# A cross-language comparison of co-word networks in Digital Library and Museum of Buddhist Studies

## Introduction

This paper reports a co-words domain analysis of Buddhism literature collected by DLMBS (Digital Library and Museum of Buddhist studies) at National Taiwan University. Established in 1995, the DLMBS is one of the most comprehensive online repository of Buddhist research materials. It currently contains over 400 thousand records of books, research papers, theses and dissertations in 45 languages and digitized Buddhist scriptures. A controlled vocabulary, which is in five languages, including Chinese, English, Japanese, German, and French, was used to help users search DLMBS's bibliographic database. Using co-occurrence data of author assigned keywords in the bibliographic records, this study attempts to generate co-word networks in three different languages, Chinese, English, and Japanese, to compare regional focuses on Buddhist studies.

Co-words analysis has been shown to be effective in mapping the intellectual structure of disciplines (He, 1999; Leydesdroff, 1989). While it has been widely used in the domains of sciences and technologies (e.g. Buitelaar, Bordea, & Coughlan, 2014; Ding, Chowdhury, & Foo, 2001; Bhattacharya & Basu,1998; Looze, & Lemarie, 1997; Peters & van Raan, 1993a, 1993b; Courtial, 1994; Tijssen, 1992; Callon, Courtial, & Laville, 1991; Rip & Courtial, 1984), to the best our knowledge, it has so far not been applied to humanities. Part of the advantage of using co-word in sciences and technologies is the highly codified subject languages, therefore a higher degree of consistency between concepts and terms in these fields. We believe that a cross-language co-word analysis of Buddhist studies literature would be a worthwhile endeavor for a couple of reasons: Firstly, it has been pointed out that there has been a wide variety of methods, perspectives and subject matters within the international communities of Buddhist studies. The heterogeneity of its scholarships can be partly traced to their geographic roots (Cabezón, 1995). It is therefore interesting to empirically study whether and how the intellectual structures reflected in the published literatures in these language communities differ from one another. A comparison of the intellectual structures can shed light on knowledge interests shared and distinct in these three language communities. From the methodological plain, co-word analysis also provides a viable alternative to citation-based network analysis in humanities where the citation structure is known to be much sparser than in sciences and technologies.

## Procedures and analysis

Three separate co-word networks were generated in three different languages where nodes denote the keywords and edges the strength of their co-occurrence. Unlike in most of the previous co-words analysis, where keywords were extracted from titles and abstracts, author assigned keywords were used here as it is believed that they are more representative to the content of the articles and tend to have higher degree of consistency than keywords in free-text. For monographs and other types of publications, the subject-headings assigned by human indexers were used as keywords. Edge weights were normalized by both the inclusion and the Jaccard index (Courtial,1986; Callon, Law, & Rip,1986).

Thus three word similarity matrixes were generated so social network analytical methods such as cohesion, centrality and community-detection (Blondel et.al., 2008) could be performed with a view to exploring the social and cognitive structure of Buddhist studies manifested in its published literature in respective languages. Specifically, the study seeks to answer the following interrelated research questions: firstly, are there recognizable branches or specialties in Buddhist studies? If so, what these areas of research might be. Cross-languages comparisons were also made to examine the similarities and differences of the intellectual structure in different language communities.

## Results

### Descriptive data

Table 1 gives the numbers of items pertinent to various publication types analyzed in three languages in DLMB.

