# Structure and Evolution Patterns of Contents of Chinese Children's Bestsellers

Jiming Hu[1,2][0000-0001-6491-1995] and Xiang Zheng[1,2][0000-0001-7933-2822]

[1] School of Information Management, Wuhan University, Wuhan, China
[2] Center for the Study of Information Resources, Wuhan University, Wuhan, China
hujiming@whu.edu.cn

**Abstract.** Children's books involve a large number of topics. This poster focuses on that in China, which is the fastest growing market for children's book in the world. This poster chose Dangdang.com, the biggest Chinese online bookstore, for data source to obtain children's bestsellers. The topic words of children's bestsellers were extracted from their brief introductions of the content on the website. With the aid of co-occurrence theory and tools of social network analysis and visualization, the distribution, correlation structures, and evolution patterns of topics were revealed and visualized. This poster shows that topics of Chinese children's bestsellers are broad and relatively concentrated, but their distribution is unbalanced. There are four distinguished topic communities (Living, Animal, World, and Child) in terms of centrality and maturity, and they all establish their individual systems and tend to be mature. The evolution of these communities tends to be stable with powerful continuity.

**Keywords:** Chinese Children's Books, Topics Distribution, Topic Correlation Structures, Evolution Trends, Visualization.

## 1 Introduction

Research on content of children's books has always been the focus. Text analysis based on works could help uncover the explicit and implicit connotation of children's books [1-3]. Achievements in terms of topics or themes have been gained. Previous studies have found that children's books with different topics could help to raise their awareness of the world, family and individual in different ways [4, 5].

With the development of Chinese children's book market, scholars begin to explore the content characteristics of children's bestsellers [6]. The findings indicate that topics of Chinese children's bestsellers are becoming increasingly diversified [7, 8]. However, previous studies were mostly qualitative analysis on the small scale. Quantitative analysis and semantic correlation analysis on topics of large-scale children's books have not yet been explored, as well as the longitudinal study of topics in Children's books is lacked.

To understand and explore the characteristics of Chinese children's books and even the Chinese children's culture, we attempt to utilize a large-scale data of Chinese chil-

dren's bestsellers to reveal the intellectual structure, evolution patterns, and development trends. We try to identify the key topics in Chinese children's books and reveal their implicit relationship by co-word analysis. The findings display the intellectual structure of topic words and the changing of topics over time with the aid of visualization tools. This will enable us to understand the development of contents in Chinese children's books, and even the reading preference of children and parents in China.

## 2 Methodology

### 2.1 Data Collection and Processing

This study selected the Top 500 children's books listed in the annual ranking from 2014 to 2018 in Dangdang.com, which is Chinese largest online book retailing platform in China [9]. After removing the repetition from the series, the brief introduction of each book was collected. Then, topic words in the introduction were manually extracted. In total, 11,766 words were selected for analysis.

In our study, the threshold frequency of word was set as 40. In addition, synonyms were replaced by the general terms. Secondly, words with frequency less than 40 were merged into the corresponding hypernyms with frequency greater than 40. Finally, 307 words were selected for the subsequent co-word analysis.

### 2.2 Co-word Network Analysis

This study is conducted based on the theory of co-word analysis, which allows us to mine the deeper meaning of content [10] and has been successfully used in many research fields [11-13].

Firstly, the overall characteristics of Chinese children's bestsellers co-word network were analyzed. A co-word network file was generated by SCI2 [14, 15]. The co-word network comprised of selected 307 words is exactly the largest connected component. Secondly, the co-word data file was imported into Pajek [16] to calculate network indicators, such as density, centralization or centrality [17], and clustering coefficient [18]. Meanwhile, based on the Louvain algorithm [19], topic communities representing the mainstream directions of Chinese children's bestsellers were detected bedded in Pajek. Finally, the correlation structure of co-word network that distinguishes each community was drawn by VOSviewer [20]. According to correlation structure of co-word network in each year, a longitudinal evolution venation was drawn in Cortext platform [21].

