Research Article

Cluster Analysis and Visualization for the Legend of the Condor Heroes Based on Social Network

Chao Fan [ID],[1,2] **Zhihui Yang,**[1,2] **and Yuyi Yuan**[1,2]

[1]*The School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China*
[2]*Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China*

Correspondence should be addressed to Chao Fan; fanchao@jiangnan.edu.cn

The Legend of the Condor Heroes (LCH) is one of the fifteen well-known Wuxia novels penned by Jin Yong. It portrays a number of characters in the background of the Southern Song Dynasty. In this research, we attempt to analyze the relationship of characters in LCH based on social network, including network feature analysis, cluster analysis, and data visualization. Moreover, the approach can be extended to other literary works because our research provides a general framework for analyzing character relationships. We first perform lexical analysis on the corpus to extract character names and then utilize co-word analysis to build a social network of character relationships. We reckon characters as nodes and count the cooccurrences of characters as weights of links. By applying the social network analysis of created network, we can obtain network features of LCH. Furthermore, a hierarchical clustering algorithm is implemented to study the structure of LCH network. Both the dendrogram and Venn diagram are used for data visualization. An improved approach of visualizing the clustering results is advanced in order to display the group and hierarchical structure better. The final experimental results demonstrate that the proposed method shows a good effect.

1. Introduction

The Legend of the Condor Heroes, written by JIN Yong, is one of the most reputable novels painting the fantasy world of Wuxia. Based on real historical events, this novel tells the story of the Southern Song Dynasty. It describes the legend of Guo Jing, who is destined to shoulder the responsibility of the country and grow into a hero with the help of Huang Rong. The novel portrays about a hundred dramatic characters.

Relationships of characters can be explored from the perspective of social network. This paper tries to build a character relationship network based on the co-word analysis. It is well known that the cooccurrence of two characters' names in a paragraph or a sentence indicates a connection between them in some ways. Hence, the co-word analysis technology can be applied to the construction of character relationship network. At the beginning, textual analysis of LCH corpus is carried out by natural language processing such as segmentation, part-of-speech tagging, etc. Utilizing the character names and their cooccurrences, we are able to build a weighted undirected network. Many network features are calculated to describe the characteristics of LCH network structure. Furthermore, the agglomerative hierarchical clustering method is implemented to analyze the relationships of characters. Some interesting conclusions of the novel can be drawn. Finally, an improved approach by employing the Venn diagram is proposed to visualize clustering results. By conducting cluster analysis and data visualization, not only does this work facilitate the exhibition of character relationships in LCH, but also it provides a framework for analyzing characters in the literary work.

The content of this paper can be presented in the following sections. Related studies are reviewed in Section 2. Social network construction is elaborated in Section 3. Section 4 states experimental results, discussions of cluster analysis, and data visualization. Section 5 gives the conclusion and future work.

2. Related Work

Quantitative research is an effective strategy to analyze literature, among which co-word analysis attracts the attention of many scholars. Co-word analysis is an important technology to build a social network. The co-word analysis approach was developed by French bibliographers [1] and soon introduced to other fields [2–4]. Some researchers [5] utilized the co-word analysis to study the theme of publication in the field of scientometrics. Their research results suggest that topics of publications often follow hot spots. Zhu and his team [1] concentrated on the scientific literature concerning social computing and found some hot research topics by using the co-word analysis. Gan et al. [6] leveraged co-word analysis tools to analyze the medical subject headings (MeSH) terms of epilepsy genetics. Nguyen [7] paid attention to the interdisciplinary literature on non-biomedical topics from 1987 to 2017. A large-scale co-word analysis was exploited for mapping knowledge. Another research [8] focused on the evolution of medical tourism. It analyzed the data from the Web of Science and Scopus based on co-word analysis and gave a visualization of all subfields. According to recent research [9], open data were employed to investigate knowledge areas, themes, and future research based on co-word analysis. Khaldi and Prado-Gascó [10] discussed international cooperation issues on migration by using co-word analysis of articles in Web of Science. They gave some frequent keywords and disclosed that most active authors in this field came from a few countries, such as the United States, the United Kingdom, Germany, etc.

Several authors explored the relationships of characters of literature by utilizing social network analysis, cluster analysis, and data visualization. Zhao et al. [11] extracted character relationships from Chinese literary based on social network. They discovered that the character network in Chinese literature had apparent small-world property and limited power-law distribution. Wang et al. [12] proposed a research method to analyze the historical characters in the Romance of the Three Kingdoms (RTK). They created a network according to the cooccurrence of character names and found some interesting phenomena in the RTK. Fan [13] established a character relationship network of the Dream of the Red Chamber. They calculated the network features and performed a cluster analysis. Some interesting conclusions can be drawn through the visualization of experimental results. Hu [14] made an effort to construct the character network in Records of the Three Kingdoms. Natural language processing (NLP) tools were taken into account to handle the text at first. Then segmentation and part-of-speech tagging were carried out. A custom dictionary was adopted to process the ancient Chinese text. Anaphora resolution was also used to determine which character a pronoun is referring to. Besides, the author applied a k-kernel decomposition approach to analyze the significant characters in the work. Xu [15] studied the classical literature Zuozhuan in the Pre-Qin period from the perspective of language network and social network. Both ancient and modern Chinese versions were selected as research objects. Language network considered the literature as language data and preprocessed it by NLP technologies, whereas social network reckoned the work as historical documents and focused on the character relationships.

