

Research Article

Culture under Complex Perspective: A Classification for Traditional Chinese Cultural Elements Based on NLP and Complex Networks

Lin Qi ,¹ Yuwei Wang,² Jindong Chen,^{1,3} Mengjie Liao,¹ and Jian Zhang ,^{1,4}

¹Beijing Information Science and Technology University, Beijing 100192, China

²Department of East Asian Studies, University of Arizona, Tucson, AZ 85719, USA

³Beijing International Science and Technology Cooperation Base of Intelligent Decision and Big Data Application, Beijing 100192, China

⁴Beijing World Urban Circular Economy System (Industry) Collaborative Innovation Center, Beijing 100192, China

Correspondence should be addressed to Lin Qi; qilin@bistu.edu.cn and Jian Zhang; zhangjian@bistu.edu.cn

Received 29 December 2020; Revised 16 February 2021; Accepted 17 March 2021; Published 8 April 2021

Academic Editor: Rosa M. Benito

Copyright © 2021 Lin Qi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The cultural element is the minimum unit of a cultural system. The systematic categorizing, organizing, and retrieval of the traditional Chinese cultural elements are essential prerequisites for the realization of effective extracting and rational utilization, as well as the prerequisite for exploiting the contemporary value of the traditional Chinese culture. To build an objective, integrated, and reliable classification method and a system of traditional Chinese cultural elements, this study takes the text of *Taiping Imperial Encyclopedia* in Northern Song Dynasty as the primary data source. The unsupervised word segmentation methods are used to detect Out-of-Vocabulary (OOV), and then the segmentation results by the THULAC tool with and without custom dictionary are compared. The TF-IDF algorithm is applied to extract the keywords of cultural elements and the Ochiai coefficient is introduced to create complex networks of traditional Chinese cultural elements. After analyzing the topological characteristics of the network, the community detection algorithm is used to identify the topics of cultural elements. Finally, a “Means-Ends” two-dimensional orthogonal classification system is established to categorize the topics. The results showed that the degree distribution in the complex network of Chinese traditional cultural elements is a scale-free network with $\gamma = 2.28$. The network shows a structure of community and hierarchy features. The top 12 communities have taken up to 91.77% of the scale of the networks. Those 12 topics of the traditional Chinese cultural elements are circularly distributed in the orthogonal system of cultural elements’ categorization.

1. Introduction

Traditional Chinese culture was formed by the precipitation and accumulation of psychological and behavioral characteristics in its long history [1]. It embodies and reflects the commonalities of the Chinese nation in behavior and temperament. Exploring traditional Chinese culture and discovering its contribution to modern culture is significant for rejuvenating the Chinese culture, promoting the Chinese cultural market, and publicizing the Chinese culture globally.

Cultural elements are the basic units that constitute the cultural system. The use of traditional Chinese cultural

elements has been increasingly valued by scholars who are in the field of cultural creation and management. Incorporating excellent elements of traditional Chinese culture in production, creative design, urban/rural planning, and construction will not only gain wider audiences and greater economic returns but also promote the creative transformation and development of traditional Chinese culture. Furthermore, it is an essential spirit in the “Notice on building a National Cultural Big Data System” by the Publicity Department of the Central Committee of the Chinese Communist Party. Finally, it becomes the consensus of industries and fields such as film and television creation,

animation creation, industrial design, tourism design, and architectural design [2–7].

Under the economic environment of the network platform, classifying, organizing, and retrieving of Chinese cultural elements have been a prerequisite for effective extraction, rational usage, and finally modernizing traditional cultural elements. There are still some deficiencies in this research aspect. First, scholars have not yet made agreements on the definition and classification of traditional Chinese cultural elements. Second, in terms of the classification of traditional Chinese cultural elements, most studies adopt a mixed method of subjective classification or natural language processing and qualitative analysis. There is a need for a reliable way to analyze the data only using the quantitative data analysis method. Also, the current classification is based on cultural element themes, while it lacks a system that illustrates relationships between categorizations, and further reshapes the picture of traditional Chinese cultural structures.

In China's long history, a vast number of books and classics have been accumulated and preserved in the written form to these days. They are the most direct vehicles for analyzing traditional Chinese culture. The systematic classification work of these classics has made some progress [8, 9]. With the increasing development of natural language processing and other technologies, it becomes easier to automatically segment ancient texts, classify questions, mine topics, extract events, and map knowledge [10–14]. The analysis of complex networks, especially community structure, can reveal and interpret the theme and community division of complex structures based on textual data [15, 16]. It provides potential technical support for the classification of traditional Chinese cultural elements.

This study proposes that the Chinese traditional cultural system is a set of material and nonmaterial element systems formed in the long-term survival practice. Traditional Chinese cultural elements are the basic unit of material and nonmaterial carriers that consist of the above system. This research proposes a theoretical method that directly extracts a complete classification of cultural elements from traditional Chinese cultural classics textual data sources using natural language processing techniques, complex network model with its community detection method. The whole process leads to an objective, complete, and valid classification system of traditional Chinese cultural elements.

Based on the literature review of previous studies, this paper is organized as follows: Section 2 gives a brief review of the current research on the classification of traditional Chinese cultural elements and the application of complex networks in natural language processing; Section 3 describes the data course; Section 4 proposes the construction of a complex network of traditional Chinese cultural elements; Section 5 analyzes its network topology; Section 6 focuses on the theme detection and complete classification of this network; and Section 7 concludes with contributions of this study and scope for future research.

2. Literature Review

2.1. Classification of Traditional Chinese Cultural Elements. Although research regarding cultural elements' extraction has made much progress, there is still a paucity of work, especially in establishing an objective, complete, and reliable classification method and system for traditional Chinese cultural elements. For example, Zhou et al. propose that cultural elements are condensations of cultural characteristics [4]. Tai argues that traditional Chinese cultural elements include the characters, symbols, or customs that most Chinese people identify with, embody the Chinese nation's traditional cultural spirit, and show national dignity and national interests. Tai divided these elements into two categories: tangible symbols and intangible spiritual symbols. In terms of application, the two categories are further divided into four parts: text, symbol, color, and spirit [17]. Using cultural elements as a tool to deconstruct culture, Wang et al. point out that cultural elements cover clothing, food, housing, transport, and other people's basic needs related to human survival. They argue that these elements should be classified into elements of tangible culture, institutional culture, and psychological culture [18]. Based on this classification, they adopted a subjective and objective mixed method of word frequency analysis, sentiment analysis, and grounded theory to investigate the classification of cultural elements in tourism. They identified six types of elements: natural landscape images, architectural style patterns, community living atmosphere, remnants of work scenes, characters in rural stories, expression of inheritance skills, and local festival performances [19].

2.2. Application of Complex Network in Natural Language Processing. Corrêa et al. use a bipartite graph to represent the semantic structure of the text. They construct edges of the bipartite network between target words and feature words of target word contexts, and ignore the relationship between the feature words. The network model constructed in this way is applied to achieve the word sense disambiguation task, which shows good results and robustness in the case of small samples [20]. Tohalino and Amancio proposed a multi-document summarization method based on a multilayer network. In this method, sentences are represented by nodes and connected according to shared words. Sentences in the same document are connected at the same layer, while sentences in different documents are connected at different layers. This method effectively improves the quality of summary generation [21]. Marinho et al. analyzed the author's attribution of documents based on the motif structure in complex networks. They used directed motifs with three nodes as features for document classification and compared the result with four machine learning methods. The results show that the method based on motif significantly improves the accuracy of document classification [22]. Qi et al. used natural language processing technology to analyze enterprise documents, established a complex network of reserved knowledge topics, and used a fuzzy comprehensive evaluation method to calculate the

ability of enterprises to meet specific knowledge needs [23]. On the whole, Cong and Liu systematically reviewed the results of language research using the complex network model, sorted out three main research clues, and finally put forward suggestions for future linguistic research from the perspective of the complex network [24].

