

Knowledge Map Construction and Visualization with Weight

Ming Zhao, Yingxu Tai, Shoujin Wang* and Yiwen Li

School of Computer Science and Engineering, Shenyang Jianzhu University, Shenyang, Liaoning, China

**Corresponding author. Email: wangshoujin@sjzu.edu.cn*

ABSTRACT

Knowledge map construction and visualization of characters and related information in novel and films can help readers or audiences understand the content of related works more accurately and quickly. Taking the novel "In the Name of the People" as an example, the occurrence frequency of main characters and co-occurrence frequency between characters are introduced into the knowledge map in the form of weights to form the knowledge map of characters and relationships with weight, and the Web programming technology of JavaScript is used to visualize it. It is compared with the knowledge map constructed by graph database Neo4j. Compared with the knowledge map constructed by Neo4j, the weighted knowledge map of people and relationships can display the importance with weight of characters and the closeness of the connection between characters more intuitively.

Keywords: Knowledge Map, Weight, Visualization

1. INTRODUCTION

With the rapid development of economy and the improvement of science and technology, people need to make full use of fragmented time to meet their spiritual needs, among which reading novels and watching movies and TV works are the most common means. The study of character relationships and their visualization, through the analysis and visualization of characters and their relationships in novels and works, makes readers and audiences can more accurately, intuitively and quickly identify the relationship between the characters, so as to better understand the novel and the content of the work without their own subjective thinking. Therefore, it has important research significance and application value. In the study of the relationship between famous characters and novels, Yang Xin et al studied the relationship between characters by using co-word analysis [1]. Chen Lei et al. analyzed the social hierarchy among characters in A Dream of Red Mansions through the snowball algorithm [2]. Xiao Tianjiu et al. analyzed the novels of Jin Yong and Gu Long and used cluster analysis to prove that the styles of jin Yong and Gu Long differ greatly in multiple dimensions [3]. In the research on the visualization of character relations, Chen Zhangzhang described complex and diverse social relations in a more intuitive and understandable form by constructing multidimensional character relations and combining visualization methods [4]. Through clustering analysis and association analysis, Si Danni deeply mined the attribute information of characters [5]. Xiang Fan et al. constructed a family tree through RT algorithm for a further and more

detailed study in the study of the relationship between people displayed by knowledge graph [6]. Feng Yuanwei constructed the knowledge graph of character relationship through design to meet the requirements of querying character relationship [7]. Chen Qian used the sliding window method to expand and optimize the relationship feature phrases and build the microblog topic relationship knowledge map [8]. Drakopoulos G et al. stored transformed character relation data through graph database, and then inquired and analyzed character relation through speech in graph database [9-11]. In the study of character relationship intimacy, Li Jiao combined the association rules in co-occurrence and data mining, and used the common keywords to reflect the degree of closeness between the two [12]. Wang Yibo et al. took Romance of The Three Kingdoms as the research object, generated the co-occurrence matrix of the novel character relations and obtained the visualization results [13].

To facilitate the readers and the audience more intuitive grasp the related work of character and character relationships to the importance of the tightness of the paper on the basis of the contemporary novels "in the name of the people" as an example, introduces the concept of weight to the knowledge map to extend it to man in frequency as the weight of the characters, the co-occurrence frequency as the weight of characters relations. At the same time, the weighted knowledge graph is visualized using Web technology, and compared with the construction and visualization effect of common knowledge graph.

2. RESEARCH FRAMEWORK

The research framework of this paper is shown in Figure 1, and the specific process is as follows:

(1) Data acquisition, processing and analysis. Download the electronic version of the novel, use word segmentation to get the list of characters, get the frequency of characters, on this basis, analyze the relationship between characters and establish the co-occurrence matrix of characters.

