

Research on Knowledge Structure of Physics Textbook

Xue-mei Cui¹, Seung Kee Han² and Shou Zhang^{3*}

¹Normal College Yanbian University, Yanji, China

²Department of Physics Chungbuk National University, Cheongju, Korea

³Department of Physics Yanbian University, Yanji, China

xmcui@ybu.edu.cn, skhan@chungbuk.ac.kr, *szhang@ybu.edu.cn

Abstract - A Physics textbook is composed of numerous physics terminologies which are organized to form new concepts of physics. For the purpose of understanding the global organization of the science-knowledge system in science textbooks, we analyzed the binary network of science textbooks, where the pair appearance of two science terminologies is taken as the link between two nodes. By analyzing the characteristics of binary network of the textbook, we can be understood the structure characteristics of physics knowledge in the textbook. We found that a physics knowledge network possesses the characteristics of complex network: 1) short mean distance between terminologies; 2) power law degree distribution; 3) hierarchical modular structure.

Index Terms - physics textbook, physics terminology, physics knowledge network, power law

1. Introduction

Since the invention of written language and printing, as the recording and transmitting carrier of information, books have been playing important role in all fields. With the rapid development of computer science and network technology, with various forms and more systematic construction mode, books have become to the learning media which are easier to be understood by readers.

After the long time accumulation of experience of human, the writing mode has been fixed gradually. To describe the specific contents of a book, the author selects the words purposefully and intentionally, and furthermore, arranges them systematically to promote the knowledge from simple level to higher level. The words used in book are selected by author, but, to be easy to understood, this selection is limited by the contents and restricted by the words frequently used by forefathers or the words allowed. The technical terms of certain trade or profession are called terminologies which are the arrangements to express the same knowledge. Describing a phenomenon is hard to be achieved with a single term, but needs several relevant terms and must use some special terms at high frequency. The closest terms are appeared in the same sentence to express the simple phenomenon, and furthermore, sentences compose a paragraph, paragraphs compose a section, sections compose a unit, and finally, units compose a book forming the largest unit of knowledge. From this point of view, the terminologies used in a book are related together and they jointly create the new knowledge. It can be said that the knowledge is composed of the terminologies needed and the relation between these terminologies. Hence, the connection graph, which is also known as network, can be established for physics terminologies based on the theory of complex network.

Commonly, the network is constructed by linking the relevant elements, and the more the elements exist, the more complex the network is. To explain the complexity of natural or social phenomena, many researches were carried out recently [1-5], and the development of computer science and network technology makes the research of complexity be possible in the fields of social network, biological network, Internet etc.

In this paper, it is focused that the knowledge in a book is composed of the relations between physical terminology, and the network of physics knowledge was established for three textbooks of general physics, and furthermore, the characteristics of the network were analyzed. In our work, a node of knowledge network is a physical terminology, and the links of the network describes the relations between physical terminology.

2. Research Materials and Methods

The research subjects of this paper are three textbooks of general physics written by Knight[6], Griffith[7] and Hewitt[8] respectively, and the definitions of physical terminology are determined according to Oxford Dictionary of Physics[9]. When extracting physical terminology, the singular and plural forms of a terminology are regarded as the same term, and multi-word terms are extracted as compounds. For example, the term of magnetic force is only extracted as compound term instead of extracting force as another term. For polysemantic terms, such as second, the semantic parts of non-physics are cancelled according to artificial verifying after extracting.

The appearance number of a term in whole book is known as frequency of the term denoted as f . The basic data of the physical terminology used in above three textbooks are listed in Tab.1. The system sizes of these three textbooks are different, which is conducive to discover the system-size-free intrinsic characteristics of the physics knowledge.

TABLE 1 Basic Statistic Data of Three Textbooks

Author	Title	Number of sentences	Number of physical terminology	Accumulated number of physical terminology
Randall D. Knight	Physics	31488	815	34743
W. Thomas Griffith	The Physics of Everyday Phenomena	19213	685	22286
Paul G. Hewitt	Conceptual Physics	13115	611	15421

The physics knowledge of a physics textbook is consist of sentences constructed by the most basic physical terminology, and several sentences compose paragraphs, sections, and so on, to illustrate natural phenomena. Therefore, it can be considered that there exit relations among the terms appeared in the same paragraph, and the relations among the terms in a sentence are closer. Based on this point of view, the knowledge network can be established according to following assumptions:

①A physical terminology appeared in textbook is represented as a node of network.

②If two physical terminology are appeared in the same sentence, then these two terms are regarded as linked term without weight.

③A knowledge network is an undirected graph.

The knowledge network established according this way is an unweighted and undirected binary network.