## 3 Results and Findings

### 3.1 Distribution of Topics

Different titles in the same series were counted as one book or series in this paper. As shown in Fig. 1., the continuous growth indicates that the number of single titles are

more than series books in the past five years. This trend also equals that Chinese readers have their preference on more diversified topics.
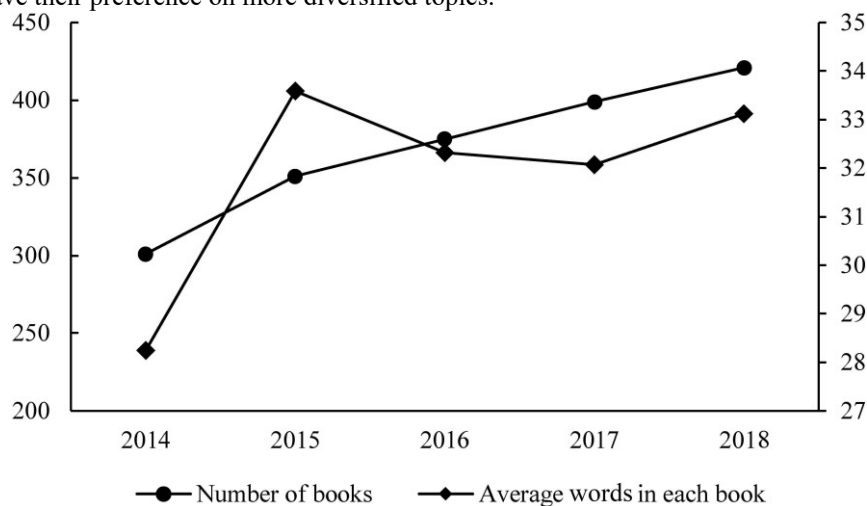


**Fig. 1.** Changing of Chinese children's bestseller titles and topic words (2014-2018)

It also indicates a serious imbalance in children's bestsellers' topics. Table 1 shows the top 20 words with frequency. Books involving the top ten topic words are more preferred because that they are related to children's life and growth. All the topics mentioned above reflect the characteristics of children's physical and mental development, ideological understanding, and reading aesthetics.

**Table 1.** Top 20 topic words in Chinese children's bestsellers

| No. | Word | Occurrence frequency | No. | Word | Occurrence frequency |
|-----|------|----------------------|-----|------|----------------------|
| 1 | Living | 447 | 11 | Dad | 266 |
| 2 | Child | 420 | 12 | Picture | 259 |
| 3 | Animal | 382 | 13 | Friend | 258 |
| 4 | Game | 331 | 14 | Picture Book | 246 |
| 5 | Mom | 315 | 15 | Knowledge | 243 |
| 6 | Growing up | 295 | 16 | Love | 242 |
| 7 | Nature | 290 | 17 | Science | 237 |
| 8 | World | 283 | 18 | Sea | 224 |
| 9 | Read | 275 | 19 | Biological Phenomena | 215 |
| 10 | Human Being | 274 | 20 | Ability | 212 |

## 3.2 Correlation Network Analysis of Topic Words

**Network Characteristics.** The indicators of the whole network shown in Table 2 reveal the structural characteristics of correlation formed by topic words. Low density shows the correlation between topic words on the whole is not very close, and the overall development of children's bestsellers is not mature in terms of content. The high degree centralization and closeness centralization indicate that the structural cohesion is very high, and topic words are more inclined to be correlated centering on a few topic words. The low betweenness centralization indicates that most topic words correlate with others not via a third one. The higher network clustering coefficient shows a great potential that topic words are aggregated into distinguished clusters centering on the few ones.

It can be said that there are four distinguished topic communities or directions of Chinese children's bestsellers.

**Table 2.** Indicators of the whole network

| Indicators | Value |
| --- | --- |
| Number of nodes | 307 |
| Number of lines | 11699 |
| Average degree | 76.21 |
| Density | 0.2491 |
| Network all degree centralization | 0.7197 |
| Network all closeness centralization | 0.7750 |
| Network all betweenness centralization | 0.0457 |
| Network clustering coefficient | 0.4894 |
| Communities | 4 |