(2) Construction of knowledge graph and knowledge graph with weight. Based on the

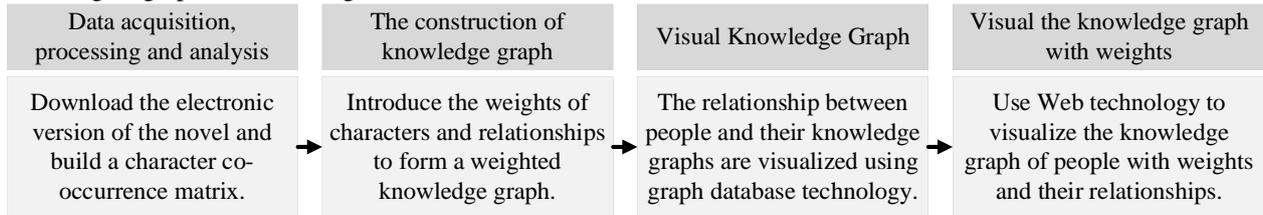


Fig 1. Research framework

3. RESEARCH PROCESS

3.1 Data Acquisition, Processing and Analysis

The data in this paper is derived from downloading high-quality electronic version of the novel In the Name of The People from the Internet, and then converting the novel into a text file. Data processing is conducted on the text document, Chinese word segmentation is conducted in the unit of period, and the main character names of the novel are identified, so as to further analyze and summarize the relationship between the characters. The steps are as follows:

(1) Download the high-quality electronic version of the novel, convert it into text format, write the program in Python language, use Jieba to perform

Table 1. Frequency of main characters in the novel (part)

Name	Hou Liangping	Li Dakang	Gao Yuliang	Qi Tongwei	Sha Ruijin
Frequency	1355	780	627	553	353
Name	Gao Xiaoqin	Chen HAI	Ding Yizhen	Lu Yike	Ou Yangjing
Frequency	329	300	294	283	271

Table 2. Co-occurrence matrix of main characters in the novel (part)

	Li Dakang	Hou Liangping	Sha Ruijin	Gao Yuliang	Qi Tongwei
Li Dakang		68	103	102	59
Hou Liangping	68		51	149	135
Sha Ruijin	103	51		78	23
Gao Yuliang	102	149	78		110
Qi Tongwei	59	135	23	110	

knowledge graph and its representation, a knowledge graph with weights is formed and represented by introducing weights of people and relations.

(3) Visualization of knowledge graph. The graph database technology is used to visualize the knowledge atlas of people and their relationships.

(4) Knowledge map visualization with weights. The weighted people and their relational knowledge maps are visualized by Web technology.

Chinese word segmentation in the unit of period, identify the main character names in the novel, and extract the main character list.

(2) Based on the character list, the Program is written in Python language to count the frequency of each character appearing in the novel.

(3) Use Chinese word segmentation, read the text of the novel sentence by sentence for identification, record the name and relationship of characters, combine synonyms, and sort out the relationship of characters.

(4) Use Python programming to analyze characters' co-occurrence matrix.

Among them, the results of the frequency list of characters are shown in Table 1 (partial data), and the co-occurrence matrix of characters is shown in Table 2 (partial data).

3.2 Construction of Knowledge Graph of Personage Relationship and Knowledge Graph with Weight

3.2.1 Construction of Knowledge Graph of Human Relationship

Knowledge graph, also called scientific knowledge graph, is a relatively general formal description framework of semantic knowledge. Triplet is the basic unit that constitutes the knowledge graph and is expressed as "entity-relationship-entity". Among them, entities are nodes and relations are edges, which constitute the network structure of the knowledge graph. For example, entity 1 "Gao Yuliang", entity 2 "Li Dakang" and relation "opponent" constitute the triplet of "Gao Yuliang - opponent - Li Dakang" in the knowledge graph of human relationship.

Through the above overview, it can be concluded that character relationship is the generalization of the relationship between people and people in society. Different observation angles will also get different generalization names. The relationship between characters is extracted from the text to construct the triplet of character relation, and the knowledge map of character relation can be obtained.

