantum dot devices or quantum control devices in quantum computing devices upstream, and describes functions of these devices in quantum computing devices downstream.
2. (2) Core technology chain "#processor→operation→#quantum processing device" describes operations that quantum processors can perform upstream, and describes impacts of these operations downstream, including quantum decoherence and scattering effects.
3. (3) Core technology chain "cool apparatus→#quantum processing device" mainly describes functions of cooling equipment in quantum computing devices, including keeping quantum processors at low temperatures to avoid qubit decoherence and reduce scattering effects.

Finally, internal core technology chains contained in the two "internal" core knowledge subnetworks S1 ("#processor" component) and S2 ("#quantum processing device" component) are identified and interpreted:

1. (1) The main internal core technology chain in S1 ("#processor" component) is "qubit decoherence→effect," mainly describing the function of cooling

equipment in avoiding scattering effects within quantum computing devices.

(2) The main internal core technology chain in S2 ("#quantum processing device" component) is "processor→computer executable component→memory→system," which describes quantum processor execution technology for computer executable components upstream, computer executable component instruction storage in memory midstream, and memory as a component of quantum computing systems downstream.

4.3.4 Result Validation Combining external and internal core technology chain identification results, this paper summarizes the core technology chains in the quantum computing field (see Figure 11 [Figure 11: see original paper]). The core technology chains in the quantum computing field are divided into upstream, midstream, and downstream sections, where the upstream mainly describes related principles or physical structures of quantum computing devices, the midstream mainly describes internal technologies of quantum computing devices, and the downstream mainly describes applications and functions of quantum computing devices.

To validate the results, the project team presented the identification results (primarily Figures 9-11) to two professors in the quantum computing field. Both considered the identified core technology chains to be relatively complete and accurately reflecting the main technologies and their relationships in the current quantum computing industry. One professor noted that certain quantum computer technology routes (such as superconducting, ion trap, and photonic) were not reflected in the identified core technology chains. Through discussion with this professor, it was found that these technology routes appear more frequently in high-level papers but less in patents, and multiple technology routes have not yet converged into mainstream technologies, thus not appearing in current core technology chains. Additionally, the authors reviewed literature summarizing quantum computing technology development and evaluated the feasibility and effectiveness of the proposed core technology chain identification method through comparison with research results.

(1) Upstream core technology chains in the quantum computing field describe principles or physical structures of quantum computing devices, including principles for improving network security, integration methods between quantum dot devices, and their spatial characteristics. Among them, the principle for improving network security in quantum computing devices mainly involves "quantum key distribution technology" for asymmetric cryptographic systems. According to related research, this technology is considered key to improving quantum communication confidentiality and belongs to key technologies in quantum computing [26-27]. Quantum dot devices are considered one way to achieve quantum computing [26, 28-30], and quantum dot materials and devices are currently among the most cutting-edge research topics internationally. The quantum well materials

included in their physical composition are foundational materials for new solid-state quantum devices [31]. Additionally, integration methods [29] and spatial characteristics [32] are also considered key research areas in quantum computing.

(2) Midstream core technology chains in the quantum computing field mainly introduce internal technologies of quantum computing devices, including quantum memory and quantum processor (quantum chip) data processing technologies, as well as cooling equipment-related technologies. The physical implementation of quantum processors and data processing technologies are both key technologies in quantum computing [24-25, 31]. Current solid-state quantum chips also face serious decoherence issues, which are difficulties in achieving quantum computers [30, 33]. Reducing quantum computing device temperatures and maintaining long-term quantum state coherence are currently the biggest technical bottlenecks [28, 33], making cooling equipment-related technologies a key research focus.

(3) Downstream core technology chains in the quantum computing field mainly introduce applications of quantum computing devices. According to related research, quantum computing applications mainly focus on large-scale data processing [33], information security protection [28, 33], and optimization problem solving [26, 33], which are highly consistent with the analysis results of this paper.

In summary, the core technology chains identified in the quantum computing field in this paper are highly consistent with results from related research.

5. Conclusion

Identifying core technology chains in industries is theoretically and practically significant for building complete technology and industrial chains. However, existing methods are insufficient in automatically and completely identifying core technology chains. This paper proposes a core technology chain identification method based on an input-output SAO network, using input-output SAO structures extracted from patent texts as technical elements to construct a domain knowledge network. Weighted k-Core algorithms are used to obtain core knowledge subnetworks containing important technical relationships, and strongly connected component extraction and main path analysis are applied to identify external and internal core technology chains, which are then interpreted based on patent texts. Using the quantum computing field as an example, core technology chains in this domain are identified and interpreted based on patent texts. Comparison with professional research results in the quantum computing field confirms the effectiveness of the proposed method.