**Table 3.** The centrality of top 10 topic words

| Ranking | Words | Degree | Words | Closeness | Words | Betweenness |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Living | 295 | Living | 0.9653 | Living | 0.0480 |
| 2 | Child | 286 | Child | 0.9387 | Child | 0.0417 |
| 3 | Animal | 278 | Animal | 0.9162 | Animal | 0.0386 |
| 4 | Mom | 264 | Mom | 0.8793 | Game | 0.0318 |
| 5 | Game | 259 | Game | 0.8669 | Mom | 0.0313 |
| 6 | Dad | 256 | Dad | 0.8596 | Human Being | 0.0277 |
| 7 | Human Being | 253 | Human Being | 0.8524 | Dad | 0.0260 |
| 8 | Nature | 250 | Nature | 0.8453 | Nature | 0.0256 |
| 9 | World | 242 | World | 0.8270 | World | 0.0253 |
| 10 | Friend | 240 | Friend | 0.8226 | Picture | 0.0226 |

Table 3 shows the top ten topic words in terms of centrality. Topic words such as "Living," "Child," "Animal," and "Mom", etc. have both higher degree centrality and closeness centrality, indicating that they are the core and the important topic words. In contrast, betweenness centrality of all topic words is very low. This again indicates that their role as a bridge to link other topic words is very weak.

In sum, it can be concluded that there are a few hot topics in Chinese children's bestsellers. Books related to these topics can be seen as the most favorite ones for Chinese market and readers.

**Topic Community.** Current Chinese children's bestsellers focus on four topic communities, as shown in Table 4. In terms of scale, there are three larger topic communities (C1-Living, C2-Animal and C4-Child) and relatively smaller one (C3-World). Each community has remarkable differences in the topic, but it is consistent in the direction, which represents the mainstream of current Chinese children's bestsellers.

**Table 4.** Four topic communities of Chinese children's bestseller content

| Community | Words |
| --- | --- |
| C1 | Living; Game; Nature; Read; Picture; Knowledge; Science; Ability; Learning; Hue; History; Art; Beauty; Cognition; Plant; Cogitation; Maths; Pleasure; Curiosity; Imagination; ect. |
| C2 | Animal; Friend; Sea; Difficulty; Mouse; Home; Cat; House; Forest; Water; Travel; Fish; Food; Bird; Rabbit; Flower; Bear; Insect; Puppy; Eye; ect. |
| C3 | World; Human Being; Wisdom; Take a Risk; Country; Girl; Secret; Wizardry; Fairy Tale; Juvenile; Myths and Legends; China; ect. |
| C4 | Child; Mom; Growing Up; Dad; Picture Book; Love; Biological Phenomena; Psychology; Affection; Kind; Paterfamilias; Brave; Boy; ect. |

**Visualization of Network.** Fig. 2. and Fig. 3. illustrate the correlations in each topic community and among four topic communities.

As shown in Fig. 2., the degree of internal correlation in each community is distinguished obviously. It also indicates the significant difference in directions of Chinese children's bestsellers. Fig. 3. shows the correlation structure formed by four topic communities. As the three largest communities, C1-Living, C2-Animal, and C4-Child are closely correlated with each other, especially the latter two. This pattern reveals that topics related to these three communities are the most popular in China.

**Fig. 2.** Four topic communities of Chinese children's bestseller content (Top 500 links)



**Fig. 3.** Correlation structure formed by topic communities

### 4.3 Topic Evolution and Development Trends

**Evolution Patterns of Topics.** As shown in Fig. 4., continuity of the topics evolution is pretty good with many obvious venations and few isolated communities or intermittent venations. It indicates that topics in Chinese children's bestsellers are stable and have not changed a lot. There are no brand-new topics that hit the market.
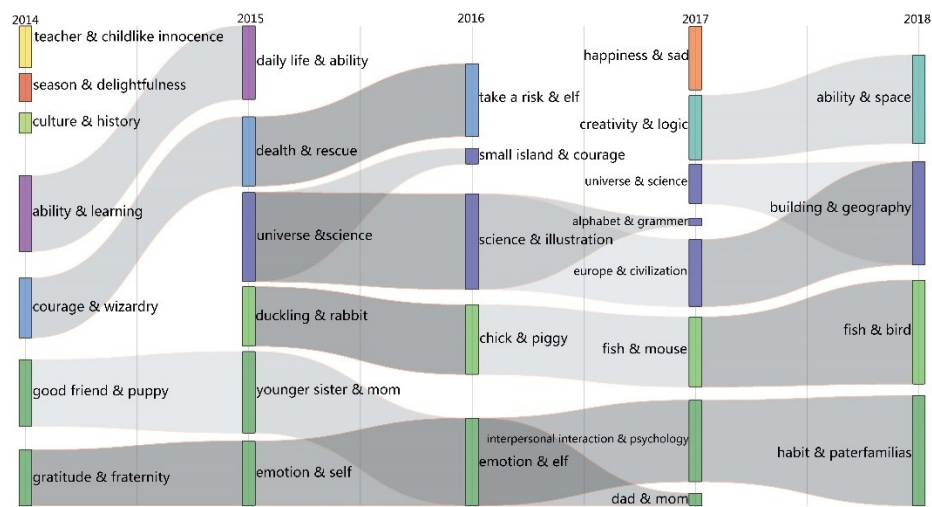