3. Data Acquisition

The dataset includes two parts. (1) All text of Taiping Imperial Encyclopedia and (2) The China Biographical Database, supporting word segmentation in the process of language processing of Taiping Imperial Encyclopedia.

3.1. Textual Data of Taiping Imperial Encyclopedia. The book—*Taiping Imperial Encyclopedia*—was compiled by officers Fang Li, Mu Li, and Xuan Xu during the Taiping Xingguo era in the Northern Song dynasty. As the most massive Chinese leishu encyclopedia, the compilation work was completed by the 8th year of the Taiping Xingguo era. The encyclopedia was divided into 55 sections and 550 subsections. It thoroughly describes the historical and literature materials before the Song Dynasty.

Electing the *Taiping Imperial Encyclopedia* as the data source lies in the following considerations. (1) Classified documentation. As a *leishu*, *Taiping Imperial Encyclopedia* is similar to current reference books and encyclopedias in many ways. On the one hand, its documentation and explanation of previous literature and knowledge have formed complete entries. On the other hand, its classification of these entries (heaven, Earth, human affairs, and other categories) provides a natural node connection for the complex network construction of traditional Chinese cultural elements. (2) Standardized annotation system. The current digital source of the *Temporal Imperial Encyclopedia* has a complete punctuation system and a patterned hierarchical documentation structure. For instance, it divides the volume number and volume name by “◎,” uses the marker “〇” before each entry, and uses brackets for each reference in the text. These markers help with the rule-based word extractions of Out-of-Vocabularies (OOVs). (3) Appropriate data size. The text of the *Temporal Imperial Encyclopedia* was obtained through the National Key Technologies R&D Program. It contains 4,799,200 Chinese characters, an appropriate size for conducting natural language processing research.

3.2. China Biographical Database. The *China Biographical Database* (CBDB) was first started by Robert Hartwell, a famous sinologist and professor at the University of Pennsylvania. The database was first based on his research and information collection. It then became a joint project, continuously maintained and developed by Fairbank Center for Chinese Studies at Harvard University, Institute of History and Philology of Academia Sinica, and Center for Research on Ancient Chinese History at Peking University. The CBDB has incorporated sources from biographical indexes for 7–19 century A.D, including dynasties, reign

title, person names, place names, official names, literature work, and social relations. The wide range of biological sources and metadata in CBDB will serve as an important supplemental custom dictionary for our further word segmentation. The CBDB used in this study was obtained through public Internet websites.

4. Complex Network of Chinese Traditional Cultural Elements’ Construction

Four steps in this construction process are OOV detection, word segmentation, themes of cultural elements extraction, and association. Mutual information and adjacency entropy are employed for rule-based OOV detection. Then, the THULAC is used to achieve word segmentation of text chapters. The TF-IDF algorithm is applied to extract the keywords of cultural elements’ topics. Finally, the association of keywords was obtained from the Ochiai co-occurrences coefficient calculation. The construction process of the complex network of traditional cultural elements is shown in Figure 1.

4.1. OOV Detection. Since there is no space between Chinese characters, it needs to use word segmentation tools to segment words before keyword extraction. No matter how large a dictionary is used to train word segmentation tools, a certain proportion of words in actual word segmentation are outside the dictionary, which are called Out-Of-Vocabulary (OOV) [25]. Before using the pretrained word segmentation tool for word segmentation, detecting OOVs and adding them to the user-defined dictionary can help to improve the accuracy of word segmentation.

4.1.1. Mutual Information and Adjacency-Entropy-based OOV Detection. Each word is an independent linguistic unit. For words consisting of more than two characters, there are associations between characters. A stronger association infers a higher possibility that these characters will form a word. Mutual information can be used to evaluate and quantify this association between characters [26]. Due to the large size of the candidate strings, we implement mutual information for multi-word extraction, which is defined as

$$MI(W) = \log_2 \frac{p(w_1 \dots w_n)}{\text{avg}(w_1 \dots w_n)},$$

$$p(w_1 \dots w_n) = \frac{f(w_1 \dots w_n)}{\text{num}},$$

$$\text{avg}(w_1 \dots w_n) = \frac{1}{n-1} \sum_{i=1}^{n-1} p(w_1 \dots w_i)p(w_i \dots w_n),$$
(1)

where $W = w_1 \dots w_n$ is the candidate string, $p(w_1 \dots w_n)$ represents the probability of the string $w_1 \dots w_n$ occurring in the corpus, $f(w_1 \dots w_n)$ denotes the occurrence number of the string $w_1 \dots w_n$ in the corpus, num is the total number of Chinese characters in the corpus, and $\text{avg}(w_1 \dots w_n)$ is the

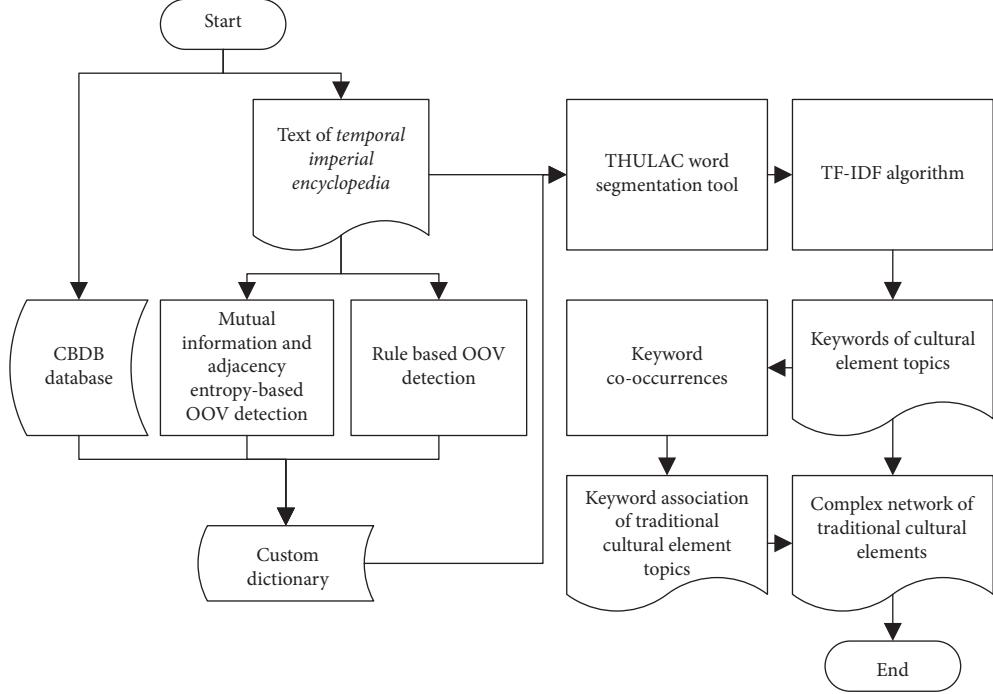


FIGURE 1: The construction process of the complex network of traditional cultural elements.

probability of different combinations of multicharacter strings.

Besides using the mutual information to measure the internal cohesiveness, the adjacency entropy is also applied to facilitate the match of word boundaries. Entropy is the measurement of uncertainty, while information entropy is used to quantify the uncertainty of information. In general, a higher value of adjacency entropy implies that neighbor words of a character or a string are more diverse, and therefore more likely to be the word boundary. Adjacency entropy can be directionally categorized into left entropy and right entropy. For instance, the calculation method of left adjacency entropy is

$$H_{\text{left}}(W) = - \sum_{w_{\text{left}} \in C_{\text{left}}} p(w_{\text{left}}|w) \log_2 p(w_{\text{left}}|w), \quad (2)$$

where C_{left} is the set of left contiguous words of the candidate word w , $p(w_{\text{left}}|w)$ is the conditional probability that the adjacency on the left is w_{left} when the candidate word w appears. If $N(w_{\text{left}})$ is the frequency at which the left contiguous word w_{left} and the candidate word w appear together, $N(w)$ is the frequency at which the candidate word w appears alone; then, $p(w_{\text{left}}|w)$ can be expressed as

$$p(w_{\text{left}}|w) = \frac{N(w_{\text{left}})}{N(w)}. \quad (3)$$

Consequently, by comparing the $H_{\text{left}}(W)$ or $H_{\text{right}}(W)$ to a given threshold, we can determine the direction from which the candidate word will be chosen. Defining

$$S(W) = MI(W) + H_{\text{left}}(W) + H_{\text{right}}(W), \quad (4)$$

and TopN words will be considered as OOVs after sorting the $S(W)$. Based on the aforementioned methods, the whole

text of *Temporal Imperial Encyclopedia* is processed and 1245 OOV words are retrieved. A randomly selected part of OOVs is shown in Table 1.