3. Creation of Social Network

3.1. Preprocessing of LCH Corpus. We downloaded a full text of the Legend of the Condor Heroes (LCH) in Chinese dataset (https://www.uidzhx.com/Shtml604.html). Data cleansing was done automatically and some punctuation errors were corrected manually. Also, mistakes of character names were revised according to the name list of LCH acquired from the Internet. After original dataset preprocessing, the created corpus can be used for lexical analysis.

The Chinese lexical analyzer ICTCLAS (http://ictclas.nlpir.org/) was employed to process the LCH corpus. Chinese segmentation was performed on each sentence and parts-of-speech were tagged for segmented words. Exploiting the results of segmentation and part-of-speech tagging, we succeeded in extracting the character names which were used for creating LCH network.

3.2. Construction of LCH Network. When generating nodes of the network, the full name, given name, and nickname of a character are regarded as one node. For instance, "Hong Qigong," "Qigong," and "Bei Gai" represent the same character. The cooccurrence of two names in a specified area (a paragraph or a sentence) denotes a link between two characters. The corresponding frequency indicates the weight of a link. In this paper, the contextual area is set as the paragraph.

We built a weighted undirected network to describe the relationships of characters in LCH based on co-word analysis. The established network incorporates 90 nodes and 1,013 links. A visualized LCH network is depicted in Figure 1 where degrees of nodes are ranked in the top 50. The size of a node represents the degree of the node.

3.3. Network Features

3.3.1. Degree Distribution. The degree of a node is defined as the count of its neighbors. The degree of a character means how many other characters are linked to him or her, thus implying the significance of a character. In the novel, a character may be directly connected to other characters or talked about in others' conversations.

As can be seen from Figure 1, Guo Jing has the biggest value of the degree. The top ten people in degree ranking are as follows: Guo Jing, Huang Rong, Yang Kang, Qiu Chuji, Ke Zhene, Wanyan Honglie, Zhu Cong, Huang Yaoshi, Ouyang Feng, and Tuolei. We calculated the average degree of LCH network and obtained a value of 22.51. This number shows that a character in LCH is linked to an average of other 22.5 characters. That is to say, characters in LCH network are highly connected to each other.

Degree distribution $p(k)$ offers the probability that a node is related to k other nodes [16]. The degree distribution of LCH network is presented in Figure 2. A
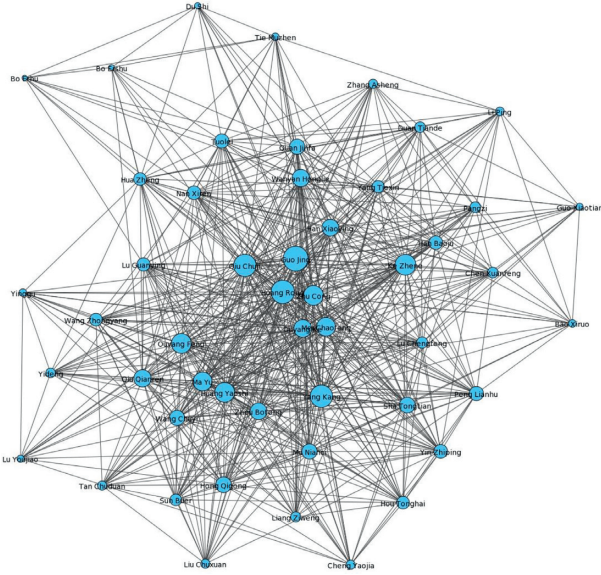
FIGURE 1: Data visualization of LCH network (top 50 in node degree).

power-law distribution cannot be found in LCH dataset because the novel of JIN Yong often portrays important characters in the Wuxia world, which is different from the results of RTK [17].

3.3.2. Average Shortest Path Length.
The shortest path length [17] measures the number of links that pass through two nodes with the shortest path. Therefore, the average shortest path length can be computed as an average of every shortest path length. This indicator evaluates the reachability of two nodes in the network. LCH network has an average shortest path length of 1.7958. In other words, a character is able to reach any other character with less than two steps. Additionally, the diameter of LCH network is 4. The diameter is defined as the largest number in all shortest path lengths. One of the longest paths is listed as follows: Qu Shagu, Yang Kang, Qiu Chuji, Tie Muzhen, Hu Duhu.

Figure 3 illuminates the distribution of the shortest path length for LCH network. According to the figure, about 67.89% of the shortest path is composed of length 2 and over 93.18% ranges from length 1 to length 2.