**Fig. 4.** Topics evolution of Chinese children's bestseller (2014-2018)

In general, the topics of Chinese children's bestsellers has good continuity. The three continuous evolution venations indicate the mainstream contents of Chinese children's bestsellers, which are "*gratitude & fraternity*," "*duckling & rabbit*," and "*universe & science*."

Meanwhile, topic venations of intermittent evolution or isolated status also exist. For example, topic venation of "*ability & learning*" stopped in 2015, indicating that Chinese parents has begun to realize that comprehensive abilities are more important. Since 2017, children's books, involving brain development, emotion management, art appreciation, and logical thinking training, have been increasingly popular.

**Development Trends.** As shown in Fig. 5., the contrast among four topic communities is very obvious in terms of degree centrality and density. In general, the density of each topic community is much higher than that of the whole network, indicating that the correlation between topic communities is not high. C2-Animal is located in Quadrant I with high degree centrality and density. Related topics are the core of Chinese children's bestsellers, and its development is relatively more mature. C3-World is located in Quadrant II with low centrality but high density. Although this topic community is mature, related topics are not the core of Chinese children's bestsellers. Similarly, C1-Living and C4-Child are located in Quadrant III with the relatively both low centrality

and density. Although their scales are relatively large, they are still in the exploration stage in Chinese market.
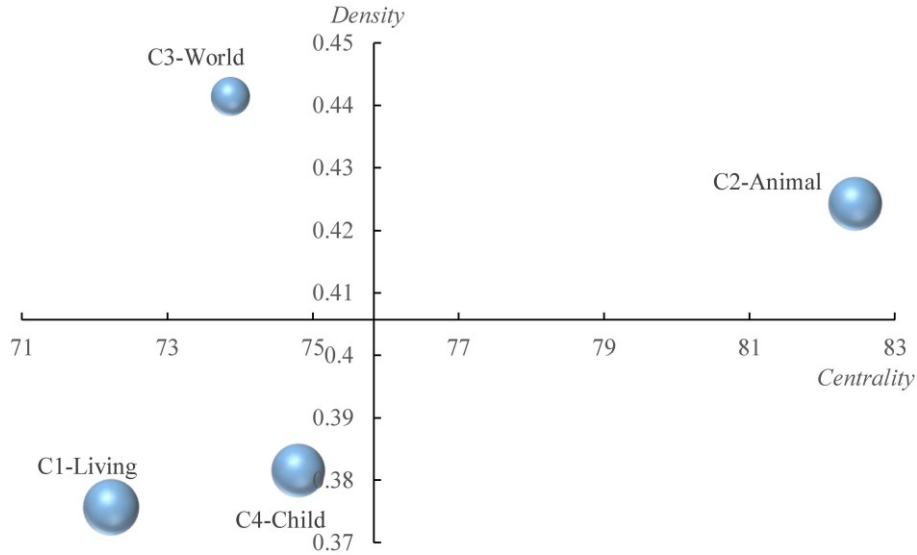


**Fig. 5.** Development trends of four topic communities of Chinese children's bestseller

## 4    Conclusions

This paper reveals the topic distribution, correlation structure, evolution patterns, and development trends of Chinese children's bestsellers through co-word analysis.