4.1.2. Rule-Based OOV Detection. As an unsupervised learning method, the accuracy and recall rate of the OOV detection based on mutual information and adjacency entropy still need to be improved. Therefore, a rule-based OOV detection method, as a supplement of the unsupervised method, is introduced to process texts of the *Temporal Imperial Encyclopedia*. As shown in Figure 2, there exists a hierarchical structure guided by symbols in the texts. Thus, we proposed corresponding rules for OOVs' detection: (1) volume names: after “◎” until the nearest Chinese character string, such as *TianBu* “section of the heaven” and *DiBu* “section of the Earth”; (2) subsection names: the character string between “○” and the nearest new line break, such as “*Yuanqi*” and “*Taiyi*”; and (3) reference names: the string between the left book title number “《” and the nearest right book title number ”》”, such as *Laozi* and *Hetu*.

In this process, OOVs are detected via a hybrid method of mutual information, adjacency entropy, and rules. Moreover, words from the CBDB database, such as dynasties, era names, person names, place names, official names, and literary works are also selected. Taken together, OOVs serve as the customized dictionary for word segmentation, which consists of 631,522 words.

4.2. Word Segmentation Validation. Three chapters, “Volume 213-Officials Section 11,” “Volume 782-Barbarian Tribes Section 3-East Barbarian Tribes Subsection 3,” and “Volume 43-Earth Section 8” from the *Temporal Imperial*

TABLE 1: A randomly selected part of OOVs in *Temporal Imperial Encyclopedia* extracted through mutual information and adjacency entropy.

Num	Category	OOVs
1	Names, ceremony, family names	Shangqing, Yishao, Fuxi, Diku, Erzhurong, Hou'andu, Taishang, Zaofu, Huanwen, Yuanjun, Anlingchan, Cuihong, Shaokang, Caogong, Erzhu, Siming, Wuding, Chiyou, Dingji, Hebo, Zhumeng, Jieli, Shikuang, Baosi
2	Ceremony, system	Dixia, Qingxu, Baowei, Qiansui, Jiaosheng, Jiaosi, Zhanshuai, Xiangxing, Yuelv, Ziyi, Shizhen, Liudian, Mingjing, Tuntian, Fengchan, Yudie, Yuanqiu, Jieyu, Lushan
3	Direction, geographical names, astronomy	Sufang, Jinyang, Lelang, Fuyu, Qianmu, Yanzhou, Gaochang, Fengxiang, Wusun, Jinbian, Fusang, Wangshi, Tuguhun, Maotou, Huoshan, Chenliu
4	Plants, animals, medicinal materials	Juyou, Zhongru, Qiongqiong, Zishiying, Longxu, Zhuyu, Junyi, Luanniao, Qilin, MaifanQingling, Xuanniao, Yushi, Pichao, Qicao
6	Official, government office	Shilang, Zongzheng, Taifu, Bieji, Yuren, Duzhi, Zuocheng, Liangzhoumu, Sikong, Duwu, Taiguan, Dajiangjun, Taishi, Huangmen, Shaofu
7	Utensils	Tuohu, Tanqi, Chupu, Shuijing, Gongshi, Guizan, Hunyi
8	Books, work of art	Yifu, Wenshi, Yuanri, Ruiyington, Tianyujing, Shangzi, Qingying, Hetu
9	Decade of a century, solar term, season	Houwei, Xiaojing, Jidong, Baique, Yiyue, Liuyue, Wugeng, Mingjia
10	Behavior	Baochou, Guqin, Xingjian, Chanyunyun, Shangshu, Gujiu

卷一◎天部一

Volume 1 ◎ Heaven Section 1

- 元气

- Vitality

《三五历纪》曰：未有天地之时，混沌状如鸡子，溟涬始牙，濛鸿滋萌，岁在摄提，元气肇始。又曰：清轻者上为天，浊重者下为地，冲和气者为人。故天地含精，万物化生。

In *Sanwuliji*, it says “Before the world, the world was chaotically like an egg. With the passage of time, the natural air, the body of heaven and earth slowly began to form. The vitality also slowly began to burst,” it also says “When the universe was first formed, some small and small things floated upward to form the sky, and the heavy matter sank and formed the earth.”

.....

- 太易

- Taiyi

《老子》曰：有物混成，先天地生。

In *Laozi*, it says “There is a chaotic and complicated thing that appeared earlier than heaven and earth.”

《易乾凿度》曰：夫有形者，生于无形；故有太易者，未见气也。

In *Yiqianzaodu*, it says “All tangible things are produced by the invisible things. Therefore, in the state of the boundless emptiness of the universe, there is no Qi.”

《帝王世纪》曰：天地未分，谓之太易。

In *Diwangshiji*, it says “When the heaven and the earth are not separated, the state of being incapable of chaos is called Taiyi.”

.....

FIGURE 2: The hierarchical textual structure in *Temporal Imperial Encyclopedia*.

Encyclopedia, are randomly selected for word segmentation validation.

Artificial word segmentation and word annotation serve as the standard. Then, automatic segmentation and annotations by applying the THULAC with and without customized dictionaries are compared and validated. The THULAC (THU Lexical Analyzer for Chinese) is a Chinese lexical analysis toolkit developed by the Natural Language Processing and Computational Social Science Lab in Tsinghua University [27]. The word segmentation model that comes with the THULAC Toolkit is trained by the

People's Daily corpus, which contains about 12 million words.

It is noticeable that segmentation with customized dictionary outstrips that without across all three chapters. All three metrics (Precision, Recall, and F-measure) remarkably improve in each section: in the Earth Section (from 0.55, 0.49, 0.50 to 0.78, 0.72, 0.73, respectively); in the Official Section (from 0.53, 0.50, and 0.50 to 0.72, 0.70, 0.70, respectively); in the East Barbarian Tribes Subsection (from 0.40, 0.29, 0.33 to 0.75, 0.65, 0.67, respectively). Considering the three indicators and the tolerance of large-scale text data

to noisy data, the THULAC tool with a custom dictionary can meet the research needs. The word-segmentation results of sample texts with the THULAC NLP tool are shown in Table 2. The segmentation of typical data sources including sections of philosophy writings (Zi), heaven, Earth, criminal law, and human affairs with the THULAC tool are shown in Table 3.

4.3. Keywords Extraction of Traditional Chinese Cultural Element Topics. Different keyword extraction algorithms can meet the needs of different scenarios. For example, the method based on intermittency is especially suitable for short text and single document text [28], while TF-IDF and TextRank are more suitable for extracting information in a long text environment with multiple documents [29, 30]. Because of the simplicity of the multi-document scene and TF-IDF algorithm, the TF-IDF algorithm is selected to extract keywords of *Temporal Imperial Encyclopedia*.