The main advantages of the proposed core technology chain identification method are: (1) The input-output SAO structure better clarifies succession relationships between technologies compared to classic SAO structures, making it more suitable for core technology chain identification and analysis; input-output SAO structure extraction has high automation and significantly reduces expert involvement. (2) The domain knowledge network uses frequency of technical relationships in the domain as edge weights, enabling identification of core technology chains at different levels (mainstream and major) based on importance, offering strong flexibility. (3) Through preprocessing using weighted k-Core algorithms and strongly connected components, the application domain of main path analysis is extended to networks with loops.

This paper has some limitations. First, patent abstracts cannot fully capture technical details. Second, current extraction rules for input-output relationships between technical elements still have some ambiguity and multiplicity. Third, relying solely on patent data identifies relatively mature current core technologies, while more frontier core technology chains may be contained in scientific papers and other data sources. Future research will improve core technology chain identification accuracy by introducing heterogeneous data to expand technology data sources and perfecting extraction rules for input-output SAO structures. Additionally, more in-depth analysis of identified core technology chains will be conducted, such as competitiveness comparison and dynamic evolution analysis of technology chains, to better serve the improvement of China's industrial technology level.

References

[1] Wu Jinming, Shao Chang. Research on the Formation Mechanism of Industrial Chains: The “4+4+4” Model [J]. China Industrial Economics, 2006(4): 36-43.

[2] Zhao Junli, Zhong Jie, Wang Lixin. Research on Development Strategies for Strategic Emerging Industrial Clusters from the Perspective of “Three Chains”: A Case Study of Xining Carbon Fiber Industrial Cluster [J]. Qinghai Social Sciences, 2020(5): 96-105.

[3] Hong Yong, Su Jingqin. Research on Coordinated Development of Core Technology Chains and Core Industrial Chains in Developing Countries [J]. China Industrial Economics, 2007(6): 38-45.

[4] Chen Si, Zhao Yuxiang, Zhu Qinghua. Analysis of Industrial Technology Competitive Intelligence Service Model Based on Technology Chain [J]. Information Studies: Theory & Application, 2020, 43(5): 31-37.

[5] Yuan Deyu. On Industrial Technology [J]. Journal of Northeastern University (Social Science), 2000(1): 22-25.

[6] Liu Qi, Ding Yunlong. Analysis of Technical Causes of Industrial Clusters [J]. China Science and Technology Forum, 2005(4): 122-125.

[7] Gao Ruxi, Ji Yuntao, Chen Zhihong. Systematic Analysis of Technology Chain and Industrial Selection [J]. R&D Management, 2006(6): 95-101.

[8] Huang Liye, Zhao Hui, Wang Jian, et al. Research on Industrial Competitive Intelligence Analysis Framework Based on Patent Analysis [J]. Information Science, 2015, 33(4): 59-63.

[9] Zhu Ruibo. Core Technology Chain, Core Industrial Chain and Regional Industrial Leapfrog Upgrading Path [J]. Economic Management, 2011, 33(4): 43-53.

[10] Fang Si, Li Guoqiu. Global Patent Analysis of Driverless Vehicles: From a Two-Dimensional Perspective of Industrial Chain and Technology Chain [J]. Competitive Intelligence, 2016, 12(5): 27-36.

[11] Wu Feifei, Luan Jingjing, Huang Lucheng, et al. Analysis and Countermeasure Research on China's Power Battery Industry from the Perspective of Industrial Technology Chain [J]. Science & Technology Progress and Policy, 2017, 34(2): 58-63.

[12] Yuan Xiaodong, Bao Yewen. Enlightenment of "ZTE Incident" for China's Industrial Development: Based on Patent Analysis [J]. Journal of Intelligence, 2019, 38(1): 23-29.

[13] Wei YM, Kang JN, Yu BY, et al. A Dynamic Forward-Citation Full Path Model for Technology Monitoring: An Empirical Study from Shale Gas Industry [J]. Applied Energy, 2017, 205: 769-780.