From 2014 to 2018, there are a wide range of topics in Chinese children's bestsellers. The distribution of topics is unbalanced, but the aggregation of them is strong. Content of books can be divided into four distinct topic communities. Besides C4-World community, other three communities are highly correlated. C2-Animal community is the core and the most mature direction in Chinese children's books. According to the calculation of network indicators, the relative development trends of these four communities are different. The evolution venations of topics are clear with a good continuity, indicating that the mainstream topics of Chinese children's bestsellers didn't change much. It is different in annual topics' distribution, but evolution patterns of topics related to similar content are stable and continuous over the years. Results above also reflect some considerable problems, including fierce market competition, the Matthew Effect, insufficient innovation in topics with rigid thinking, and lack of originality.

# References

1. Braden, E. G., Rodriguez, S. C.: Beyond mirrors and windows: A critical content analysis of Latinx children's books. Journal of Language and Literacy Education 12(2), 56-83 (2016).
2. Yilmaz Genc, M. M., Akinci Cosgun, A., Pala, S.: A Study of mathematical content provided in illustrated children's books. Eurasian Journal of Educational Research 69, 159-175 （2017).
3. Nor, M. R. M., Abidin, Z., Nasaruddin, N. B.: Features of Islamic children's books in English: A case study of books published in Malaysia. Publishing Research Quarterly 34(3), 540-553 (2018).
4. Mijung, K.: Rethinking K-6 scientific literacy: Study of using science books as tool to a fundamental sense of scientific literacy. Journal of the Korean Association for in Science Education 27(8), 711-723 (2007).
5. Hsiao, C. Y.: Enhancing children's artistic and creative thinking and drawing performance through appreciating picture books. International Journal of Art & Design Education 29(2), 143-152 (2010).
6. Zhang, J. C.: Research on the Restrictive Factors and Solutions of the Localization of Children's Introduction Books. master's thesis, Beijing Institute of Graphic Communication (2017).
7. Shan, D. P., Zhang, W. H.: The regular analysis of children's best seller in our country-based on the data of the best-selling books for children in the past two years. View on Publishing. 14, 21-25 (2018).
8. Meng, H. W.: Application and development of VR technology in children's book publishing. Propagation and copyright. (in China) 6: 73-74,77 (2018).
9. Chen, L.: Dangdang sells 1.5 billion copies in 2018 (in Chinese language) (2019). http://www.sohu.com/a/287858182_267807, last accessed 2019/11/22.
10. An, X. Y., Wu, Q. Q.: Co-word analysis of the trends in stem cells field based on subject heading weighting. Scientometrics. 88 (1), 133-144 (2011).
11. Leydesdorff, L., Zhou, P.: Co-word analysis using the Chinese character set. Journal of The American Society for Information Science and Technology. 59 (9), 1528-1530 (2008).
12. Hu, J., Zhang, Y.: Research patterns and trends of Recommendation System in China using co-word analysis. Information Processing & Management. 51(4), 329-339 (2015).
13. Khasseh, A. A., Soheili, F., Moghaddam, H. S., Chelak, A. M.: Intellectual structure of knowledge in iMetrics: A co-word analysis. Information Processing & Management. 53(3), 705-720 (2017).
14. Boerner, K.: Plug-and-play macroscopes. Communications of the ACM. 54(3): 60-69. (2011).
15. Hu, C., Hu, J., Deng, S., Liu, Y.: A co-word analysis of library and information science in China. Scientometrics. 97(2), 369-382 (2013).
16. Mrvar, A., Batagelj, V.: Analysis and visualization of large networks with program package Pajek. Complex Adaptive Systems Modeling. 4 (1), 6 (2016).
17. Freeman, L. C.: Centrality in social networks conceptual clarification. Social Networks. 1(3): 215-239 (1979).
18. Doreian, P., Lloyd, P., and Mrvar, A.: Partitioning large signed two-mode networks: Problems and prospects. Social Networks 35(2):178-203 (2013）.
19. Blondel, V. D., Guillaume, J. L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 8 (10),10008 (2008).

20. Van Eck, N.J., Waltman, L.: Software survey: VOSviewer, a computer program for biblio-metric mapping. Scientometrics 84 (2): 523-538 (2010).
21. Leydesdorff, L., Goldstone, R. L.: Interdisciplinarity at the journal and specialty level: The changing knowledge bases of the journal Cognitive Science. Journal of the Association for Information Science and Technology 65 (1), 164-177 (2014).