

## Article

# Observations on the Intertextuality of Selected Abhidharma Texts Preserved in Chinese Translation

Sebastian Nehrdich

Berkeley Artificial Intelligence Research Lab, University of California, Berkeley, CA 94720, USA; nehrdich@berkeley.edu

**Abstract:** Textual reuse is a fundamental characteristic of traditional Buddhist literature preserved in various languages. Given the sheer volume of preserved Buddhist literature and the often-unmarked instances of textual reuse, the thorough analysis and evaluation of this material without computational assistance are virtually impossible. This study investigates the application of computer-aided methods for detecting approximately similar passages within Xuanzang's translation corpus and a selection of Abhidharma treatises preserved in Chinese translation. It presents visualizations of the generated network graphs and conducts a detailed examination of patterns of textual reuse among selected works within the Abhidharma tradition. This study demonstrates that the general picture of textual reuse within Xuanzang's translation corpus and the selected Abhidharma texts, based on computational analysis, aligns well with established scholarship. Thus, it provides a robust foundation for conducting more detailed studies on individual text sets. The methods employed in this study to create and analyze citation network graphs can also be applied to other texts preserved in Chinese and, with some modifications, to texts in other languages.

**Keywords:** Buddhist literature; Abhidharma; textual reuse; citation network graph; citation network



**Citation:** Nehrdich, Sebastian. 2023. Observations on the Intertextuality of Selected Abhidharma Texts Preserved in Chinese Translation. *Religions* 14: 911. <https://doi.org/10.3390/rel14070911>

Academic Editor: Marcus Bingenheimer

Received: 7 March 2023

Revised: 10 July 2023

Accepted: 11 July 2023

Published: 14 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Research into Sarvāstivāda Abhidharma literature heavily relies on Chinese translations because much of the relevant literature is accessible only in this form, as the Indic originals have either been lost or are not currently accessible. The sheer volume of these translations can be initially overwhelming, with complex relationships existing between individual texts, particularly within the canonical Sarvāstivāda Abhidharma works. Texts of the Buddhist tradition in general and Sarvāstivāda Abhidharma literature in particular are characterized by a high degree of intertextuality, leading to frequent and usually not clearly marked instances of textual reuse. Unlike contemporary Western cultures, where unmarked citations and textual borrowings are typically considered plagiarism, such practices are generally viewed as commendable in the Indian tradition. Even certain Sanskrit educational texts, such as Kṣemendra's *Kavikanṭhabharana*, explicitly encourage the art of copying, rewriting, and reusing material from previous works as an expression of stylistic prowess.

In a similar vein, the intertextuality of Sarvāstivāda Abhidharma literature heavily features unmarked textual reuse. Recent publications, such as Sakuma and Kragh (2013) and Kramer (2014), acknowledge the complex educational landscape and the varied attitudes towards textual reuse encountered in the study of Buddhist treatises. This situation makes it impossible for individual researchers to examine the material exhaustively without computational assistance.

Dedicated tools for the study of the intertextuality of Buddhist source texts such as BuddhaNexus became available recently.<sup>1</sup> However, further systematic processing is necessary in order to make the data accumulated by these tools accessible for corpus-level examination. This part of the study used methods of network graph analysis on top of the data accumulated in the BuddhaNexus. The aim was to visualize the layout

of the translation corpus of Xuanzang 玄奘 (602–664) as well as to give insights into the intertextuality of selected Abhidharma works preserved in Chinese translation, including texts not translated by Xuanzang. Given the well-established influence of the north-western Abhidharma tradition not only on Buddhism but on the greater South Asian intellectual history in general,<sup>2</sup> analyzing its literature can enhance our understanding of this crucial period of South Asian intellectual history. The network graph analysis of textual reuse offers valuable insights into the overarching structure of the intertextuality of a corpus—insights that are challenging to gain when focusing solely on individual texts.

A number of earlier studies utilizing algorithmic fuzzy matching for the intertextual analysis of Chinese texts exist. [Vierthaler and Gelein \(2019\)](#) described a local alignment-based textual reuse detection algorithm and its application to Ming literature, including visualizations of the networks of intertextuality. [Tharsen and Gladstone \(2020\)](#) applied algorithmic intertextual analysis based on the PhiloLogic system to the *Twenty-four Chinese Histories* corpus. [Nicoll-Johnson \(2018\)](#) utilized manually collected data on citations in early medieval Chinese literature for the visualization of citation networks.