3.3.3. Clustering Coefficient.
The clustering coefficient [18] is defined as the average probability that neighbors of one node are connected. According to Figure 4, the clustering coefficient of node A in Figure 4(b) is larger than the one in Figure 4(a) because its adjacent nodes are connected to each other.

Suppose d_A represents the degree of node A. The number of potential links among its neighbors can be calculated by $(1/2)d_A(d_A - 1)$. If S_A is the number of existing links among its neighbors, clustering coefficient C_A of node A can be written by the following formula:

$$C_A = \frac{2S_A}{d_A(d_A - 1)}. \tag{1}$$

As a consequence, the clustering coefficient of a network is the average of all C_A, where N is the number of all nodes.

$$C = \frac{1}{N} \sum_A C_A. \tag{2}$$

In this paper, Erdős-Rényi model [19, 20] is utilized to create a random network by adopting the same number of nodes and links as LCH network. According to the experimental result illustrated in Table 1, LCH network has a relatively high clustering coefficient compared with a random network, thereby showing a tendency that characters incline to cluster together.

3.3.4. Density.
The network's density is the number of existing links divided by the number of possible links [21]. If N and E denote the number of nodes and links in a network, the density of the network can be given by formula (3). There may be up to 4,005 potential links for 90 nodes, among which 1,013 links actually exist. Consequently, the density of the LCH network is 0.2529:

$$\text{density} = \frac{2E}{N(N-1)}. \tag{3}$$

3.3.5. Centrality.
Three centrality indicators are applicative for measuring the significance of a node, including degree, betweenness, and closeness centrality. They are often used to discover the most influential people in a network.

Degree centrality is the node degree divided by the highest degree in a network. For instance, if the node with the highest degree owns 100 connected neighbors in a network, a node with 20 links would have a degree centrality of 0.2. A node having a high degree is reckoned as a local hub in the network. The top ten characters in degree centrality ranking are listed in the second column of Table 2. The centrality value is included in parentheses. Guo Jing has the largest degree centrality 0.9213, which is consistent with the result in Figure 1.

For every pair of nodes in a network, there is at least one shortest path between them. The betweenness centrality for a node is the ratio of these shortest paths that pass through the node. A high-betweenness node is able to control the interactions of other nonadjacent nodes [22]. The top ten characters with the highest betweenness centralities are presented in the third column of Table 2. Tie Muzhen, aka Genghis Khan, has a high-betweenness centrality even though his degree centrality is not in the top ten. According to Figure 5, a heavy-tailed distribution can be seen in the distribution of betweenness for characters in LCH. The sum of the top 10 covers 70.17% of all betweenness centralities. The sum of the first 20 betweenness accounts for 85.52% of the total values.
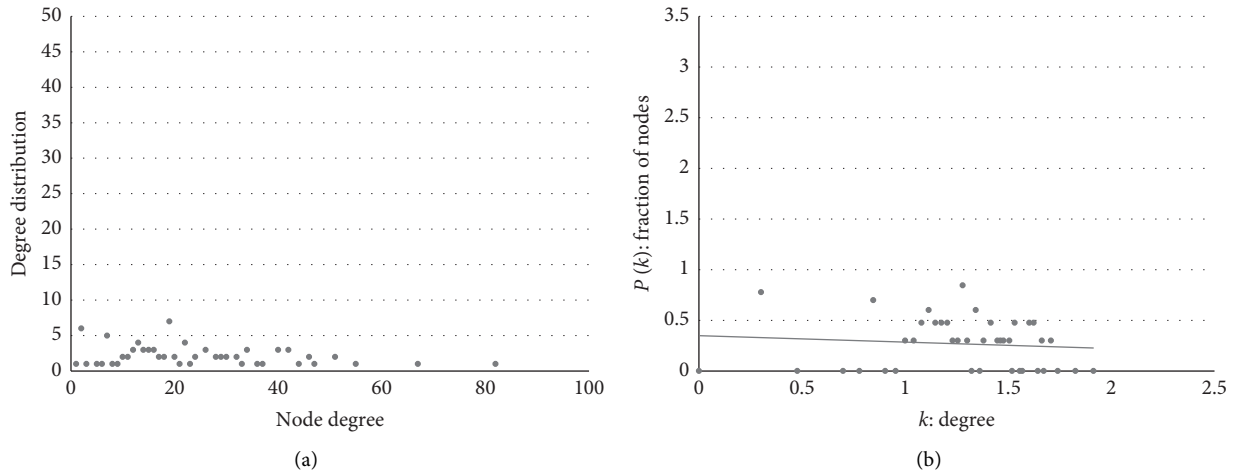
(a)

(b)

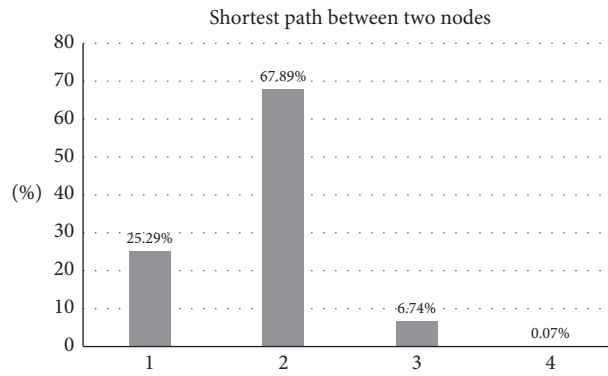FIGURE 2: Degree distribution of LCH network and its log-log plot.