The TF-IDF algorithm considers both the word frequency and the reverse document frequency: from the perspective of word frequency, the higher the frequency of a word appearing in a single document, the more prominent the topic is represented by the word; from the perspective of reverse document frequency, it is considered that a word appears in all documents. The frequency of occurrence is high, the general importance of the word is high, and the topic represented is less significant [29, 30]. TF-IDF assumes that the theme represented by a word is notable if this word has high word frequency in a single chapter and low inverse document frequency in all texts. Word frequency tf_{ij} is defined as

$$tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}}, \quad (5)$$

where n_{ij} is the frequency of word v_i shown in text d_j ; n_{kj} is the occurrences of all words in the text d_j . The inverse document frequency $i df_i$ can be expressed as

$$i df_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|}, \quad (6)$$

where $|D|$ is the total number of texts in *Temporal Imperial Encyclopedia*, $|\{j: t_i \in d_j\}|$ refers to the number of texts including v_i . Considering both word frequency tf_{ij} and inverse document frequency $i df_i$, the prominence of a cultural element theme word is

$$tfi df_i = tf_{ij} * i df_i. \quad (7)$$

Let α be the prominence threshold, then the collection of cultural-elements themes is $V'' = \{v_i | v_i \in V \cap tfi * df_i > \alpha\}$. Furthermore, we can describe the weight coefficient of the theme v_i in V'' as

$$tfidf'_i = \frac{tfi df_i}{\sum_{j=1}^n t f i d f_j}. \quad (8)$$

At last, the weight coefficient matrix of the theme collection of cultural elements V'' can be written as

TABLE 2: Word-segmentation results of sample texts with the THULAC NLP tool.

Metrics	No custom dictionary			Custom dictionary		
	Earth	Dongyi	Officials	Earth	Dongyi	Officials
Precision	0.55	0.40	0.53	0.78	0.75	0.72
Recall	0.49	0.29	0.50	0.72	0.65	0.70
F-measure	0.50	0.33	0.50	0.73	0.67	0.70

$$A = (tfidf'_1, t f i d f'_2, \dots, t f i d f'_n), \quad (9)$$

where n denotes the number of elements in V'' . In this study, the derived n is 22496 with $|D| = 1000$ and $\alpha = 0.05$.

4.4. Keywords Association of Traditional Chinese Cultural Element Topics. Keywords extraction of topics provides candidate nodes for the complex network of Chinese traditional cultural elements. This section will focus on the association between extracted keywords to establish complex relationships between nodes and reflect the relevance of the topic of traditional cultural elements. The network of traditional Chinese culture element $G = (V, E)$ is an undirected and unweighted network, where $V = \{v_i | i = 1, 2, \dots, n\}$ represents the nodes or vertices, and $E = \{e_{ij} | i, j = 1, 2, \dots, n\}$ are the edges. The edges imply the co-occurrence relationship between connected themes. The co-occurrence correlation is obtained based on the analysis of co-occurrent frequencies and quantified through Ochiai, which can be express as

$$O_{ij} = \frac{tf_{ij}}{\sqrt{tf_i \times tf_j}}, \quad (10)$$

where tf_{ij} denotes how many times that the cultural-element topic words v_i and v_j occur in one text, tf_i is the total number of occurrences of v_i , tf_j represents the total frequency of the topic-word v_j , and O_{ij} represents the Ochiai coefficient between topic words v_i and v_j . A stronger association between themes v_i and v_j will lead to a higher O_{ij} [31, 32]. Let β denote the threshold of connection strength, and equation (11) is implemented to realize the mapping between connection strength and weight of edge in the unweighted network.

$$e_{ij} = \begin{cases} 1 & O_{ij} > \beta \\ 0 & O_{ij} \leq \beta \end{cases} \quad (11)$$

In the present study, it is set that $\beta = 0.02$ and 22492 nodes are extracted. After excluding the nodes with $k = 0$, a network with 10423 nodes and 68923 undirected edges is generated.

5. Complex Network Topological Analysis of Traditional Chinese Cultural Elements

5.1. Statistical Characteristics of the Topological Network. Several main statistical characteristics, including average degree $\langle k \rangle$, average path length $\langle d \rangle$, diameter of the network

TABLE 3: Word-segmentation results of typical data source with the THULAC tool.

Number	Section	Word-segmentation examples
1	Philosophy writing (Zi)	Lizhètiyērēngqyōuāiyūwuxingŷouūxīngmiegūlxīangyinzhīzhīhaopinzhīhīaožunbeishangxiāyouti “Propriety is the decency. Human emotions have sorrow and joy; the five elements have rise and fall. Therefore, it sets up feast customs, funeral customs, marriage customs, and an imperial appointment system. Everything is in hierarchical order, from top to bottom”
2	Heaven	dunjiakaishantuyuehuoshanānanyeyouyunshiyuhu In <i>Hetuditongji</i> , it says the rain nourishes everything in the world. In Dunjiaishaantu, it documents there exist silk words named Yunshi and Yuhu.
3	Earth	huayangguozhīyuejianweīn’ānxiannananyōuēmeishānqūxiānbashildituyunyōuxianyaohanwūqiubūnengde In <i>Huayangguozhi</i> , there is a Mount Emei in the southern area, 40 kilometers from the county. The map says there exists an elixir. Emperor Wu of Han pursued it while not was able to get it.
4	Criminal law	duiyuechēnhāoyēsōngshīyūnūputianzhixiāmofeiwangtu Shuaituzhibinmofeiwangshenjinhunjuntianzhenzēwotianzhenyé Answered “I have been familiar with the “Book of Songs” since childhood, and there is a poem in the book saying: “the whole world, is it the king’s soil; all people on the Earth, are the king’s ministers”. Now that the king of Zhou is the ruler of the world, then I am a minister of the emperor”
5	Officials	ShijīyueFansuīshuoQinzhaowangyueWuzixūTuozaierchūzhāoqngyēxīngérzhāofūzhīyulingshūiwyihqikoutixīnggōpfūBāf jishouroutangtūchuihuaoxiqiyiwushizuwoguheluebihoshichen dējinmouruizixujiazhiyyouquzhongshenbufujianshichenzihshuoxingye The <i>Record of the Grand Historian</i> says “Wu Zixu persuaded Qin Shi Huang and said “Wu Zixu hid in a bag and got out of Zhaoguan. He drove at night and was concealed during the day. When he arrived at Linghui, he had nothing to eat. He sat and walked, crawled, and begged in Wu City. Finally, he rejuvenated the country of Wu, and Wu Wang Lu became the overlord. If the courtier’s strategy of advancing and offering can be like Wu Zixu’s, it is to confine me and never see the king for life, as long as the courtier’s proposal is implemented.””

D , average clustering coefficient $\langle C \rangle$, are calculated to evaluate the network. Let k_i be the degree of the vertex v_i , which refers to the number of neighbor vertices connected to the vertex v_i . The average degree of the network can be expressed as

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^n k_i, \quad (12)$$

where $\langle k \rangle$ denotes the degree of association between vertices in the network. A larger $\langle k \rangle$ indicates closer and more diverse connections between traditional Chinese cultural elements.

The average path length is also known as characteristic path length, which is defined as the average number of edges in the shortest paths between all vertex pairs, given by

$$\langle d \rangle = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}, \quad (13)$$

where d_{ij} is the number of edges in the shortest path between vertex v_i and vertex v_j . The diameter of the network D is defined as the largest distance between any two vertices in the network

$$D = \max(d_{ij}). \quad (14)$$

In this study, the average path length $\langle d \rangle$ and diameter of the network D are employed as the diversity metrics for elements comparison. Larger dissimilarities between traditional Chinese cultural elements lead to the higher average path length and diameter of the network.

The clustering coefficient of a vertex in the network quantifies how close its neighbors are. In this study, the clustering coefficient of a Chinese traditional cultural element represents the extent to which its neighbor cultural elements tend to cluster together. A large clustering coefficient resulting from closely related cultural elements indicates a specific topic has emerged. A clustering coefficient of a vertex can be expressed as

$$C_i = \frac{2l_i}{k_i(k_i - 1)}, \quad (15)$$

where l_i is the edge number of the vertex v_i . Furthermore, the clustering coefficient of the network is defined as

$$\langle C \rangle = \frac{1}{n} \sum_{i=1}^n C_i. \quad (16)$$

The main statistical characteristics of the complex network of traditional Chinese cultural elements are summarized in Table 4.