Table 1. Types of publications analyzed

| Language<br>Publication Type | Chinese | English | Japanese |
|---|---|---|---|
| Journal Article | 45,025 | 18,703 | 33,311 |
| Book | 12,644 | 13,604 | 8,869 |
| Thesis and Dissertation | 3,757 | 1,894 | 49 |
| Research Paper | 3,372 | 200 | 663 |
| Proceeding Article | 2,409 | 248 | 84 |
| Journal Article; Book Review | 158 | 2,201 | 377 |
| Sound Recording | 265 | 516 | 16 |
| Serial | 427 | 240 | 96 |
| Reference Book | 393 | 139 | 39 |
| Audiovisual | 45 | 304 | 16 |
| Book Review | 49 | 225 | 22 |
| Internet Resource | 59 | 207 | 0 |
| Collected Papers | 65 | 15 | 83 |
| Others | 15 | 33 | 0 |
| E-Book | 2 | 21 | 0 |
| Book; Internet Resource | 0 | 7 | 0 |
| Book; Sound Recording | 1 | 0 | 0 |
| Internet Resource; Book Review | 2 | 0 | 0 |
| Internet Resource; Journal Article | 0 | 2 | 0 |
| Book Review; Internet Resource | 0 | 1 | 0 |
| Total | 68,688 | 38,560 | 43,625 |

Due to the enormous size of the networks, some sorts of filtering are required to make the groupings intelligible; node degrees was used as the filter as many of the little connected nodes tend to generate noises that impairs meaningful interpretation. To determine the proper threshold of node degree, one needs to consider three criteria: the quality of the clustering, the interpretability of the individual clusters, and the preservation of information. A high threshold would filter out large amount of nodes hence the greater loss of information and low modularity values. On the other hand, a low threshold would result in difficulty in interpreting individual clusters as they tend to lump together heterogeneous topics. Thus a trade-off needs to be made. We approached this matter by performing modularity analysis at different threshold levels so the values of their modularity, the resulting number of communities, as well as the size of the networks could be recorded. The following heuristics were used to select the proper thresholds: to preserve about 10 percent of the total nodes, to limit the number of communities from 10 to 20, and to preserve a high degree of modularity, which is commonly used as the indicator of clustering quality.