FIGURE 3: Distribution of shortest path length in LCH network.
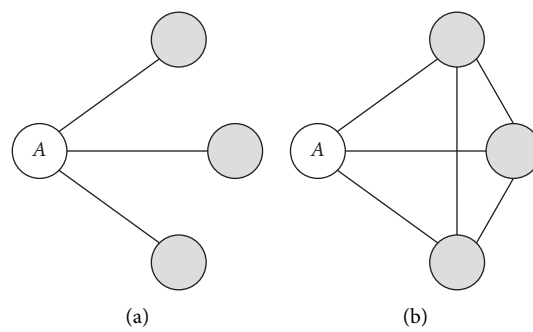

(a)   (b)

FIGURE 4: A comparison of clustering coefficient for node A.

TABLE 1: Network features of LCH network and random network.

	Number of nodes	Number of links	Clustering coefficient
LCH network	90	1013	0.7262
Random network	90	1013	0.2524

TABLE 2: Top 10 characters in centrality ranking.

| Ranking	Degree centrality	Betweenness centrality	Closeness centrality
1	Guo Jing (0.9213)	Guo Jing (0.2323)	Guo Jing (0.9271)
2	Huang Rong (0.7528)	Huang Rong (0.0863)	Huang Rong (0.7946)
3	Yang Kang (0.6180)	Wanyan Honglie (0.0551)	Yang Kang (0.7177)
4	Qiu Chuji (0.5730)	Yang Kang (0.0484)	Qiu Chuji (0.7008)
5	Ke Zhene (0.5730)	Yang Tiexin (0.0333)	Ke Zhene (0.7008)
6	Wanyan Honglie (0.5281)	Ke Zhene (0.0296)	Wanyan Honglie (0.6794)
7	Zhu Cong (0.5169)	Qiu Chuji (0.0277)	Zhu Cong (0.6742)
8	Huang Yaoshi (0.5169)	Tuolei (0.0264)	Huang Yaoshi (0.6593)
9	Ouyang Feng (0.4944)	Tie Muzhen (0.0241)	Tuolei (0.6544)
10	Tuolei (0.4719)	Hong Qigong (0.0224)	Ouyang Feng (0.6496)
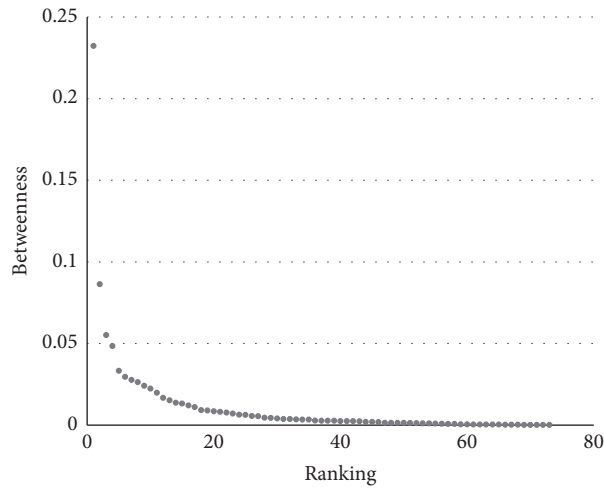
FIGURE 5: Distribution of betweenness for characters in LCH.

Closeness centrality [7] is specified by the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the network. It has the capability to estimate the degree of closeness for a node to reach other nodes in the network. The values of the top 10 closeness centralities in LCH network are specified in the fourth column of Table 2. Characters with high closeness centralities may rapidly communicate with others in the network.

According to Table 2, seven characters appear in the three columns of centrality ranking. They are Guo Jing, Huang Rong, Yang Kang, Qiu Chuji, Ke Zhene, Wanyan Honglie, and Tuolei, who are considered as the important characters in the novel.

4. Cluster Analysis and Visualization

4.1. Hierarchical Clustering

4.1.1. Cooccurrence Matrix and Similarity Matrix. Two matrices are raised to provide data for clustering. Given the frequency that two nodes appear in the same context, the cooccurrence matrix can be constructed based on co-word analysis. An example of cooccurrence matrix for six major characters in the LCH is illustrated in Table 3. The number on the diagonal indicates the count of character names appearing in the novel.

Since the weight of a link is affected by its adjacent nodes' degrees, we build the similarity matrix by calculating Ochiai coefficient [22]. Formula (4) gives the similarity of sets X and Y:

$$\text{sim}(X, Y) = \frac{n(X \cap Y)}{\sqrt{n(X) \times n(Y)}}. \tag{4}$$

Here, $n(X)$ represents the frequency of X. $n(X \cap Y)$ is the cooccurrence number of X and Y. By applying the Ochiai coefficient, the cooccurrence matrix can be transformed into a correlation matrix, which is a similarity matrix. The result of the similarity matrix for the six main characters is presented in Table 4.