In the first part of this study, an undirected citation network graph was constructed in order to better observe the layout of the works translated by the prolific translation team guided by Xuanzang. For this purpose, continuous word embeddings and vector indexing were used in order to detect similar passages between different texts, which were then used to generate a graph of parallel matches of the Buddhist scriptures and treatises translated by the team of Xuanzang preserved in the Taishō canon. The generated data are presented in the form of a network graph visualization. In the second part, a selection of texts from the Abhidharma genre was analyzed in detail with the help of the maximum spanning tree algorithm. In the third part of this study, the results of the maximum spanning tree analysis were used in combination with unsupervised text segmentation to generate a set of bipartite network graphs that illustrate the intertextuality of individual sections of a selected number of Abhidharma treatises ranging from the *Saṅgītiparyāya* and the *Dharmaskandha* up to the *Abhidharmakośabhāṣya*.

### *Preliminary Remarks*

Given the difficult transmission history of Buddhist textual material, especially of the Abhidharma genre, the material accessible nowadays is unlikely to be representative of the text production of the time of its origin. It is reasonable to assume that only a fraction of the once-existing texts have been transmitted, and one can only speculate about what might have been lost. On top of the already complex intertextual situation of the material at the time of its composition in India, its transmission to China added another layer of complexity. Chinese translators exhibited a wide range of attitudes towards instances of textual reuse in the translated material. Regarding the translation team of Xuanzang, which lies at the center of this study, it is generally acknowledged that it produced translations of excellent quality.<sup>3</sup> At times, these translations include source information on quoted passages within the translated texts that are not found in the Indic versions of the texts, suggesting that this source information was supplied by the translators. However, these supplementary notations are not systematically encountered, and unattributed instances of textual reuse frequently occur in these translations. The vast majority of instances are inherited from the Indic originals. This study operates under the assumption that Xuanzang's team had the ability to consistently translate similar passages in different texts and maintain uniform vocabulary for technical terms. Naturally, it would be unreasonable to assume perfect consistency in all cases, but it is assumed that such consistency occurred in the majority of cases.<sup>4</sup>

## **2. Visualization of the Translation Canon of Xuanzang**

The data utilized for the visualization and comparison of individual texts in this paper were derived from BuddhaNexus.<sup>5</sup> Briefly, a three-step method was employed to locate approximate textual matches among individual texts:

1. FastText Vector representations (Bojanowski et al. 2016) of individual tokens were created and averaged over a fixed window size in order to generate phrase representations.
2. A k-nearest neighbors search was then used to find similar phrase representations within a fixed corpus. Continuous chains of similar phrase representations were merged together to form longer matches.
3. Local alignment (the Smith–Waterman algorithm) was then used to find the exact beginning and end points of the detected matches.

The matches generated through this process typically exhibited a high degree of lexical similarity and could, therefore, serve as indicators of potential textual reuse. This was particularly true for matches that extended beyond a few characters, for instance:

<i>Dhātukāya</i> (T1540)	<i>Vijñānakāya</i> (T1536)
色為緣生於眼識，三和合故觸，觸為緣故受，受為緣愛。此中眼為增上、色為所緣，於眼所識色，諸貪等貪，執藏防護愛樂耽著，	色為緣生眼識，三和合故觸，觸為緣故受，受為緣故愛。此中眼為增上、色為所緣，於眼所識色，諸貪等貪執藏防護耽著愛樂，

With shorter matches, possible textual reuse was not always the appropriate explanation. The sheer collocation of similar vocabulary, especially of technical terms comprising multiple characters, could already lead to short matches that are of little research value. For example:

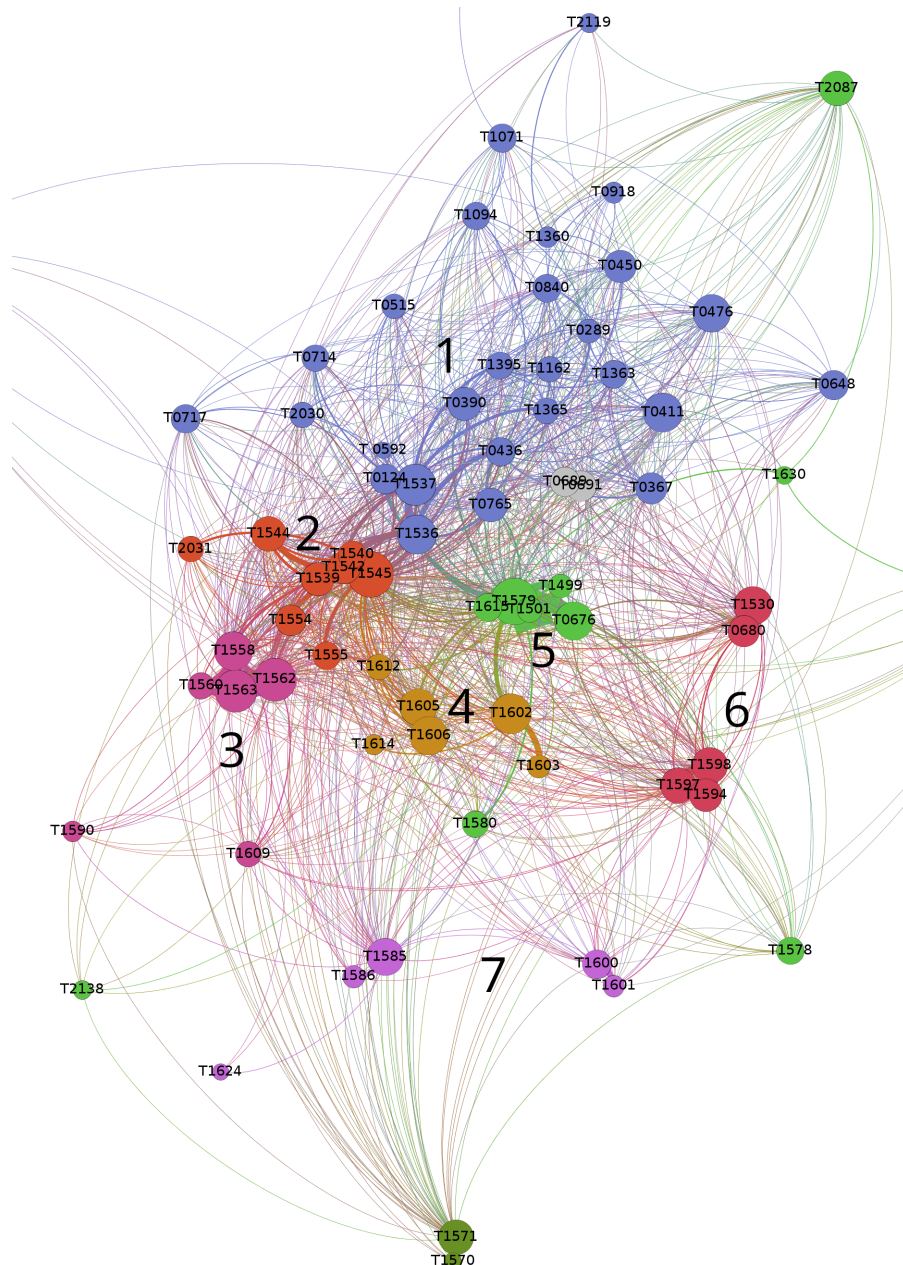
<i>Dhātukāya</i> (T1540)	<i>Vijñānakāya</i> (T1536)
觸不相應十八界、十二處、五蘊。	復次於欲界繫十八界十二處五蘊諸法中諸貪等貪，

Determining the boundary between valid instances of textual reuse and data noise composed of similar vocabulary collocations is challenging. Generally, longer matches are more likely to indicate instances of textual reuse. To minimize the impact of short, noisy matches, this study only considered matches exceeding 12 characters in length. Matches that signified text titles, author names, translator names, etc., were also excluded, as they were likely to distort the results.