3. Results of Research

A. Statistical Characteristics of physical terminology

According to different emphases of textbooks, there exist different physical terminology in various textbooks. But it was found in research work that the term appeared at maximum frequency in the three textbooks simultaneously is force, and moreover, the terms such as energy, light, time, mass, charge, motion, speed, wave, current, atom have relatively high frequency in the three textbooks. It shows that these physical terminology play relatively important roles in physics knowledge system.

Frequency distribution of physical terminology is shown in Fig. 1 to understand the frequency distribution of physical terminology in the textbooks. It may be seen that small number of terms are with high frequency, and the frequencies of most of terms are between 1~10 although there exist a few terms which frequencies are over 1000. The distribution follows the typical power law of $P(f) \sim f^{-\alpha}$, and the exponents α of three textbook are all about 1.0, which shows that these textbooks possess similar properties in frequency distribution of physical terminology.

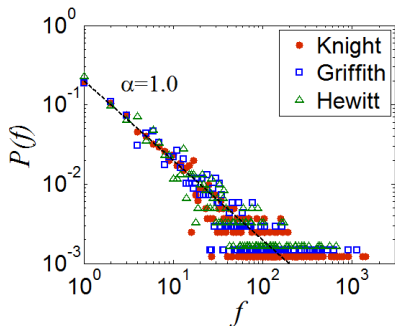


Fig. 1 Distribution of term frequency f for textbooks by Knight, Griffith and Hewitt (log-log plot)

B. Characteristics of Physics Knowledge Network

Fig.2 is drawn using Pajek[10] to show the binary

network established for the textbook by Knight. Because of large number of physical terminology, the characteristics of network will not be showed clearly if whole terms are displayed in a figure simultaneously. Therefore, in Fig.2, only 132 terms which are appeared in the same sentence simultaneously and hold frequency values greater than or equal to 15 are drawn. It can be seen from Fig.2 that there exist the physical terminology with high frequency and large degree, and there exist more link edges among the nodes having the same color which forms modular structure. With each part of textbook as a unit, several modules are composed in the network, and the terms with relatively high frequency act as hub of module, which shows that the terms with high frequency play an important role in forming the network of physics knowledge. The terms as the hub of each part are *force*, *charge*, *current*, *energy*, *particle*, *atom*, *light*, *wave*, etc. respectively, and these terms can be regarded as the central terms representing each part. Therefore, the knowledge structure behind the textbook may be revealed with the form of knowledge network.

The statistic data representing the characteristics of binary network for physics knowledge are listed in Tab.2.

The node numbers of binary networks, denoted by N , for three textbooks are 772, 654 and 562 respectively, and because of largest size of system the textbook by Knight has the largest number of nodes compared with other two textbooks.

Mean degrees, denoted by $\langle k \rangle$, of three textbooks are 20.1, 17.6 and 15.9 respectively, which are relatively large values and show that there exist more link edges among nodes of these binary networks, and indicate that many corresponding physical terminology participate in illustrating a physics concept. Analyzing the property of degree distribution is a typical analysis method of network structure characteristics. The degree distribution functions of binary network for the three textbooks are drawn in Fig.3, which shows that degree for the textbooks are also follow the power law of $P(k) \sim k^{-\beta}$, and the exponents β of three textbook fall into the range of 0.8~1.1, i.e. conform to the power law degree distribution. This indicates that binary networks for the three textbooks possess the property of scale free network.

The distance between two nodes of network is defined as the number of link edge in the shortest path connecting these two nodes, then it can be seen from Tab.2 that mean distances, denoted by $\langle D \rangle$, between nodes for the three textbooks fall into the range 2.7-2.9, i.e. less than 3.0.

TABLE 2 Statistic Data of Binary Networks for Three Textbooks

Author	N	$\langle D \rangle$	$\langle k \rangle$	$\langle C \rangle$
Randall D. Knight	772	2.8	20.1	0.3
W. Thomas Griffith	654	2.7	17.6	0.2
Paul G. Hewitt	562	2.9	15.9	0.3