4.1.2. Clustering Algorithm. Hierarchical clustering is utilized to create a tree-like structure. It embraces two strategies: divisive [23] and agglomerative [24] clustering. We devised an agglomerative algorithm to cluster LCH characters into groups. Similarity matrix based on Ochiai coefficient is taken as input dataset of clustering. Agglomerative method is selected to implement the clustering algorithm. To start with, we regard each character as one group representing the cluster. At each time, two clusters that have the largest similarity are merged into a

TABLE 3: Cooccurrence matrix of six characters.

Cooccurrence	Guo Jing	Huang Rong	Yang Kang	Mu Nianci	Huang Yaoshi	Ouyang Feng
Guo Jing	1995	1197	145	50	242	361
Huang Rong	1197	1856	117	84	160	298
Yang Kang	145	117	424	104	20	38
Mu Nianci	50	84	104	251	7	6
Huang Yaoshi	242	160	20	7	586	157
Ouyang Feng	361	298	38	6	157	860

TABLE 4: Similarity matrix of six characters.

Cooccurrence	Guo Jing	Huang Rong	Yang Kang	Mu Nianci	Huang Yaoshi	Ouyang Feng
Guo Jing	1	0.622062	0.157657	0.070658	0.223818	0.275605
Huang Rong	0.622062	1	0.131891	0.12307	0.15342	0.235873
Yang Kang	0.157657	0.131891	1	0.318797	0.040123	0.062929
Mu Nianci	0.070658	0.12307	0.318797	1	0.018252	0.012914
Huang Yaoshi	0.223818	0.15342	0.040123	0.018252	1	0.221157
Ouyang Feng	0.275605	0.235873	0.062929	0.012914	0.221157	1

larger cluster. When all clusters are grouped into one cluster, the clustering process finishes. In addition, a cluster-stopping threshold can also be set so as to control the final number of clusters.

4.2. Result of Clustering. The clustering result of LCH characters can be visualized through a tree-like structure called dendrogram [21]. In this paper, we separated the whole dendrogram into groups with different numbers. A dendrogram example of LCH characters is depicted in Figure 6.

According to the figure, six groups are marked by red lines. Most characters in H1 are from "Quanzhen Taoism," such as Ma Yu, Wang Chuyi, Liu Chuxuan, etc. H2 contains the most important characters of the novel, like Guo Jing, Huang Rong, Yang Kang, and Mu Nianci. Many martial arts masters are also included in this group: Huang Yaoshi, Ouyang Feng, Hong Qigong, and Yideng. H3 incorporates the "seven heroes" and the "Twice Foul Dark Wind" in the novel. Both of them are closely linked groups and can merge into a bigger group. Further, most characters in H4 are from a gang named "Gai." They are homeless but organized by some leaders. Characters in the H5 group are from the Mongolian nation, e.g., Tie Muzhen (Genghis Khan) and Tuolei (Tolui). Finally, H6 is composed of "Four ghosts of the Yellow River" and the remainder. Six parts make up the Wuxia world of LCH.

4.3. Improved Visualization Approach. The result of hierarchical clustering is in essence a binary tree. As shown in Figure 7, a tree can be displayed in the form of a Venn diagram [25] with the hierarchical structure. The nodes C, D, and E are leaf nodes, representing characters in the novel, whereas the nodes A and B refer to groups with

different sizes. Using the Venn diagram to display hierarchical clustering results has many advantages. For instance, it can clearly express the group features and hierarchical structure.

Nonetheless, there may be some disadvantages to this visualization approach. Since 90 characters appear in the novel, the number of nodes should be 90 and the number of groups would be 89. As a result, there will be 179 circles in the Venn diagram, leading to a bad visualization effect (see left side in Figure 8).

When displaying the characters, it is not necessary to draw all group circles in the hierarchy. In order to reduce the number of group circles, a merge operation is introduced in this research, which is shown in Figure 8. By removing the group circle B and G, or even A and F, two forms of diagrams are obtained in the right side of Figure 8.