The temporal relationship of the material analyzed in this study was largely unclear and therefore presented an important research question in itself. A substantial number of the early Abhidharma texts were likely composed by groups of people rather than individual authors, developed over a longer span of time, were subject to varying levels of interpolation, and at times consist of very heterogeneous material that was compiled from various sources. One can indeed call into question whether proposing a theory of chronology for this material makes sense at all or not.<sup>6</sup> Various attempts, both from within the tradition as well as from contemporary scholarship, at proposing a relative chronology of these texts have been made. While the methods used in this study were not sufficient to give answers to this question, they could give a general overview of the doctrinal relationship and progression within these texts, which could contribute to the establishment of a theory of their relative chronology. For this reason, the observations made here with the network-graph-based methods were compared with the established opinion of recent scholarship to see whether the picture matched up or whether obvious disagreement was observed. Because of the difficult and unclear nature of the temporal relationship among the analyzed texts, the data were approached in the form of an undirected network graph that left the question of their chronology open.

Figure 1 on page 4 shows the visualization of the undirected citation network graph of Xuanzang’s translation corpus produced by Gephi’s<sup>7</sup> ForceAtlas 2 algorithm. The nodes represent the respective texts of the Taishō. In order to make the display of this large number of nodes feasible, only the Taishō numbers are shown in the visualization, but not the titles of the individual texts. The edges represent the connection between individual texts weighted by the number of shared parallel matches. The weight of each edge was calculated in the following way: For the texts  $t_i$  and  $t_j$ , the total length of all tokens for  $t_i$

is  $l_i$  and  $t_j$  was  $l_j$ . The total number  $n_{ij}$  of shared tokens between  $t_i$  and  $t_j$  was the sum of all tokens contained in parallel matches between  $t_i$  and  $t_j$ . The weight  $w$  of the edge was thus calculated by dividing  $n_{ij}$  by the length of each text separately and summing up the result:  $w = \frac{n_{ij}}{l_i} + \frac{n_{ij}}{l_j}$ . Inspired by the work of Hellwig (2013), community detection based on modularity was performed, and the nodes and edges were then colored based on the detected communities. Additionally, the PageRank algorithm (Brin and Page 1998) was used to indicate the possible importance of a text based on how strongly it was connected to other texts. The results of the PageRank algorithm are reflected in the size of the nodes; the bigger the node, the higher the PageRank score compared to the other texts.



**Figure 1.** Visualization of the translations by Xuanzang. In order to make the display feasible, the nodes T0251 (upper left), T1034 (upper middle), and T1628 (middle right) were cropped. Numbers 1–7 indicate seven distinct communities that are the result of the application of community detection.



### 3. Analysis of the Visualization

The detected communities did indeed reflect to some extent the traditional classification of the texts, even though some interesting deviations could also be observed. The main communities could be summarized as follows:

1. The largest community, which consisted mainly of Sūtra translations in blue in the upper half of the visualization.
2. The red community to the bottom left-hand side of community 1, which centered around the \**Mahāvibhāṣā* (T1545) and included the canonical Sarvāstivāda texts *Dhātukāya* (T1540), the *Prakaraṇapāda* (T1542), the *Vijñanakāya* (T1539), and *Jñānaprasthāna* (T1544).
3. The purple community centering around the *Abhidharmakośabhāṣya* (T1558) to the bottom left of community 2. Apart from Saṅgabhadra's commentaries, two Yogācāra works attributed to Vasubandhu, the *Viṃśikā* (T1590) and the *Karmasiddhiprakaraṇa* (T1609), were also included here.
4. The dark orange community to the right of communities 2 and 3, which was dominated by the works attributed to Asaṅga, centering around the *Abhidharmasamuccaya* (T1605).
5. The green community, which was dominated by the *Yogācārabhūmi* (T1579), above community 4 and to the right of community 2.
6. The dark red community on the right side of the visualization, which consisted of the texts related to the *Mahāyānasamgraha* (T1594) and the \**Buddhabhūmisūtraśāstra* (T1530).
7. The purple community at the bottom of the visualization, with the Chéng Wéishì Lùn 成唯識論 (T1585) and the *Madhyāntavibhāṅgabhāṣya* (T1600) as its centers.