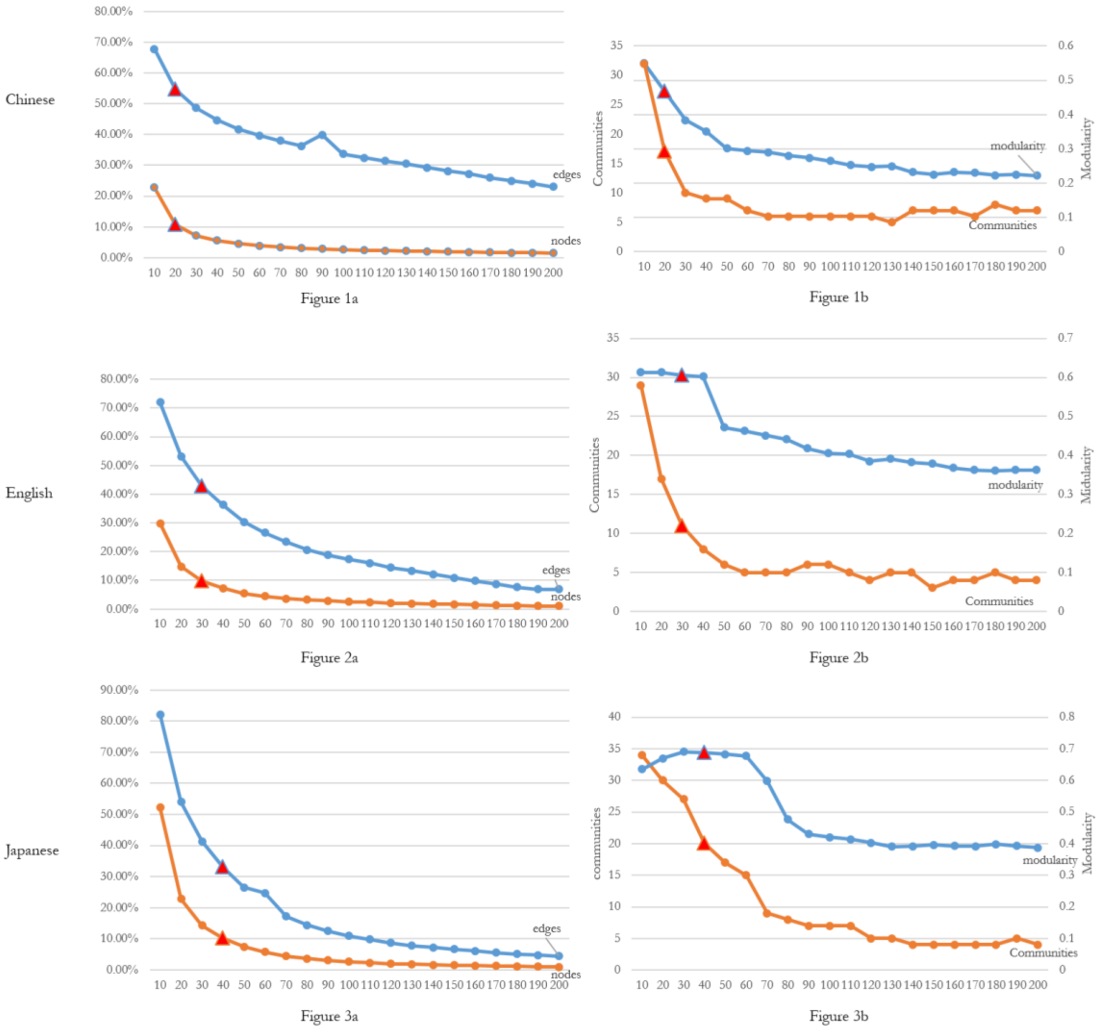

Figure 1a

Figure 1b

Figure 2a

Figure 2b

Figure 3a

Figure 3b

Figure(1-3). Node degree thresholds and resulting network attributes in three languages

Table 2 reports the thresholds and their corresponding attributes of the resulting networks.

Table 2. Descriptive statistics of the three networks

| | # of nodes | # of edge | Threshold | # of Communities | % of nodes | % of edges |
|---|---|---|---|---|---|---|
| Chinese | 58,808 | 787,682 | 20 | 17 | 10.66 | 54.61 |
| English | 34,093 | 325,161 | 30 | 11 | 9.66 | 42.74 |
| Japanese | 75,728 | 859,469 | 40 | 20 | 10.09 | 33.04 |

After filtering out lesser connected nodes, modularity maximizing community detection method was then performed to identify the subdomains in each language network. A two-stage approach was adopted here. As some of the communities resulted from the first-round of clustering can still be very broad and heterogeneous, a second modularity analysis was performed on these relative "super" clusters (i.e. clusters with more than 400 nodes), the joint results produce a two-levels hierarchical structure. Three experts in Buddhist studies were then interviewed to help us interpret the clusters (See Figure 4 and 5).