5.2. Degree Distribution of the Network. Degree distribution $P(k)$ can be interpreted as the probability that a vertex is randomly chosen from a network that has a degree of k , or the fraction of vertices in a network that has degree k . Different degree distributions signify the distinctions in both the structure and the function of the two networks, even

TABLE 4: The main statistical characteristics of the complex network of traditional Chinese cultural elements.

Statistical characteristics	Vertex	Edge	$\langle k \rangle$	$\langle d \rangle$	D	$\langle C \rangle$
Value	10423	68923	13.23	3.41	9	0.69

though they have the same vertex number and average degree $\langle k \rangle$.

Figure 3 shows the log-log scatter plot of degree distribution $P(k)$ versus degree k in the complex network of traditional Chinese cultural elements. These scatters exhibit a linear behavior in the log-log plot. Most vertices have a very low degree, whereas only a few vertices have extremely large degrees. $P(k)$ rapidly declines as degree k increases. Thus, the complex network of traditional Chinese cultural elements can be claimed to be a scale-free network [33]. According to the maximum likelihood fitting method by Clauset et al. [34], the fraction of nodes with degree k follows a power law, $P(k) \sim k^{-\gamma}$, where $\gamma = 2.28$ and the corresponding Kolmogorov-Smirnov goodness-of-fit (GoF) is 0.0288.

Scale-free networks have broad implications for the structure and dynamics of complex systems, one of which is the heterogeneity of vertex. That is, minority high-degree core nodes are in the dominant position while low-degree vertices are located to the periphery of the network. Applying this perspective, the influence of Chinese traditional cultural elements is heterogeneous, and most of the elements only appear in this cultural system. However, some elements appear frequently together with the above elements, and it seems that their influence is always reflected everywhere. These elements constitute the core elements of Chinese traditional culture. The top 30 cultural elements with the highest degree and dominant effect in the Chinese traditional complex network are listed in Table 5. Almost all elements fall into the political life category. However, the effects that the original subject matter of *Temporal Imperial Encyclopedia* has on node degree cannot be neglected. Therefore, these elements in Table 5 do not necessarily imply the core elements of Chinese traditional culture but merely belong to a specific cultural theme. This aspect will be elaborated in the module feature analysis.

5.3. Clustering Coefficient-Degree Correlation Analysis of the Network. Clustering coefficient-degree correlation of the network indicates the relationship between degree and clustering coefficient of the node. Generally, provided that in scale-free networks, the clustering coefficient of a node with degree k follows the power law, $C(k) \sim k^{-\alpha}$, a hierarchy of nodes with different degrees exists in the networks [35]. In such networks, small clusters, which are formed by densely interconnected low-degree nodes, are combined with high-degree nodes to form larger but less interconnected groups. These less interconnected groups combine again to form even larger and even less interconnected clusters. This self-similarity nesting of different groups leads to a hierarchical structure on the network. Previous investigation has shown that this hierarchical structure is also present in real

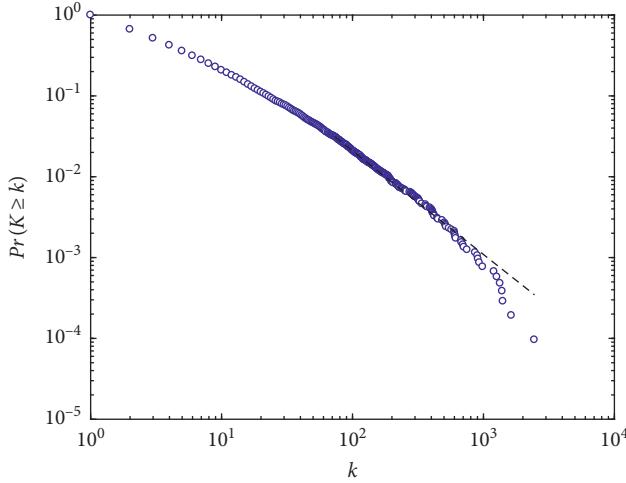


FIGURE 3: Log-log scatter plot of degree distribution for the complex network of Traditional Chinese cultural elements.

TABLE 5: Top 30 cultural elements with highest degree.

No.	Element	k
1	Tianxia (Under heaven)	2453
2	Taizi (Crown prince)	1638
3	Jiangjun (General)	1412
4	Tianzi (Son of Heaven)	1394
5	Gaozu (Remote ancestor)	1343
6	Zhuhou (Feudal vassal)	1270
7	Taishou (Chief of a prefecture)	1206
8	Taizu (Great Ancestor)	989
9	Cishi (Provincial governor)	933
10	Huanghou (Queen)	913
11	Dafu (Senior official)	902
12	Yuannian (First year)	867
13	Dashe (Amnesty)	752
14	Huangdi (Emperor)	711
15	Junzi (Nobleman)	700
16	Kongzi (Confucius)	682
17	Furen (Lady of high rank)	679
18	Shangshu (High official)	617
19	Taizong (Second emperor of a dynasty)	608
20	Bixia (Your Majesty)	603
21	Hanshu (History of Western Han)	602
22	Houhanshu (History of Eastern Han)	572
23	Jinshu (History of Jin Dynasty)	547
24	Chang'an (Xi'an)	520
25	Baixing (Common people)	513
26	Taihou (Empress Dowager)	506
27	Gongzhu (Princess)	486
28	Shidaozhi (Book name)	486
29	Xiongdi (Brother)	452
30	Sima (Minister of War)	446

networks, including Actor networks, Language networks, World Wide Web, Internet at the AS or router level [35].

The scatter distribution of the clustering coefficient of the node $C(k)$ versus k in a dual logarithmic coordinate axis is shown in Figure 4. It is clear that $C(k)$ splits into two branches. The first branch has low k and high $C(k)$,

indicating the densely interlinked clusters. While the other branch, the linear relation between $C(k)$ and k at high k , implies that $C(k)$ scales as k^{-1} , representing the existence of the hierarchical topology in the network. According to the conclusion of literature [35], this discovery reveals the self-similarity characteristics of the complex network hierarchy of Chinese cultural elements. This feature implies that if the preliminary classification of the Chinese cultural elements system can be identified by the method of community detection, then the cultural elements network of each independent category should be further classified. However, in this paper, only the overall network is classified in the following sections.

6. Detection and Classification of Topics of Traditional Chinese Cultural Elements

6.1. Cultural Element Topics Detection. Community detection in complex networks as employed in this study can provide a method to distinguish the topics of traditional Chinese cultural elements from each other. Community and its detection have been widely applied in social networks and other complex network analyses. In a complex network, communities are defined as the sets of nodes where each set of nodes is densely interconnected while sparser connections between the sets [36, 37]. Community structures and hierarchical structures between communities have been analyzed through the network clustering coefficient in the above sections. Detection and naming of these communities' structures are vital to uncover unknown functional modules. Owing to the large network, the fast community unfolding algorithm, proposed by Blondel et al. is utilized to extract the community structure [38]. The modularity serves as the metric for detecting the communities when different resolutions are applied due to unknown community numbers. A higher modularity implies a better detection as denser connections internally but sparser connections between communities. The modularity is defined as

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta_{g_i g_j}, \quad (17)$$

where m represents the number of edges in the network, A_{ij} is the adjacency matrix, and g_i and g_j are the communities to which node i and j are assigned, respectively. In the δ -function, $\delta_{g_i g_j}$ equals to 1 when $g_i = g_j$ and 0 otherwise.