Therefore, the triplet of human relations can be defined as: Let the set of human relations be $R=\{r_1,r_2,r_3,\dots,r_k\}$, there exists a person's name pair $P=(p_i,p_j)$ that satisfies the relation $r_k \in R$, then the triplet PR can be described as: $PR=\{p_i,p_j,r_k\}$, where p_i and p_j are the person's name entities, and r_k is the relation between them. The relationship files between the main characters in The Name of the People obtained through Python programming are shown in Figure 2.

```
"Gao Yuliang", "Xiao Gangyu", "subordinate"
"Sha Ruijin", "Wu Chunlin", "subordinate"
"Li Dakang", "Wang DaLu", "former assistant"
"Wang DaLu", "Ou Yangjing", "classmate"
"Wang Wenge", "Zheng Xipo", "master worker"
"Sha Ruijin", "Bai ChuZhang", "secretary"
"Li Dakang", "Zhang Shuli", "subordinate"
"Qin Juzhang", "Hou Liangping", "subordinate"
"Zhang Shuli", "Chen Qingquan", "classmate"
"Cai Chenggong", "Chang Xiaohu", "opponent"
"Gao Xiaoqin", "Liu Qingzhu", "subordinate"
```

Fig 2. Relationship between people's nominal figures

```
"Li Dakang",780
"Hou Liangping",1355
"Sha Ruijin",353
"Gao Yuliang",627
"Qi Tongwei",553
"Lu Yike",283
"Zhong Xiaoi",49
"Gao Xiaoqin",329
"Wu Caixia",57
"Ou Yangjing",271
"Zhao Ruilong",151
"Zheng Xipo",151
"Zhao Donglai",154
"Liang Lu",141
"Ji Changming",34
"Lao Cheng",252
```

Fig 3. Weighted Characters in The Name of the People

```
"Sha Ruijin", "Li Dakang",103,"subordinate"
"Li Dakang", "Gao Yuliang",102,"opponent"
"Zhao Lichun", "Li Dakang",27,"former secretary"
"Hou Liangping", "Gao Yuliang",149,"teacher"
"Ji Changming", "Hou Liangping",107,"subordinate"
"Sha Ruijin", "Gao Yuliang",78,"subordinate"
"Zhao Lichun", "Gao Yuliang",22,"former subordinate"
"Sha Ruijin", "Qi Tongwei",23,"subordinate"
"Qi Tongwei", "Gao Yuliang",110,"teacher"
"Chen Hai", "Qi Tongwei",41,"upperclassman"
"Chen Hai", "Lu Yike",21,"subordinate"
"Zhong Xiaoi", "Hou Liangping",27,"husband"
"Wu Huifen", "Gao Yuliang",29,"husband"
"Gao Xiaoqin", "Gao Yuliang",28,"brother in law"
"Qi Tongwei", "Gao Xiaoqin",65,"lover"
"Zhao Ruilong", "Gao Xiaoqin",25,"commercial intercourse"
"Wu Caixia", "Liu Qingzhu",19,"husband"
"Ou Yangjing", "Li Dakang",83,"husband"
"Zhao Ruilong", "Zhao Lichun",20,"father"
"Duo Bozhong", "Zhao Ruilong",11,"classmate"
```

Fig 4. Character Relationships with Weights

3.2.2 Construction of Knowledge Graph with Weight of Personage Relationship

The frequency of the appearance of characters represents the importance of characters in the work, and the co-occurrence frequency of characters represents the closeness of the connection of characters. The two are introduced into the knowledge graph in the form of weight to expand it. Let the set of human relations be $R=\{r_1,r_2,r_3,\dots,r_k\}$, relation r_i There exists a person's name pair $P=(p_i,p_j)$ that satisfies relation $r_k \in R$, then the quad PR with weight can be described as: $PR=\{p_i,p_j,w_{ij},r_k\}$, where p_i and p_j are named entity, w_{ij} is the weight of the relationship between them, and each named entity binary P with weight is described as $P=\{p_i,w_i\}$, where w_i is the weight of named entity p_i . For example, the duality of personage weight is {Li Dakang, 780}, {Hou Liangping, 1355}, and the quaternion of personage relation with weight is {Sha Ruijin, Li Dakang, subordinate, 103}, {Hou Liangping, Gao Yuliang, 149, teacher}. Figures 3 and 4 are weighted people files and weighted people relationship files obtained through Python programming.

3.3 Use Neo4j to Store and Visualize the Knowledge Map of Human Relationship

To construct the knowledge graph of character relationship, nodes are used to express characters, and edges are used to express the relationship between characters. Graph database is generally used to store knowledge graph. Using Neo4j as a tool, the characters and their relationships in The Name of The People are stored and visualized. The steps are as follows:

- (1) Load the Py2NE library in Python (a third-party library for Python to interact with Neo4j).
- (2) Use Python programming to read the character list stored in txt file format, and create character node in Neo4j.