In hierarchical clustering, we can control the number of clusters and merge nodes with the same cluster label. A visualization of LCH characters in five clusters is given in Figure 9. Different groups are filled with different colors. According to the figure, our proposed visualization approach is capable of showing the hierarchical structure of clusters in a clear manner. From the visualized clustering results, we can discover two major groups: Chinese groups and Mongolian groups. As Guo Jing grew up in Mongolia, the novel portrays a number of characters in Mongolian groups in the early chapters. In Chinese groups, the largest group contains the most significant roles in the LCH. Then, "the seven heroes" group and the largest group can merge into a bigger one. As the merge process goes on, the "Quanzhen Taoism" and "Four ghosts of the Yellow River" groups also joined the Chinese groups. Finally, characters in both Chinese and Mongolian groups make up the Wuxia world of LCH.
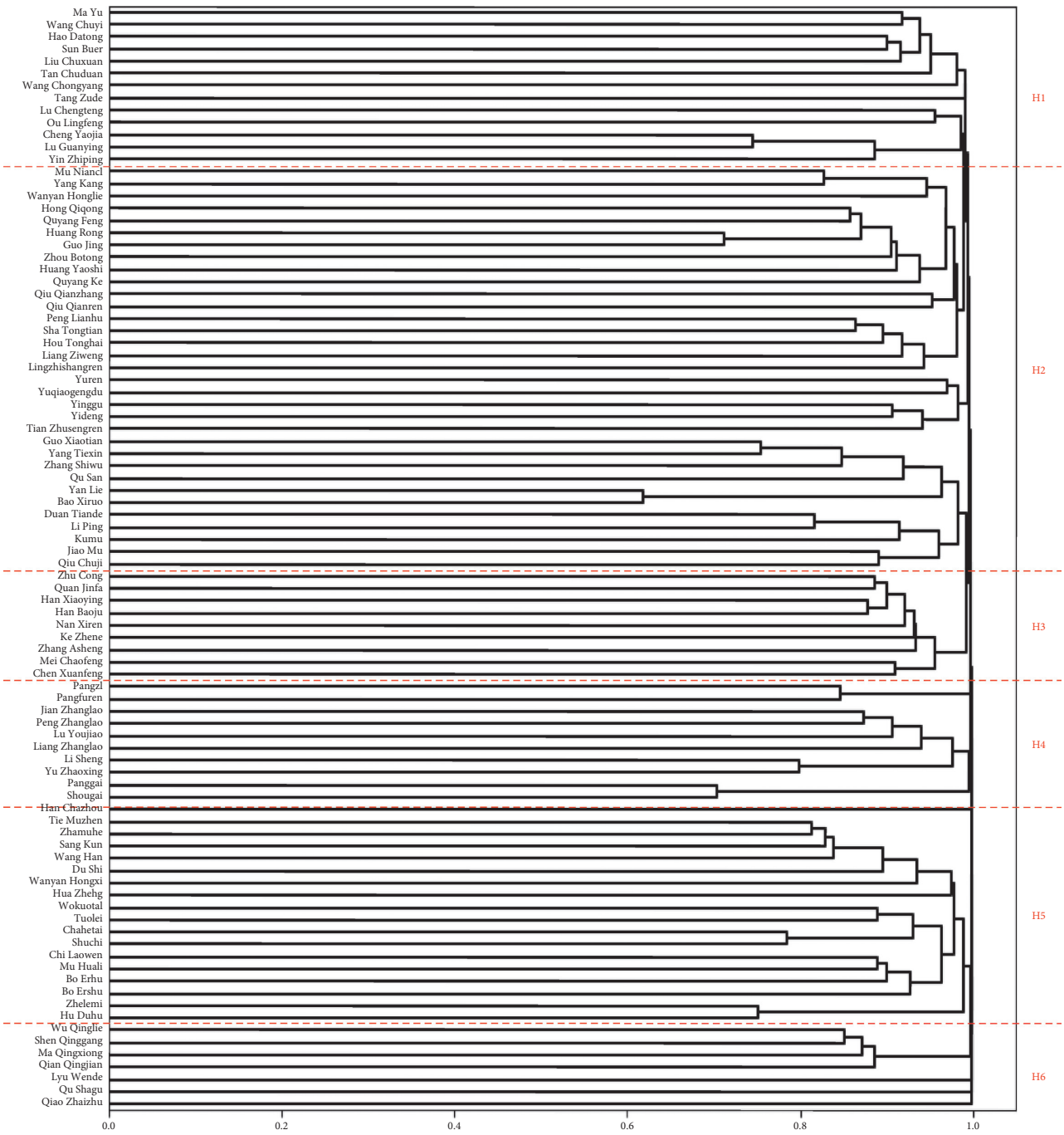
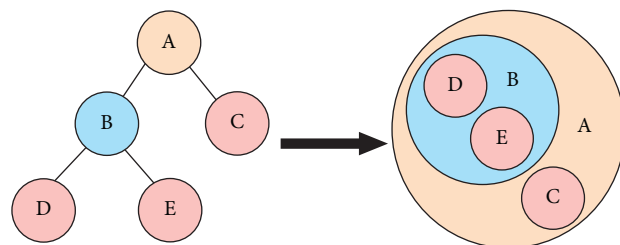FIGURE 6: Dendrogram of clustering result for LCH network.