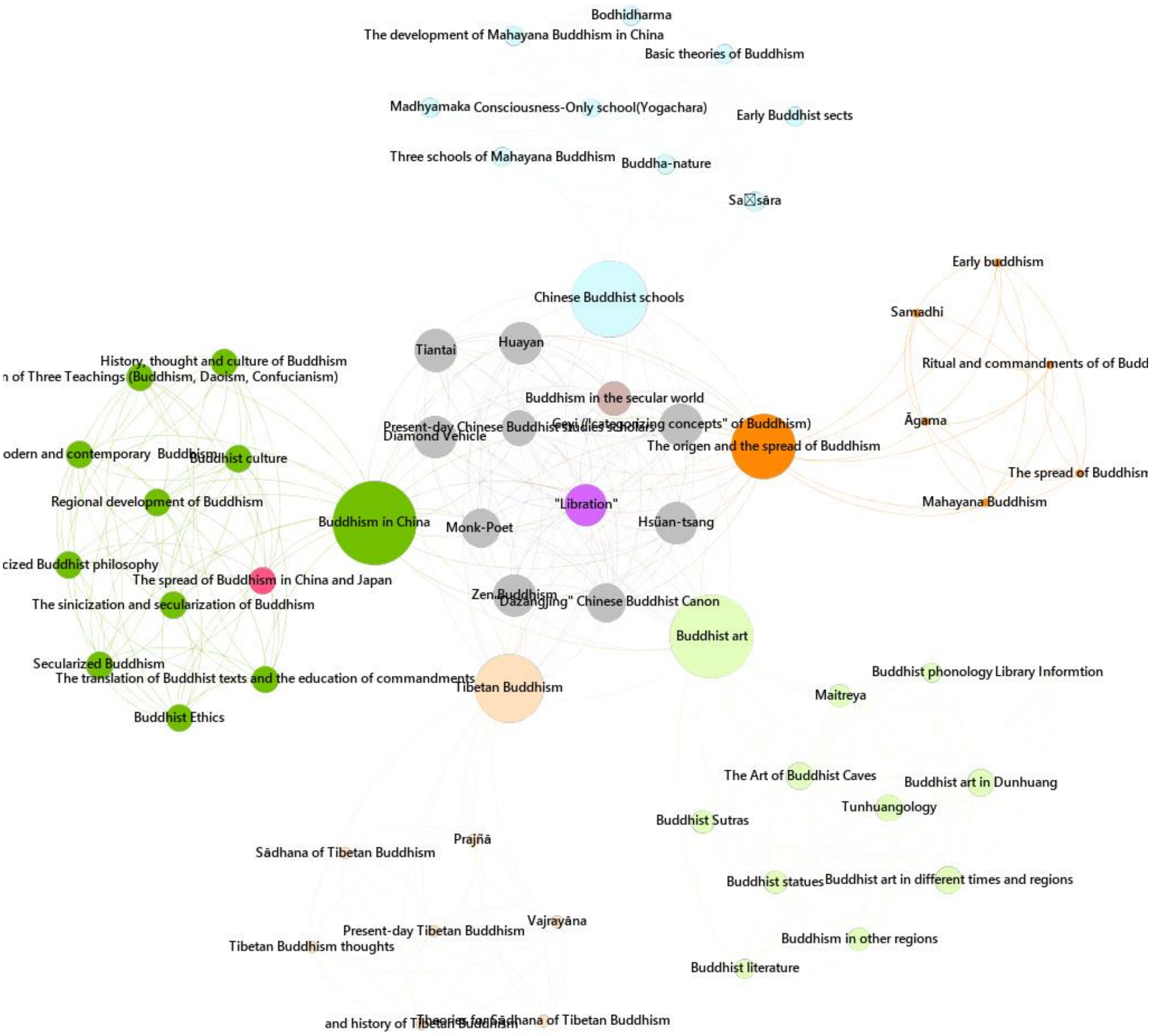
Bodhidharma
The development of Mahayana Buddhism in China
Basic theories of Buddhism
Madhyamaka Consciousness-Only school(Yogachara)
Early Buddhist sects
Three schools of Mahayana Buddhism
Buddha-nature
Sa⊠sāra
Early buddhism
Chinese Buddhist schools
Samadhi
Tiantai
Huayan
History, thought and culture of Buddhism
n of Three Teachings (Buddhism, Daoism, Confucianism)
Ritual and commandments of of Budd
Buddhism in the secular world
Present-day Chinese Buddhist studies scholars
Diamond Vehicle
Geyi ("categorizing concepts" of Buddhism)
Āgama
odern and contemporary Buddhism
Buddhist culture
The origen and the spread of Buddhism
The spread of Buddhisn
Regional development of Buddhism
"Libration"
Mahayana Buddhism
Buddhism in China
Monk-Poet
Hsüan-tsang
cized Buddhist philosophy
The spread of Buddhism in China and Japan
The sinicization and secularization of Buddhism
Zen Buddhism
"Dazangjing" Chinese Buddhist Canon
Buddhist art
Secularized Buddhism
The translation of Buddhist texts and the education of commandments
Tibetan Buddhism
Buddhist phonology Library Informtion
Maitreya
Buddhist Ethics
The Art of Buddhist Caves
Buddhist art in Dunhuang
Tunhuangology
Buddhist Sutras
Prajñā
Sādhana of Tibetan Buddhism
Buddhist statues
Buddhist art in different times and regions
Vajrayāna
Present-day Tibetan Buddhism
Tibetan Buddhism thoughts
Buddhism in other regions
Buddhist literature
and history of Tibetan Buddhism
Theories for Sādhana of Tibetan Buddhism

Figure 4. Visualization of intellectual structure in Chinse Buddhist studies.