[14] Zhao Rongying, Li Xinlai, Li Danyang. Research on Core Patents from the Perspective of Patent Citation: A Case Study of Artificial Intelligence [J]. Information Studies: Theory & Application, 2019, 42(3): 78-84.

[15] Qi Yun, Tang Heng, Shi Junguo. Research on Core Technology Identification Based on Small-World Network Characteristics: A Case Study of Graphene [J]. Journal of Intelligence, 2020, 39(2): 50-55.

[16] Choi S, Park H, Kang D, et al. An SAO-Based Text Mining Approach to Building a Technology Tree for Technology Planning [J]. Expert Systems with Applications, 2012, 39(13): 11443-11455.

[17] Zhai Dongsheng, Xia Jun, Zhang Jie, et al. Research on Technology Tree Construction Method Based on Patent Feature Extraction [J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(7): 717-724.

[18] Guo Junfang, Wang Xuefeng, Li Qianrui, et al. A Novel Technology Morphology Identification Method Based on SAO Semantic Mining [J]. Studies in Science of Science, 2016, 34(1): 13-20.

[19] Yang C, Huang C, Su J. An Improved SAO Network-Based Method for Technology Trend Analysis: A Case Study of Graphene [J]. Journal of Informetrics, 2018, 12(1): 271-286.

[20] Lai Chao'an, Hou Yanxing, Wen Xionghui. Technology Foresight Method Based on Patent SAO Chains: A Case Study of Industrial Robot Field [J]. Science and Technology Management Research, 2020, 40(7): 171-179.

[21] Li Xiaoman, Zhang Xuefu, Song Hongyan, et al. Research on Technical Element Identification Method in Patent Literature: A Case Study of Nano-Fertilizer Field [J]. Library and Information Service, 2020, 64(6): 59-68.

[22] Wei Ying, Li Feng. Research on Key Node Identification and Rumor Diffusion Suppression Based on Main Path Analysis [J]. Information Science, 2018, 36(6): 13-19.

[23] Liu JS, Lu LYY. An Integrated Approach for Main Path Analysis: Development of the Hirsch Index as an Example [J]. Journal of the American Society for Information Science and Technology, 2012, 63(3): 528-542.

[24] Wang Lina, Tang Chuan, Tian Qianfei, et al. Analysis of Global Quantum Computing Development Trends [J]. World Sci-Tech R&D, 2019, 41(6): 569-584.

[25] Zhang Haiyi, Cui Xiao, Wu Bingbing. Development Status and Application Analysis of Quantum Computing Technology Industry [J]. Information and Communications Technology and Policy, 2020(7): 20-26.

[26] Zheng Jianguo, Qin Chaoyong. Progress and Prospects of Quantum Computing [J]. Application Research of Computers, 2002(3): 641-645.

[27] Zhang J, Tian Q, Tang C, et al. Study on Worldwide Development and Trends of Quantum Technologies Based on Patent Data [J]. International Journal of Information and Education Technology, 2020, 10(3): 239-244.

[28] Zhong Cheng, Chen Guoliang. Quantum Computing and Its Applications [J]. Journal of Guangxi University (Natural Science Edition), 2002(1): 83-86.

[29] Chen Pingxing, Wu Wei, Wu Chunwang, et al. Research Status and Development Trends of Quantum Computing [J]. National Defense Science & Technology, 2014, 35(6): 3-6.

[30] Tian Qianfei, Wang Lina, Tang Chuan, et al. International Development Trend Analysis of Quantum Computing Research Based on Bibliometrics [J]. Science Focus, 2019, 14(6): 1-9.

[31] Liu Xiaoping, Li Zexia. Frontier Hotspot Analysis of Quantum Informatics Based on Co-word Analysis [J]. Science Focus, 2014, 9(5): 13-22.

[32] Zhong J, Olekhno NA, Ke Y, et al. Photon-Mediated Localization in Two-Level Qubit Arrays [J]. Physical Review Letters, 2020, 124(9): 1-6.

[33] Li Lianning. Through the Future World: Quantum Computers [M]. Beijing: Tsinghua University Press, 2019.

Author Contributions

Ren Haiying: Formulated the research topic, designed the research framework, determined research methods, revised the paper, and finalized the manuscript.

Li Zhen: Researched related methods and theories, collected data, determined relevant methods, and wrote the initial draft.

Note: Figure translations are in progress. See original paper for figures.