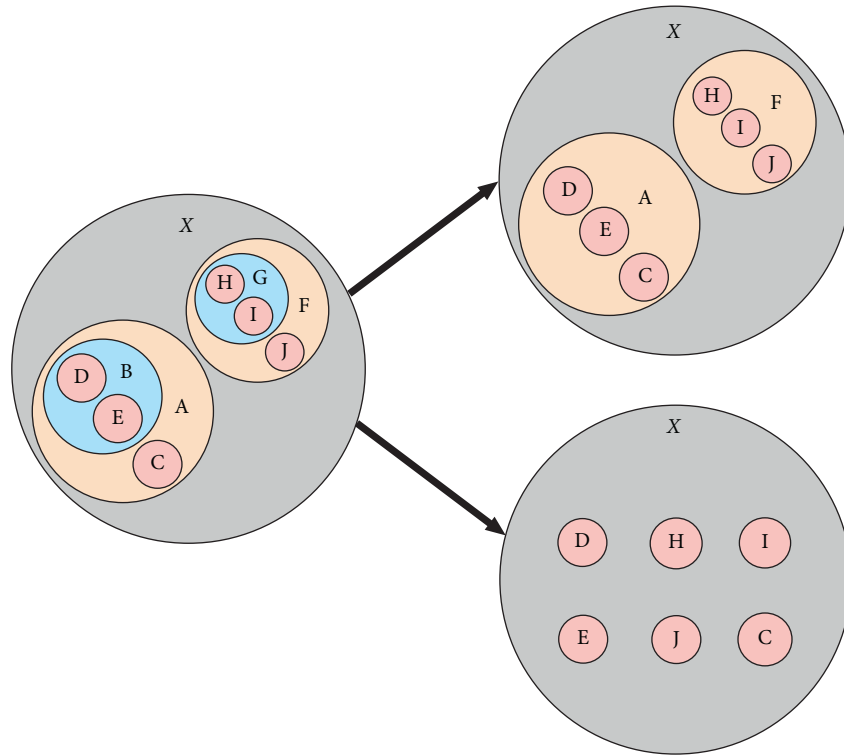FIGURE 7: Two visualization approaches of tree structure.

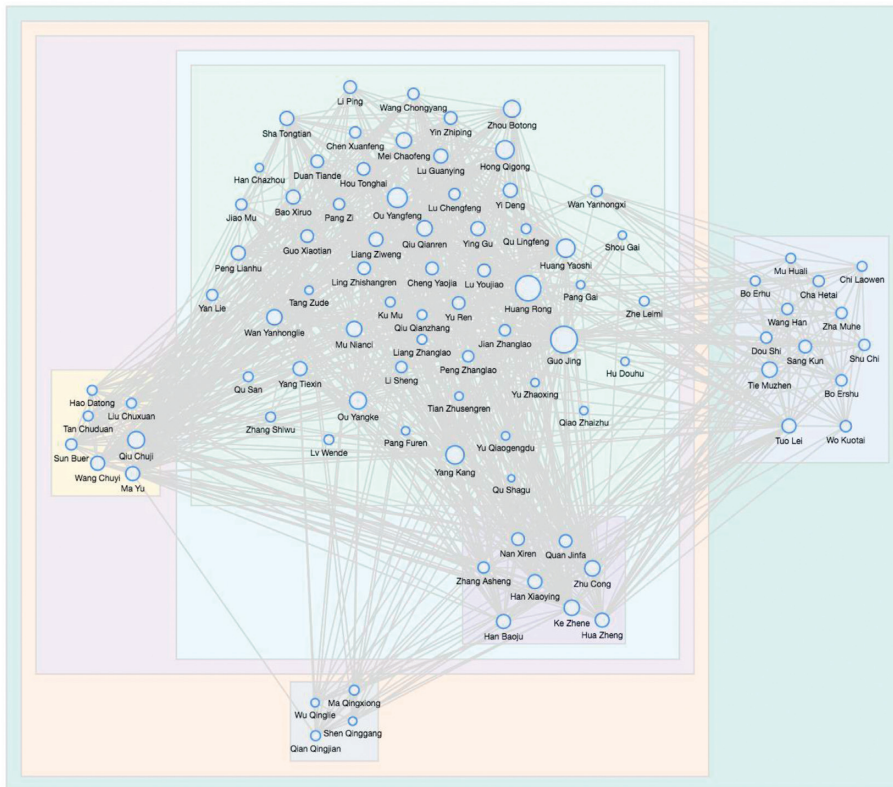FIGURE 8: The merge operation of small groups in the Venn diagram.

FIGURE 9: An example of the improved visualization method.

5. Conclusions

This paper concentrates on the analysis and visualization of character relationships in the "Legend of the Condor Heroes (LCH)." The Chinese version of LCH was selected as the research object. At the beginning, Chinese segmentation and part-of-speech tagging were enforced to preprocess the LCH corpus. Adopting the technology of co-word analysis, character names and cooccurrence of names in the context were thought to be nodes and links in the network. We then constructed a weighted undirected network of relationships of characters in LCH. Based on the social network analysis, computation of network features was completed such as centrality, clustering coefficient, density, etc. Furthermore, hierarchical clustering was performed and a dendrogram was sketched. An improved visualization approach of Venn diagram was also proposed to exhibit the effect of hierarchical clustering. From the experimental result, we can identify two major groups (Chinese and Mongolian groups) and the hierarchical structure in Chinese groups.