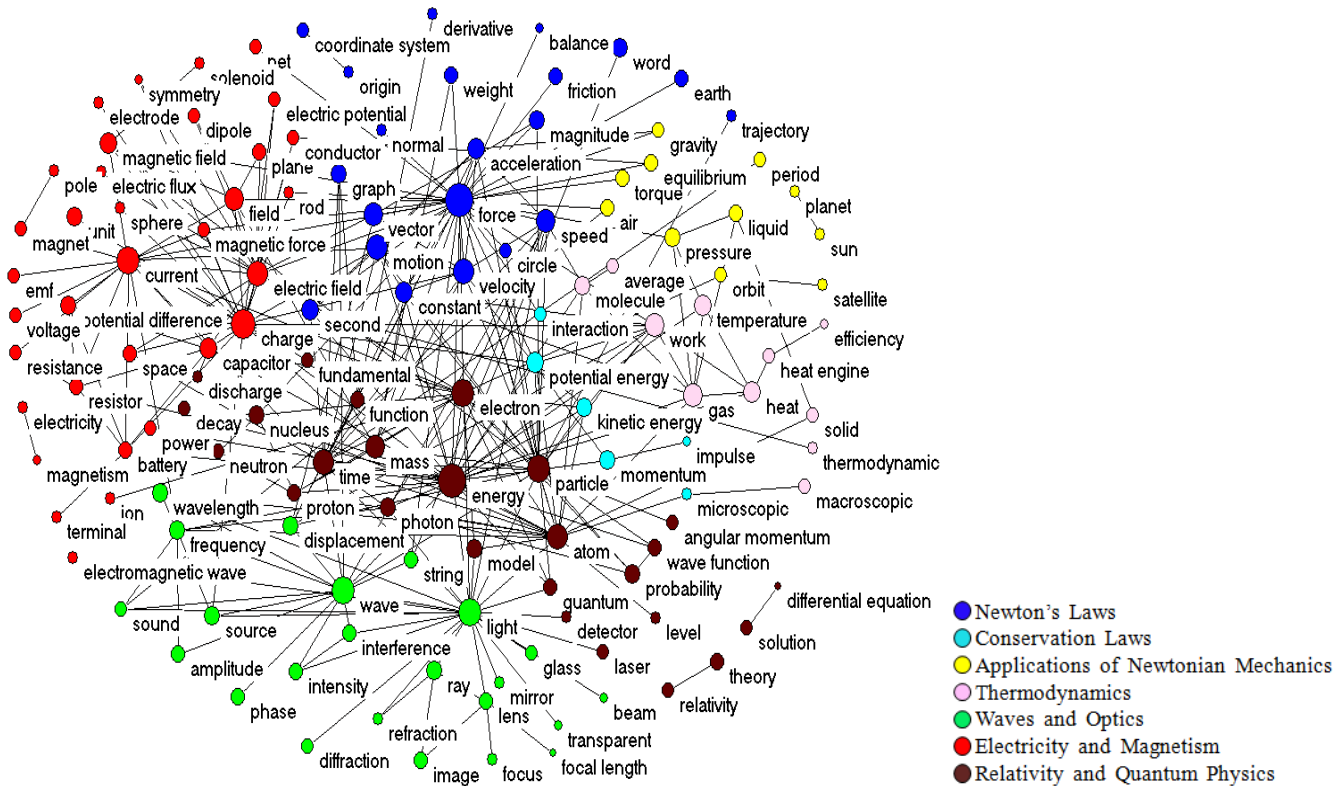


Fig.2 Binary network for the textbook by Knight. For clarity of network structure, only 132 terms appeared in the same sentence simultaneously and with frequency values greater than or equal to 15 are drawn. The size of node is proportional to the term frequency f , and the color represents the part the term belongs to. A term is belongs to the part in which the term has maximum frequency over all parts.

In graph theory, a clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together[11]. Evidence suggests that in most real-world networks, and in particular social networks, nodes tend to create tightly knit groups characterized by a relatively high density of ties; this likelihood tends to be greater than the average probability of a tie randomly established between two nodes [12-13]. Two versions of this measure exist: the global and the local. The global version was designed to give an overall indication of the clustering in the network, whereas the local gives an indication of the embeddedness of single nodes[11]. The average of the local clustering coefficients $\langle C \rangle$ of all the vertices n . Metabolic network ($\langle C \rangle = 0.7$)[2] and film actors network ($\langle C \rangle = 0.79$)[14], etc. have larger values of clustering coefficient which means that the nodes are compact in these networks. The values of $\langle C \rangle$ are in the range of 0.2~0.3 which indicates that the binary networks for the three textbooks also have certain degree of aggregation. The fact that mean distance $\langle D \rangle$ is less than 3.0 and clustering coefficient $\langle C \rangle$ is in 0.2~0.3 indicates that the binary networks for the three textbooks possess the properties of small world network.

The Clustering coefficient distribution diagram (see Fig.4) shows that there exist the power laws between clustering coefficient $C(k)$ and degree k , i.e. $C(k) \sim k^{-\gamma}$, and γ is in the

range of 0.4~0.5. This value is less than the value of metabolic network ($\gamma \approx 1$) but it is enough to illustrate that the binary networks for the three textbooks have hierarchical modular structure, i.e. several linking physical terminology compose small size concept modules and these modules compose the larger size concept modules, and with this hierarchical modular structure, the whole physics knowledge network can be constructed.