(3) Use Python programming to read the list of character relationships stored in TXT file format, match character nodes and create relationships in Neo4j.

(4) Open the Web interface of Neo4j and view the visualization results, as shown in Figure 5.

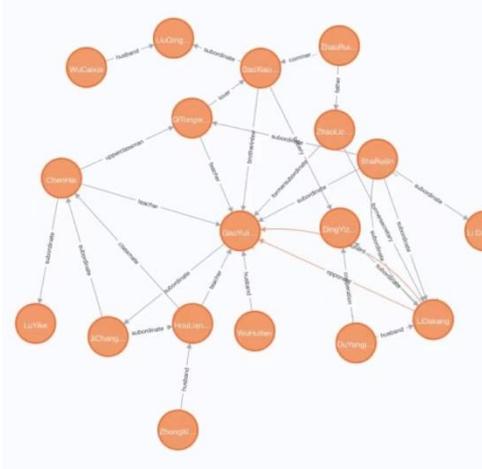


Fig 5. The Knowledge Map of Neo4j for the Relationship

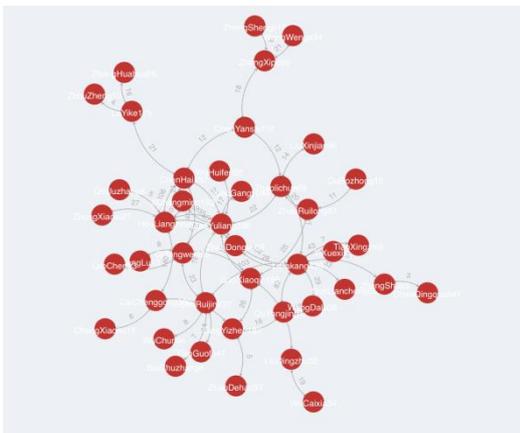
```
{ "name": "Ding Yizhen" , "value": 183 },
{ "name": "Chen Hai" , "value": 182 },
{ "name": "Zhao Dehan" , "value": 37 },
{ "name": "Zhang Huahua" , "value": 26 },
{ "name": "Zhou Zheng" , "value": 10 },
{ "name": "Chen Yanshi" , "value": 116 },
{ "name": "Cai Chenggong" "value": 260 },
```

Fig 6. Data array in json file

3.4 Use Web programming to visualize the weighted knowledge map of people and relationships

```
{"source": "Gao Yuliang" , "target": "Ji Changming" , "value": 35 , "label": "subordinate"},
{"source": "Ji Changming" , "target": "Chen Hai" , "value": 21 , "label": "subordinate"},
{"source": "Qi Tongwei" , "target": "Lao Cheng" , "value": 1 , "label": "attendant"},
{"source": "Li Dakang" , "target": "Ding Yizhen" , "value": 56 , "label": "subordinate"},
{"source": "Gao Xiaoqin" , "target": "Ding Yizhen" , "value": 26 , "label": "bribery"},
{"source": "Ou Yangjing" , "target": "Ding Yizhen" , "value": 16 , "label": "cooperation"},
{"source": "Cai Chenggong" , "target": "Ding Yizhen" , "value": 36 , "label": "bribery"},
{"source": "Hou Liangping" , "target": "Chen Hai" , "value": 106 , "label": "classmate"},
{"source": "Chen Hai" , "target": "Gao Yuliang" , "value": 39 , "label": "teacher"},
{"source": "Chen Hai" , "target": "Chen Yanshi" , "value": 12 , "label": "father"},
```

Fig 7. Links array in json file



Knowledge graphs with weights need not only to create character nodes and relationships, but also to represent the weights of characters and relationships, and to set different display effects according to different weights during visualization. Therefore, the Web programming technology of JavaScript is used to visualize the weighted human relation knowledge graph. The steps are as follows:

(1) Use Python programming to read the list of characters and their weights and the Excel file that stores the co-occurrence matrix of character relations, and build JSON files, store characters and their weights in the data array of the file, and store character relations and their weights in the links array of the file.

(2) Use Javascript programming to build the Echart chart of Web, whose type is Graph, and read the constructed JSON file.