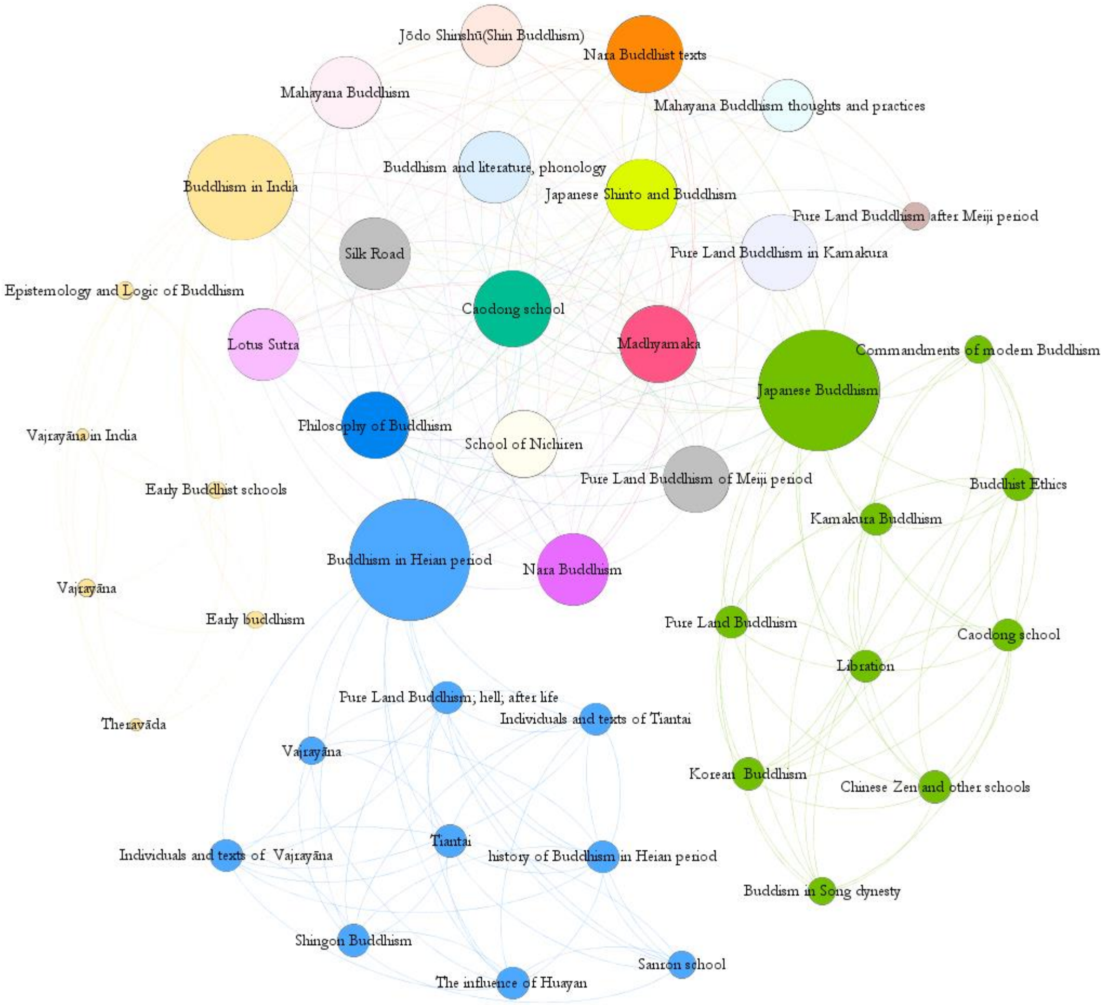
Jōdo Shinshū(Shin Buddhism)
Nara Buddhist texts
Mahayana Buddhism
Mahayana Buddhism thoughts and practices
Buddhism in India
Buddhism and literature, phonology
Japanese Shinto and Buddhism
Pure Land Buddhism after Meiji period
Silk Road
Pure Land Buddhism in Kamakura
Epistemology and Logic of Buddhism
Caodong school
Lotus Sutra
Madhyamaka
Commandments of modern Buddhism
Japanese Buddhism
Philosophy of Buddhism
Vajrayāna in India
School of Nichiren
Pure Land Buddhism of Meiji period
Early Buddhist schools
Buddhist Ethics
Kamakura Buddhism
Buddhism in Heian period
Nara Buddhism
Vajrayāna
Early buddhism
Pure Land Buddhism
Caodong school
Libration
Pure Land Buddhism; hell; after life
Individuals and texts of Tiantai
Theravāda
Vajrayāna
Korean Buddhism
Chinese Zen and other schools
Tiantai
Individuals and texts of Vajrayāna
history of Buddhism in Heian period
Buddism in Song dynesty
Shingon Buddhism
Sanron school
The influence of Huayan