However, there are some disadvantages in this research. The cooccurrence of character names in the context may not represent the real relationships of characters. Nevertheless, the proposed quantitative research method provides a new perspective to analyze the characters in the novel. In the future, coreference resolution will be taken into consideration to build accurate relationships of characters. Moreover, we expect to explore the semantic analysis of texts to refine the experimental result.

Data Availability

The LCH dataset used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] Q. Zhu, X. Peng, and X. Liu, "Research topics in social computing area based on co-word analysis," *Information Studies: Theory & Application*, vol. 12, pp. 7–11, 2012.

[2] M. Callon, J.-P. Courtial, W. A. Turner, and S. Bauin, "From translations to problematic networks: an introduction to co-word analysis," *Social Science Information*, vol. 22, no. 2, pp. 191–235, 1983.

[3] Q. He, "Knowledge discovery through co-word analysis," *Library Trends*, vol. 48, no. 1, pp. 133–159, 1999.

[4] X. Wang, "Formation and evolution of science knowledge network (I): a new research method based on co-word network," *Journal of the China Society for Scientific and Technical Information*, vol. 28, no. 4, pp. 599–605, 2009.

[5] S. Ravikumar, A. Agrahari, and S. N. Singh, "Mapping the intellectual structure of scientometrics: a co-word analysis of the journal scientometrics (2005–2010)," *Scientometrics*, vol. 102, no. 1, pp. 929–955, 2015.

[6] J. Gan, Q. Cai, P. Galer et al., "Mapping the knowledge structure and trends of epilepsy genetics over the past decade: a co-word analysis based on medical subject headings terms," *Medicine*, vol. 98, no. 32, Article ID e16782, 2019.

[7] D. Nguyen, "Mapping knowledge domains of non-biomedical modalities: a large-scale co-word analysis of literature 1987–2017," *Social Science & Medicine*, vol. 233, pp. 1–12, 2019.

[8] A. de la Hoz-Correa, F. Muñoz-Leiva, and M. Bakucz, "Past themes and future trends in medical tourism research: a co-word analysis," *Tourism Management*, vol. 65, pp. 200–211, 2018.

[9] D. Corrales-Garay, M. Ortiz-de-Urbina-Criado, and E.-M. Mora-Valentín, "Knowledge areas, themes and future research on open data: a co-word analysis," *Government Information Quarterly*, vol. 36, no. 1, pp. 77–87, 2019.

[10] H. Khaldi and V. Prado-Gascó, "Bibliometric maps and co-word analysis of the literature on international cooperation on migration," *Quality and Quantity*, vol. 2021, no. 1, pp. 1–25, 2021.

[11] J. Zhao, L. Zhang, Q. Zhu, and G. Zhou, "Extracting and analyzing social networks from Chinese literary," *Journal of Chinese Information Processing*, vol. 31, no. 2, pp. 99–106, 2017.

[12] Y. Wang, J. Yu, and C. Zhao, "Research on application of co-word analysis on relationships of characters in the romance of the three kingdoms," *Information Research*, vol. 7, pp. 52–56, 2017.

[13] C. Fan, "Research on relationships of characters in the dream of the red chamber based on co-word analysis," *ICIC Express Letters Part B: Applications*, vol. 11, no. 5, pp. 1–8, 2020.

[14] Y. Hu, *Construction and Analysis of Social Networks in Chinese Literature: A Case Study on Three Kingdoms*, Zhejiang Gongshang University, Hangzhou, China, 2019.

[15] C. Xu, *Research into the Language Network and Social Network of Zuozhuan*, Nanjing Normal University, Nanjing, China, 2014.

[16] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[17] C. Fan and Y. Li, "Coword and cluster analysis for the romance of the three kingdoms," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5553635, 8 pages, 2021.

[18] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[19] B. Bollobás, *Random Graphs*, Cambridge University Press, Cambridge, UK, 2nd edition, 2001.

[20] M. E. J. Newman, *Networks*, Oxford University Press, Oxford, UK, 2nd edition, 2018.

[21] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, UK, 1994.

[22] Q. Zhou and L. Leydesdorff, "The normalization of occurrence and co-occurrence matrices in bibliometrics using cosine similarities and Ochiai coefficients," *Journal of the Association for Information Science & Technology*, vol. 67, no. 11, pp. 1–25, 2016.

[23] A. Ishizaka, B. Lokman, and M. Tasiou, "A stochastic multi-criteria divisive hierarchical clustering algorithm," *Omega*, vol. 103, Article ID 102370, 2021.

[24] N. Liu, Z. Xu, X.-J. Zeng, and P. Ren, "An agglomerative hierarchical clustering algorithm for linear ordinal rankings," *Information Sciences*, vol. 557, pp. 170–193, 2021.

[25] N. B. J. Bipin and M. S. Sarath, "Evolutionary clustering annotation of ortho-paralogous gene in a multi species using Venn diagram visualization," *International Journal of Engineering & Technology*, vol. 7, no. 1, pp. 162–166, 2018.