(3) Map the figures in the data array to the nodes in Graph, and show how often each character appears. The screenshots of the JSON file is shown in Figure 6.

(4) Map the character relation in the links array to the edge between the corresponding character nodes in Graph, the size of the numbers on the sides reflects the closeness of the characters, with the closer the relationship, the larger the number. The screenshots of JSON file is shown in Figure 7.

(5) Use a Web browser to display visual results, as shown in Figure 8.

Fig 8. Web visualization

4. ANALYSIS AND DISCUSSION

According to the constructed weighted knowledge map of character relationship and the visualization results, it can be concluded that "Hou Liangping" is the main character of the novel, that is, the protagonist of the novel. Other character nodes spread around based on him, and the surrounding characters such as "Li Dakang", "Gao Yuliang", "Qi Tongwei" and "Sha Ruijin" have obvious relations with him and are closely related. It can be concluded that these main characters form the main frame of the novel, and then

these characters spread around to form the network of relationships of the full novel, which is also consistent with the description of the novel in the Name of the People. Including Hou Liang gentle good higher education for teachers and students relations, Hou Liang peaceful qi wei to classmates, Li Dakang good and high education to in-fighting between rival relationship closely linked, such as the closely relationship in the knowledge map line is relatively coarse, the rest of the characters between, such as bribery, husband and wife, interest relationship between the superior and the subordinate can express through knowledge map can be more intuitive, And can directly see how closely the characters are connected.

Through the comparison of the two knowledge graphs, it can be seen that the knowledge graph with weight can more intuitively show the importance of characters in the novel and whether the relationship between characters is close, so that readers and audiences can more accurately locate characters and their core network, so as to understand the plot of the work more quickly and accurately.

REFERENCES

- [1] Yang, X., Song, Z. (2019) "Research on Character Relationship in Novels Based on Co-word Analysis", School of Information, Guizhou University of Finance and Economics
- [2] Chen, L., Hu, Y., Ai, W., Hu, J. (2015) "Social Power Relationship extraction and Network Construction in A Dream of Red Mansions", *Journal of Chinese information science*, 29(05), pp.185-193 +203.
- [3] Xiao, T., Liu, Y., (2015) "An Analysis of Jin Yong's and Gu Long's Novel Styles based on Clustering and Classification", *Journal of Chinese information science*, 29(05), pp.167-177.
- [4] Chen, Z.(2020) "Research on Multidimensional Human Relation Construction and Visualization Method", Zhejiang University of Technology.
- [5] Si, D., (2018) "Research on Character Relationship Analysis and Visualization", xi'an University of Science and Technology.
- [6] Xiang, F., Zhu, S., (2018) "Experimental Report on the Drawing of Chinese family tree -- Based on the Database of Biographies of Chinese Historical Figures in Visual Practice Decoration", 306(10), pp.92-95.
- [7] Feng, Y.(2016) "Design and Implementation of Human Relation Construction Based on Knowledge Graph", Chongqing University.
- [8] Chen, Q., (2019) "Research on the Construction of Knowledge Map of Microblog Topic people Relationship", Anhui University of Science and Technology.
- [9] Drakopoulos, G., Baroutiadi, A. (2019) "Megalooikonomou V. Higher order graph centrality measures for Neo4j", *International Conference on Information*, 68(201), pp.6-7.
- [10] 10. Drakopoulos, G., Kanavos, A., Mylonas, P. (2017) "Defining and evaluating Twitter influence metrics: a higher-order approach in Neo4j", *Social Network Analysis & Mining*, 7(1), pp.52-53.
- [11] 11. Karidi, D. (2020) "From user graph to Topics Graph: Towards twitter followee recommendation based on knowledge graphs", *IEEE International Conference on Data Engineering Workshops*, Helsinki, Finland, 16(20201). 163-164.
- [12] 12. Li, J. (2019) "Construction process of human relation graph based on co-occurrence and association mining", *Chinese academy of national information technology*, 01(01), 110-112.
- [13] 13. Wang, Y., Yu, J., Zhao, C. (2017) "Research on the Relationship between Chinese characters and Chinese Characters", *Information Exploration*, 2017(07), pp.52-56.