Figure 5. Visualization of intellectual structure in Japanese Buddhist studies.

In this study we utilized co-word network to visually represent the domain of Buddhist studies. A heuristic was proposed to help select the proper threshold in order to filter less significant keywords. A two-stage clustering approach was adopted, which arguably provides a finer representation of the intellectual structure of the domain. Further analysis will be done, with the help of domain experts, to compare the differences in the intellectual structure reflects in three language communities.

## References

Bhattacharya, S. and Basu, P. (1998), "Mapping a research area at the micro level using co-word analysis", *Scientometrics,* 43(3), pp. 359-372.

Blondel, V., Guillaume, J., Lambiotte, R. and Lefebvre, E. (2008), "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), pp. 10008.

Buitelaar, P., Bordea, G., & Coughlan, B. (2014), "Hot Topics and Schisms in NLP: Community and Trend Analysis with Saffron on ACL and LREC Proceedings". In *9th Edition of Language Resources and Evaluation Conference (LREC2014).*

Cabezón, J. I. (1995), "Buddhist Studies as a Discipline and the Role of Theory", *Journal of the International Association of Buddhist Studies*, 18(2), pp. 231-268.

Callon, M., Courtial, J. and Laville, F. (1991), "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry", *Scientometrics*, 22(1), pp. 155-205.

Callon, M., Law, J., & Rip, A. (1986), "Qualitative scientometrics", in *Mapping the dynamics of science and technology*, Palgrave Macmillan UK, pp. 103-123.

Courtial, J. (1994), "A coword analysis of scientometrics". *Scientometrics*, 31(3), pp. 251-260.

Courtial, J. P. (1986), "Technical issues and developments in methodology", in *Mapping the Dynamics of Science and Technology*, Palgrave Macmillan UK, pp. 189-210.

Ding, Y., Chowdhury, G. G., & Foo, S. (2001), "Bibliometric cartography of information retrieval research by using co-word analysis". *Information processing & management*, 37(6), pp. 817-842.

He, Q. (1999), "Knowledge discovery through co-word analysis", *Library trends*, 48(1), pp. 133-133.

Leydesdroff, L. (1989), "Words and co-words as indicators of intellectual organization", *Research policy*, *18*(4), pp. 209-223.

Peters, H. P. F., & van Raan, A. F. (1993), "Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling", *Research Policy*, 22(1), pp. 23-45.

Tijssen, R. J. (1992), "A quantitative assessment of interdisciplinary structures in science and technology: co-classification analysis of energy research", *Research policy*, 21(1), pp. 27-44.