Figure 5 depicts the modularity and community number versus resolution in which resolution ranges from 0.6 to 1.0, with an increment of 0.2. At each chosen resolution, modularity and community numbers are repeatedly calculated 20 times to acquire the corresponding average value and standard deviation. It is insightful to note that the modularity level is escalated as resolution increases and reaches the maximum value of 0.42 when resolution equals 0.94, where the community number reaches the minimum of 38. Thus, the communities are detected with a resolution of 0.94. The results of the detected community, including the size, proportion, and cumulative proportion, are shown in Table 6. As can be seen, the cumulative proportion of the Top

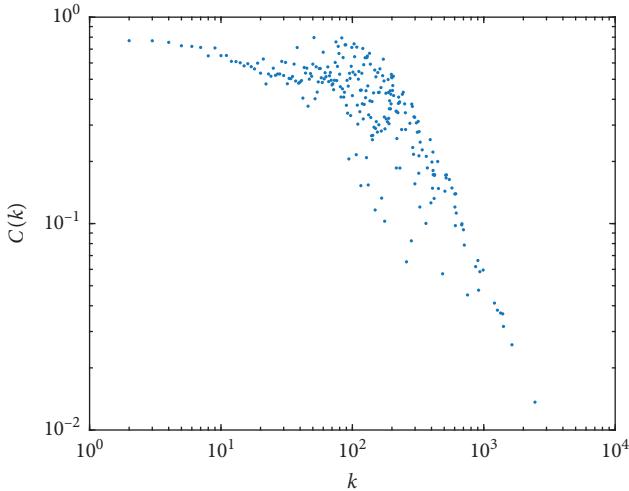


FIGURE 4: The scaling of $C(k)$ with k in traditional Chinese cultural element network.

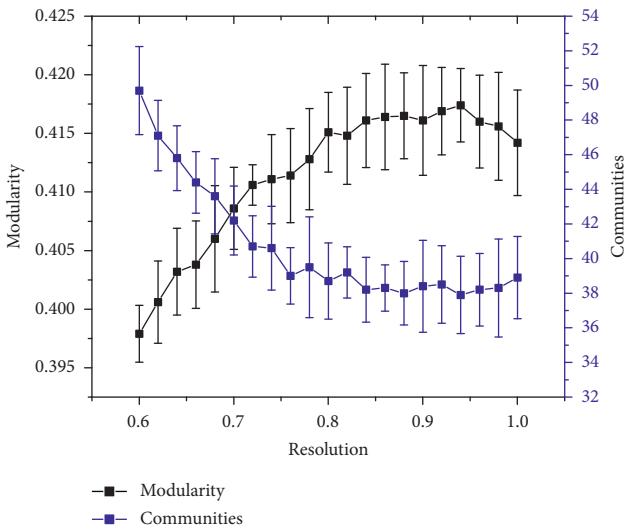


FIGURE 5: Dual-line plot with modularity (on left y-axis) and community number (on right y-axis) versus resolution.

TABLE 6: Size, proportion, and cumulative proportion of communities (topics) in the complex network of traditional Chinese cultural elements.

No.	Size	Proportion (%)	Cumulative proportion (%)
1	1548	14.85	14.85
2	1138	10.92	25.77
3	1111	10.66	36.43
4	1059	10.16	46.59
5	1030	9.88	56.47
6	798	7.66	64.13
7	690	6.62	70.75
8	618	5.93	76.68
9	460	4.41	81.09
10	425	4.08	85.17
11	419	4.02	89.19
12	269	2.58	91.77
Others	858	8.23	100

12 communities is 91.77%. Visualization and analyses of these 12 communities are realized through the Force Atlas algorithm in Gephi. Names and hierarchy structures of the cultural topics, community distribution and names of themes, and typical cultural elements are shown in Table 7, Figures 6 and 7, respectively.

6.2. Cultural Element Topics Classification. The Force Atlas algorithm lays out the nodes according to the extent of interdependence between correlated nodes. Thus, nodes with stronger correlations and the communities formed by these nodes are closer to each other in the performed layout. More specifically, colored communities and cultural topics in close positions imply a strong correlation between them, as in Figure 6. This study constructs a two-dimensional orthogonal cultural element topic classification system—a Means-End classification system to systematically explain this association.

Edgar Schein believes that culture serves the purpose of external adaption and internal integration [39]. The horizontal dimension of the Competing Values Framework (CVF) for cultural assessment by Quinn and Rorbaugh also pointed out that the organization focuses inward or outward [40]. Mao is convinced that practice is based on knowledge, of which objects include the nature phenomenon, nature property, natural regularity, and the relationship between nature and humans as well as that between people [41]. Hence, the culture element topics can be classified either as internal integration or external adaption based on its objective or nature-oriented and social-oriented practice according to its means. By creating an orthogonal coordinate system with means and objectives as its axes, the cultural element topics quadrant and location can offer useful insights for the explanation.

As shown in Figure 6, Confucianism, the core of the Chinese traditional culture value system, is located at the network center. The cultural elements of Confucianism also have a strong correlation with the other 11 thematic elements, which implies a near-perfect adaption of Confucianism in Chinese society's internal and external environments. Being able to solve the internal and external problems, as a result, Confucianism is collectively recommended as the central theme by others.

In contrast to Confucianism, however, Taoism is marginalized and located in the quadrant II. Its position reveals its emphasis on the relationship between humans and nature that spontaneously formed from social life; the so-called “People follow Earth which follows Heaven; Heaven follows Dao which follows the Nature.” Spontaneous formation refers to Taoist concepts and elements based on myths and legends that are sublimated from craft artifacts and survival skills in people's everyday lives. From this perspective, craft artifacts are the materialization of social life, while myths and legends are the sublimation of social life and craft artifacts. Furthermore, Taoism is the theoretical origin of the myths and legends and connects social life to nature. The natural herbs topic, located right above quadrants I and II in Figure 6, exhibits various products from nature. The

TABLE 7: Names and hierarchical structures of cultural topics.

Hierarchy	Name	Typical elements
Values	Confucianism	Figure 7(a)
	Taoism	Figure 7(b)
Institutional behavior	Military and diplomacy	Figure 7(c)
	Social rule	Figure 7(d)
	Political life	Figure 7(e)
	Cosmos and astrology	Figure 7(f)
	Natural disaster	Figure 7(g)
Symbols	Landforms and geography	Figure 7(h)
	Exotic customs	Figure 7(i)
	Craft artifacts	Figure 7(j)
	Myths and legends	Figure 7(k)
	Natural herbs	Figure 7(l)