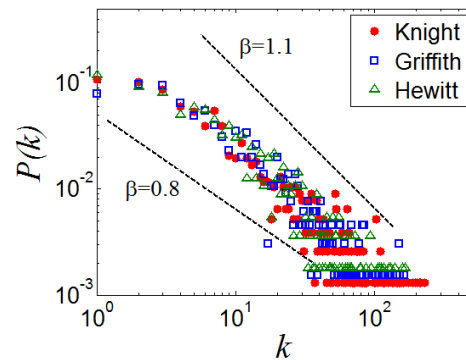


Fig. 3 Degree distribution diagram of knowledge networks for three textbooks (in dual log coordinate)

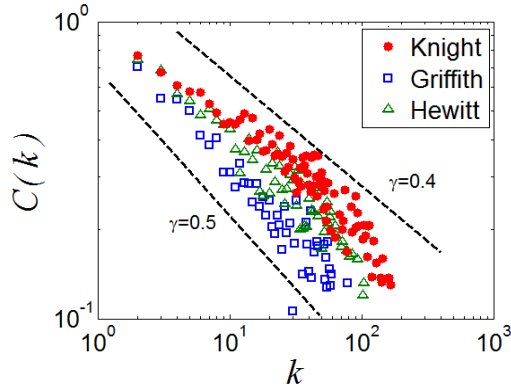


Fig. 4 Clustering coefficient distribution diagram of knowledge networks for three textbooks (in dual log coordinate)

4. Conclusions

Based on the physical terminology defined by Oxford Dictionary of Physics, three English edition textbooks of general physics were analyzed in the work. It was found that the frequency distribution of physical terminology follows the power law, $P(f) \sim f^{-\alpha}$, where $\alpha \approx 1.0$, which shows that there exist few repeatedly used terms such as *force* and *energy*, but most terms have low frequency. It is consistent with the result of analyzing for common life terms [15]. In the binary networks established for the three textbooks, the degree distribution follows $P(k) \sim k^{-\beta}$, where $0.8 < \beta < 1.0$, which indicates that the binary networks of textbooks have the property of scale free network, and possess the characteristics of growing network. Mean distances of the binary networks for the three textbooks are less than 3.0 and clustering coefficient is in the range of 0.2~0.3, which indicates that the networks possess the properties of small world network. In addition, the distribution of clustering coefficient follows the power law, $C(k) \sim k^{-\gamma}$, where $0.4 < \gamma < 0.5$, which shows the characteristics of hierarchical modular structure. This shows that there exist close relation between physics knowledge and physical

terminology, and furthermore, there exist relations among these physics knowledge.

Our work provided the method of establishing basic statistic data by collecting the physical terminology appeared in the textbooks, and the basic relations among physical terminology are found by constructing the knowledge networks. The proposed method of data collecting and analyzing may be widely used in other learning fields with its referential significance.

Acknowledgements

Our work was supported by Project of Development of Science and Technology of Yanbian University under Grant No.2012-12.

References

- [1] A-L. Barabási, R. Alber, H. Jeong, G. and Bianconi, "Power-law distribution of the world wide web," *Science*, vol. 287, pp. 2115, 2000.
- [2] H. Jeong, S.P. Mason, A-L. Barabási, and Z.N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, pp. 41-42, 2001.
- [3] C. A. Hidalgo, C. Rodriguez-Sickert, "The dynamics of a mobile phone network," *Physica A*, vol. 387, pp.3017-3024, 2008.
- [4] A. E. Motter, A. P. S. de Moura, Y. C. Lai, and P. Dasgupta, "Topology of the conceptual network of language," *Phys. Rev. E*, vol. 65, pp. 065102, 2002.
- [5] R. F. Cancho, and R. V. Sole, "The small world of human language," *Proc. R. Soc. Lond B*, vol. 268, pp. 2603-2606, 2001.
- [6] R.D. Knight, *Physics*, Addison-Wesley, New York, 2003.
- [7] W.T. Griffith, *The physics of everyday phenomena*, McGraw-Hill, New York, 1998.
- [8] P.G. Hewitt, *Conceptual physics*, Addison-Wesley, New York, 1997.
- [9] J. Daintith J (Editor), *Oxford dictionary of physics*, Oxford University Press, New York, 2005.
- [10] <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>.
- [11] http://en.wikipedia.org/wiki/Clustering_coefficient.
- [12] P. W. Holland, and S. Leinhardt, "Transitivity in structural models of small groups," *Comparative Group Studies*, vol. 2, pp. 107-124, 1971.
- [13] D. J. Watts, and Steven Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440-442, 1998.
- [14] D.J. Watts, and S.H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440-442, 1998.
- [15] R.F. Cancho, and R.V. Solé, "Two regimes in the frequency of words and the origins of complex lexicons: Zipf's law revisited," *J. Quan Ling*, vol. 8, pp. 165-173, 2001.