Within the Sūtra community 1, two Abhidharma treatises, the *Saṅgītiparyāya* (T1536) and the *Dharmaskandha* (T1537), were found. Collet Cox remarked that the style and content of these early Abhidharma texts are hard to distinguish from those of the Sūtras, and it is therefore indeed challenging to determine the demarcation line between early Abhidharma and Sūtra literature based on style and/or content alone (Cox 1998, pp. 168–70). Both texts are commonly regarded as belonging to the earliest layer of Sarvāstivāda Abhidharma literature.<sup>8</sup> For these reasons, it is not entirely surprising that the algorithm was not able to properly distinguish these texts from other Sūtra literature.<sup>9</sup>

Interestingly, the travel report *Dà Táng xīyù jì* 大唐西域記 (T2087) formed its own separate community and was placed at the top right of community 1. Regarding community 3, which was centered around the *Abhidharmakośabhāṣya* (T1558) and contained further works ascribed to Vasubandhu, another Yogācāra text attributed to Vasubandhu, the *Pañcaskandhaka* (T1612), was found in close vicinity, even though it was clustered into community 4. It is worth noting that T1590, T1609, and T1612 are works of Vasubandhu that were argued in Schmithausen (1967) to exhibit Sautrāntika doctrinal influences and therefore likely have the same author as the *Abhidharmakośabhāṣya* (T1558), while the commentaries on Maitreya/Asaṅga-works,<sup>10</sup> which are traditionally attributed to Vasubandhu as well but which Schmithausen argued to belong to a separate group, appeared in community 6 in the rightmost part of this visualization.<sup>11</sup> Below community 3, a small, distinctly green community consisted of one text of the non-Buddhist Vaiśeṣika school, the *Shèngzōng shíjùyì lùn* 勝宗十句義論 (T2138). Given that both communities primarily consisted of Abhidharma treatises, it is unsurprising that community 2 and community 3 were depicted in close proximity. The *Abhidharmakośabhāṣya* (T1558), which lay at the center of community 3, built on the doctrinal basis of the Vaibhāṣika tradition, which was laid out in the texts of community 2. Community 4, which was dominated by the works attributed to Asaṅga, was located at the crossroads between the Abhidharma treatises to the left of the visualization and the Yogācāric in the middle and to the right.

Two general observations could be made regarding the visualization:

1. Sūtra material tended to accumulate in community 1 at the top, while Śāstra material gathered in communities 2–7 at the bottom.

2. The left side was dominated by Abhidharma works and the “Sautrāntika” Yogācāra works attributed to Vasubandhu in community 3, the middle by works ascribed to Asaṅga and the *Yogācārabhūmi* (T1579) in community 4 and 5, and the Maitreya texts and their respective commentaries were found on the right side of the visualization in community 6 and at the bottom in community 7.

It could even be proposed that there was a chronological trend, with older material located at the top and younger material at the bottom of the visualization. This holds true for at least the progression of *Saṅgītiparyāya* (T1536) and *Dharmaskandha* (T1537) -> *Dhātukāya* (T1540) -> *Prakaraṇapāda* (T1542) -> *\*Mahāvibhāṣā* (T1545) and *Yogācārabhūmi* (T1579) -> *Abhidharmakośabhāṣya* (T1558) and *Abhidharmasamuccaya* (T1605) -> Chéng Wéishì Lùn 成唯識論 (T1585), as there is largely a consensus about the relative chronology of these texts.<sup>12</sup> There were also outliers to this observation. For instance, the position of the *Jñānaprasthāna* (T1544) did not align well with the arrangement of the other works in community 2. Similarly, the positions of the *Triṃśikā* (T1586) and the *Madhyāntavibhāga* (T1600) and its commentary *Madhyāntavibhāgabhāṣya* (T1601) in community 7 were also not very convincing when seen in the context of other Yogācāra works. The influence of the many common passages between the undoubtedly late Chéng Wéishì Lùn 成唯識論 (T1585) and these texts might serve as a possible explanation for their placement at the bottom of the visualization.