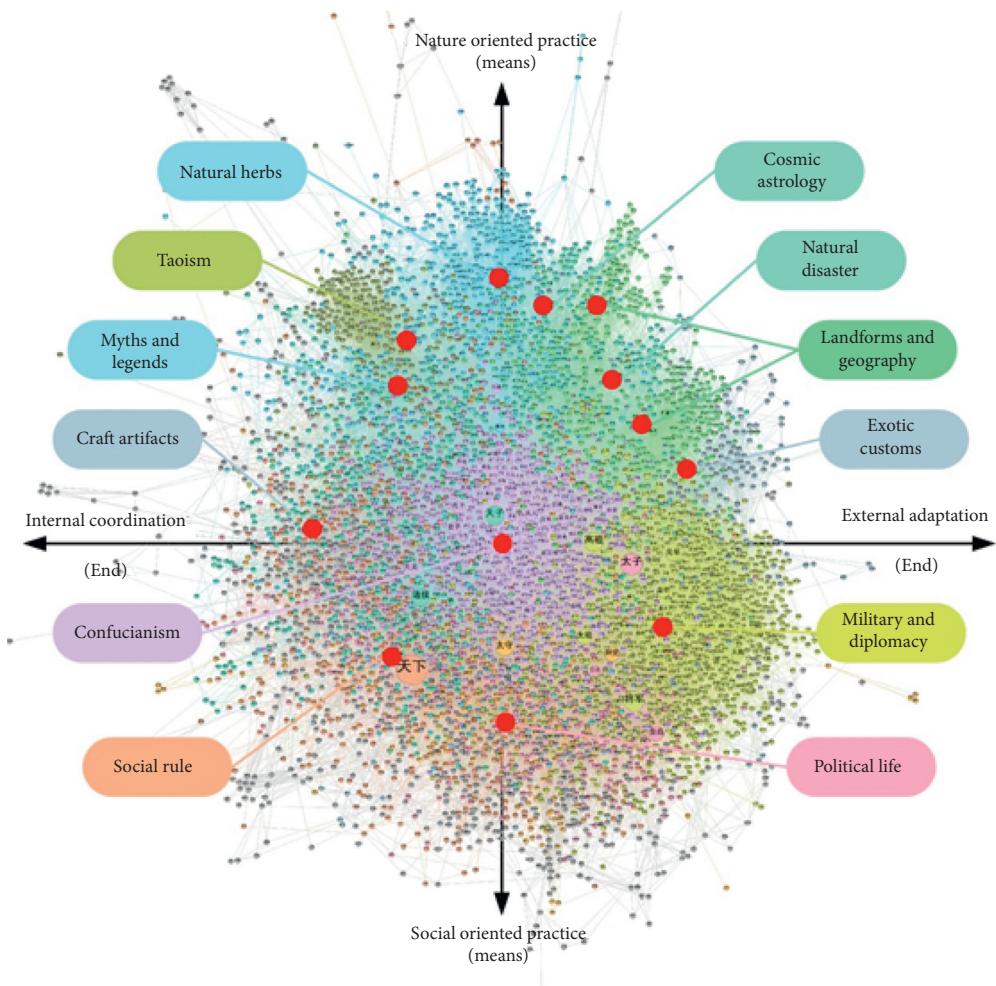


FIGURE 6: Communities and topics in the complex network of traditional Chinese cultural elements.

elements in this topic can be widely used by people; they nevertheless have the least impact on human activities.

On the right side of the natural herb theme, topics in quadrant I involve the natural and exotic environments that affect people's activities. The topics of cosmos and astrology narrate astronomical phenomena related to divination and

disasters; the topic of natural disaster describes various catastrophic natural events, such as wind, rain, thunder, hail, and earthquake, which caused actual social, economic, and property losses. Both topics reflect that people tend to profit and avoid loss when natural environments change at that time. The topic of landforms and geography records



FIGURE 7: Typical elements in traditional Chinese cultural element complex network's topics. (a) Confucianism. (b) Taoism. (c) Military and diplomacy. (d) Social rule. (e) Political life. (f) Cosmic astrology. (g) Natural disaster. (h) Landforms and geography. (i) Exotic customs. (j) Craft artifacts. (k) Myths and legends. (l) Natural herbs.

mountains, rivers, and administrative divisions within the ruling class. One interesting outcome is that natural disaster is surrounded by the elements of the landforms geography topic due to the association between disaster and geographical conditions. Finally, the topic of exotic customs reports the ethnic minorities, external geographical environment, and foreign cultures outside the ruling region. Thus, although Buddhism has been introduced to China

since the Han Dynasty and is inextricably linked with traditional Chinese culture in every aspect, the unshakable dominant role of Confucianism is not affected. Buddhism's influences focus on external environment adaption of people but the penetration is less at the political and military levels.

The elements from military and diplomacy mainly occupy quadrant IV to coordinate relations with foreign ethnic minorities. These elements detail not only external

aggressions, antiaggressions, conquest, negotiation, and indirect rule but also delineate military activities that occurred during dynasty alternation. Because as far as the ruling class is concerned, there is no essential difference between incursions and internal disturbance, as both require decisive intervention.

The third quadrant is primarily involved with the social rule, which aims at internal society coordination and government. This topics' cultural elements include but are not limited to criminal names, laws, rituals, moral concepts, and specific events. On the other hand, the political life topic, settled at the center of quadrants III and IV, consists of cultural elements from the royal court, officials, government offices, important empresses, relatives, and politician domains. It serves to coordinate domestic affairs, military, and diplomacy, and reflects the highest form of social practice.

It is also found that there exists an intertwined relationship between political life and social rule. There are no apparent color divisions like those among other topics, which to a certain extent reflect the hereditary monarchy thought and governance models of the ruling class. On the right side of the abscissa, social rule results in the orderly progress of daily social production activities, and hence the topic returns to craft artifacts which stand for daily life.

Typical cultural elements of each topic are shown in Figures 7(a)–7(l).

7. Conclusion and Limitations

Cultural elements are minimal units of a cultural system. The goal of this study was to build an objective, complete, and reliable categorization system of traditional Chinese cultural elements. Using *Temporal Imperial Encyclopedia* from the Northern Song dynasty as the major data source, this research adopted the unsupervised word segmentation method to detect OOVs, used THULAC NLP tools to achieve word segmentation, and then extracted keywords of cultural elements topics. After that, a complex network of traditional Chinese cultural elements based on the relationships between the keywords is constructed, and the characteristics of network topology are analyzed. The network's community detection algorithm is applied to detect the topics of cultural elements. Based on these, this research built a two-dimensional orthogonal Means-Ends cultural elements classification system to classify major cultural topics. The following conclusions were made:

(1) The research randomly chose three text files and put them to test. An OOV detection method was used, which was based on mutual information and adjacency entropy to process all unregistered words that were retrieved from a custom dictionary. The custom dictionary was used for word segmentation of *Temporal Imperial Encyclopedia*. With the test on the THULAC NLP tool, the Precision, Recall and F-measure all improved, compared with that of no-custom dictionary being used.

The TF-IDF algorithm is used to extract keywords of traditional Chinese cultural element topics, and the Ochiai coefficient is applied to calculate relationships between

keywords. Setting the number of texts as 1000, the importance threshold of topics as $\alpha = 0.05$, and the strength of relations between themes as $\beta = 0.02$, we extracted 22492 keywords of cultural element topics. After removing nodes with $k = 0$, we obtained 10432 nodes from the complex network of traditional Chinese cultural elements and 68923 undirected edges. The average degree of network is 13.26, the average path length is 3.41, and the average clustering coefficient is 0.69, which is considered as a tightly connected network between nodes.

(2) The fraction of nodes with degree distribution of the complex network of traditional Chinese cultural elements follow the power-law distribution, where $\gamma = 2.28$, indicating the heterogeneity of the network. That is, minority core nodes, in the dominant position of the network, exert significant influence and point out the direction of traditional cultural mining. The cluster-degree correlation analysis shows that the cluster coefficients of nodes with lower degrees are distributed more scattered and that of higher degree nodes follows the power law. Moreover, low-degree nodes can have a high cluster coefficient between which community structures connect to each other. The results mentioned above corroborate the existence of a hierarchical structure in the network.

(3) The results of community detection suggest that 91.77% of the elements in the complex network of traditional Chinese culture are from the top 12 communities (cultural topics), which are Confucianism, Taoism, military and diplomacy, social rule, political life, cosmos and astrology, natural disaster, landforms and geography, exotic customs, craft artifacts, myths and legends, and natural herbs. These topics are ring-wise distributed in Force Atlas force-directed layout. At last, applied to an orthogonal coordinate system with the Means-Ends of the culture, each topic's connotation and the relationships between topics are systematically explained.

This study provides a completely quantitative and reliable method of traditional Chinese cultural elements categorization. There are still a few limitations. (1) The textual data source is limited in *Temporal Imperial Encyclopedia* where the cultural topics may be imperfect. (2) The classification system is static. Although it reflects the structure of a cultural system, it may not be able to reflect the dynamic process of cultural system formation. (3) Further studies are needed in terms of the fitness of this system to other parts of Chinese culture such as the revolutionary culture, multiculturalism in modern society, and the cultural system of typical western, developed countries.

Data Availability

The data presented in this study are available on request from the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Key R&D Program of China (grant no. 2017YFB1400400), Youth Talent Promotion Program of Beijing Association for Science and Technology (grant no.2020-2022-16), Social Science Foundation of Beijing (grant no. 19GLC068), and Program for Promoting the Connotative Development of Beijing Information Science & Technology University (grant nos. 521201090A and 5026010961).

References

- [1] D. Qin, "An analytical model of Chinese national cultural characteristics," *Academics*, vol. 2, pp. 18–25, 2010.
- [2] S. Zhang, J. Zhong, and S. Zhou, "The application of traditional cultural elements in the opening of TV plays," *Contemporary TV*, vol. 12, pp. 62–64, 2019.
- [3] M. Lu, "The application of traditional cultural elements in domestic animation – drawing the bad people," *Contemporary TV*, vol. 3, pp. 94–95, 2017.
- [4] Y. Zhou, Y. Xu, and X. Shen, "Integration and inheritance: the use of Chinese cultural elements in animation brands," *Study and Practice*, vol. 6, pp. 129–133, 2019.
- [5] J. Jin and L. Ren, "The application of Chinese traditional cultural elements in the design of visual communication—commenting on study on modeling elements in visual communication design-the beauty of pictures and texts," *Journal of The Chinese Society of Education*, vol. 3, p. 119, 2019.
- [6] Z. Li, "The aesthetic construction of traditional cultural elements in tea packaging design," *Packaging Engineering*, vol. 41, no. 8, pp. 286–289, 2020.
- [7] C. Tian, Y. Wang, and J. Yang, "The impact of using Chinese cultural elements in hotels on customer perceived value," *Journal of Beijing Union University (Humanities and Social Sciences)*, vol. 18, no. 2, pp. 32–38, 2020.
- [8] S. Guan and M. Li, "A study on the classification of agricultural books in ancient Chinese bibliography," *Library Development*, no. 1, pp. 30–38, 2021.
- [9] L. Zhang and J. Wang, "Design of faceted classification system of ancient book databases," *Library Development*, no. 2, pp. 1–9, 2021.
- [10] S. Wang, D. Wang, S. Huang, and L. He, "Research on the automatic word segmentation of the book of songs under multi-dimensional domain knowledge," *Journal of The China Society for Scientific and Technical Information*, vol. 37, no. 2, pp. 183–193, 2018.
- [11] D. Wang, R. Gao, S. Shen, and B. Li, "Deep learning-based classification of pre-qin classics questions," *Journal of The China Society for Scientific and Technical Information*, vol. 37, no. 11, pp. 1114–1122, 2018.
- [12] L. He, Y. Qiao, and X. Liu, "Topic mining and evolution analysis of social development in spring and autumn period—a case of studying Zuo Zhuan," *Library and Information Service*, vol. 64, no. 7, pp. 30–38, 2020.
- [13] Z. Li, Z. Li, and L. He, "Study on the extraction method of war events in Zuo Zhuan," *Library and Information Service*, vol. 64, no. 7, pp. 20–29, 2020.
- [14] Z. Liu, J. Dang, and Z. Zhang, "Research on automatic extraction of historical events and construction of event graph based on historical records," *Library and Information Service*, vol. 64, no. 11, pp. 116–124, 2020.
- [15] R. Zhang, Y. Chen, and Y. Deng, "A review of community discovery in hybrid network for science structure analysis," *Library and Information Service*, vol. 63, no. 4, pp. 135–141, 2019.
- [16] H. Guo, B. Kong, and Z. Zhang, "Study on textual topic identification by clustering clique structure in multi-relationship text graph," *Journal of The China Society for Scientific and Technical Information*, vol. 36, no. 5, pp. 433–442, 2017.
- [17] Y. Tai, "The application of Chinese elements in cultural and creative industries," *China Economist*, vol. 7, pp. 232–233+235, 2019.
- [18] X. Wang, H. Yu, and T. Chen, "Identifying elements of nostalgia culture from a tourism perspective: taking the Ancient Huizhou cultural tourism area as case study," *Resources Science*, vol. 41, no. 12, pp. 2237–2247, 2019.
- [19] X. Wang, X. Zhang, and T. Chen, "Influencing factors of tourists' cognition of local nostalgic cultural elements: take Huizhou region as a case study," *Geographical Research*, vol. 39, no. 3, pp. 682–695, 2020.
- [20] E. A. Corrêa, A. A. Lopes, and D. R. Amancio, "Word sense disambiguation: a complex network approach," *Information Sciences*, vol. 442–443, pp. 103–113, 2018.
- [21] J. V. Tohalino and D. R. Amancio, "Extractive multi-document summarization using multilayer networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 503, pp. 526–539, 2018.
- [22] V. Q. Marinho, G. Hirst, and D. R. Amancio, "Authorship attribution via network motifs identification," in *Proceedings of the 2016 5th Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 355–360, Recife, Brazil, October 2016.
- [23] L. Qi, X. An, S. Zhang, and X. Wang, "Research on knowledge gap identification method in innovative organizations under the ‘Internet+’ environment," *Information*, vol. 11, no. 12, p. 572, 2020.
- [24] J. Cong and H. Liu, "Approaching human language with complex networks," *Physics of Life Reviews*, vol. 11, no. 4, pp. 598–618, 2014.
- [25] I. Bazzi, *Modelling Out-Of-Vocabulary Words for Robust Speech Recognition*, Massachusetts Institute of Technology, Cambridge, MA, USA, 2002.
- [26] L. Du, X. Li, G. Yu, C. Liu, and R. Liu, "New word detection based on an improved PMI algorithm for enhancing segmentation system," *Acta Scientiarum Naturalium Universitatis Pekinensis*, vol. 52, no. 1, pp. 35–40, 2016.
- [27] Z. Li and M. Sun, "Punctuation as implicit annotations for Chinese word segmentation," *Computational Linguistics*, vol. 35, no. 4, pp. 505–512, 2009.
- [28] M. Ortúñoz, P. Carpéna, P. Bernaola-Galván, E. Muñoz, and A. M. Somoza, "Keyword detection in natural languages and DNA," *Europhysics Letters (EPL)*, vol. 57, no. 5, pp. 759–764, 2002.
- [29] J. Zhang, "A method of intelligence Key words extraction based on improved TF-IDF," *Journal of Intelligence*, vol. 33, no. 4, pp. 153–155, 2014.
- [30] J. Wang and T. Qiu, "Focused topic Web crawler based on improved TF-IDF algorithm," *Journal of Computer Applications*, vol. 35, no. 10, pp. 2901–2904+2919, 2015.
- [31] L. Leydesdorff, "On the normalization and visualization of author co-citation data: salton's Cosineversus the Jaccard index," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 1, pp. 77–85, 2008.
- [32] F. Li, J. Zhang, and Z. Wang, "Review of social recommendation with bibliometrics and social network analysis," *Data*

- Analysis and Knowledge Discovery*, vol. 1, no. 6, pp. 22–35, 2017.
- [33] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
 - [34] A. Clauset, C. R. Shalizi, and M. E. J. Newman, “Power-law distributions in empirical data,” *SIAM Review*, vol. 51, no. 4, pp. 661–703, 2009.
 - [35] E. Ravasz and A.-L. Barabási, “Hierarchical organization in complex networks,” *Physical Review E*, vol. 67, no. 2, p. 026112, 2003.
 - [36] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York, NY, USA, 1994.
 - [37] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical Review E*, vol. 69, no. 2, p. 026113, 2004.
 - [38] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 12, no. 10, p. P1000, 2008.
 - [39] E. Schein, “Organizational culture and leadership: a dynamic view,” *Organization Studies*, vol. 7, pp. 199–201, 1985.
 - [40] R. E. Quinn and J. Rohrbaugh, “A spatial model of effectiveness criteria: towards a competing values approach to organizational analysis,” *Management Science*, vol. 29, no. 3, pp. 363–377, 1983.
 - [41] T. Mao, *Selected Works of Mao Tse-Tung*, Elsevier, Amsterdam, Netherlands, 2014.