Table 1 shows the edges with the highest weights in the citation network graph limited to Abhidharma and Yogācāra works; Sūtra material was left out, since it was not a focus of this study. It was not very surprising that two of these edges, that between the *Yogācārabhūmi* (T1579) and the Púsà jièběn 菩薩戒本 (T1501) as well as that between the *Yogācārabhūmi* (T1579) and the Wángfǎ zhènglǐ lùn 王法正理論 (T1615), were between texts that contained lengthy translations of the same Indic source material. With the exception of the edge between the *Prakaraṇapāda* (T1542) and the *Dhātukāya* (T1540), all of these edges were connected to either the *Yogācārabhūmi* or the *\*Mahāvibhāṣā*, which is a testimony to how central these texts were for the treatises that Xuanzang’s team translated. Table 2 shows the five texts with the highest PageRank score in this network graph. Here, again, the *Yogācārabhūmi* and the *\*Mahāvibhāṣā* occupied the highest positions. It might be slightly surprising at first sight that two works attributed to Saṅgabhadra, the *\*Nyāyānusāriṇī* and the *\*Samayapradīpikā*, were found in this list and not the *Abhidharmakośabhāṣya* on which they comment. This is explainable by the fact that they are both longer, 36.158 characters for the *\*Nyāyānusāriṇī* and 17188 characters for the *\*Samayapradīpikā*, than the *Abhidharmakośabhāṣya*, with 14633 characters. And since both commentaries contain lengthy literal matches with the *Abhidharmakośabhāṣya* in addition to matches with other texts, these contributed to a higher overall PageRank score. The placement of the *Dharmaskandha* in this list is interesting and might be explained by the fact that this text mainly consists of commentaries on doctrinal concepts scattered throughout the texts of the Sūtra category, with which it therefore exhibits a significant number of shared edges, contributing to its comparatively high PageRank score.

**Table 1.** Highest edge weights in the undirected citation network graph.

Text A	Text B	Weight
<i>Yogācārabhūmi</i> (T1579)	Púsà jièběn 菩薩戒本 (T1501)	5.91
<i>Prakaraṇapāda</i> (T1542)	<i>*Mahāvibhāṣā</i> (T1545)	3.46
<i>Prakaraṇapāda</i> (T1542)	<i>Dhātukāya</i> (T1540)	3.40
<i>Yogācārabhūmi</i> (T1579)	Wángfǎ zhènglǐ lùn 王法正理論 (T1615)	2.74
<i>Dhātukāya</i> (T1540)	<i>*Mahāvibhāṣā</i> (T1545)	2.59

**Table 2.** Texts with the highest PageRank scores in the undirected citation network graph.

Text	PageRank Score
<i>Yogācārabhūmi</i> (T1579)	0.028
* <i>Mahāvibhāṣa</i> (T1545)	0.027
* <i>Nyāyānusāriṇī</i> (T1562)	0.025
* <i>Samayapradīpikā</i> (T1563)	0.024
<i>Dharmaskandha</i> (T1537)	0.023

#### 4. Maximum Spanning Tree Analysis of Selected Abhidharma Works

Figure 2 on page 8 shows a selection of nodes beginning with the canonical Abhidharma texts of the Sarvāstivāda school up until the commentaries on the *Abhidharmakośabhāṣya* after the application of the maximum spanning tree algorithm (MST, Kruskal (1956)). The maximum spanning tree is a subset of the edges of a connected, edge-weighted undirected network graph that connects all the vertices together, without any cycles, and with the maximum possible total edge weight. The community detection in this analysis identified four distinctive groups. The first group, in green, consisted of *Dharmaskandha* and *Saṅgītiparyāya*. The second group, in orange, included *Dhātukāya* and *Prakaraṇapāda*. The third group, in blue, comprised *Vijñānakāya*, *Jñānaprasthāna*, \**Mahāvibhāṣā*, and T1554 and T1555. The final group, in red, consisted of *Abhidharmakośabhāṣya* and its commentaries. The arrangement of the texts into these four groups followed a pattern that fit well with the consensus regarding their relative chronology / doctrinal development in contemporary scholarship (Cox 1998, p. 171 ff.), with the least developed group at the bottom left and the most refined group at the top right of the visualization. The first group, consisting of the *Dharmaskandha* and *Saṅgītiparyāya*, is generally regarded as forming the earliest layer of the canonical Abhidharma texts. Regarding the relative chronology of these two texts, there is disagreement (Cox 1998, p. 172). Frauwallner regards the *Saṅgītiparyāya* as the older of the two (Frauwallner 1995, p. 20), while Fukuhara sees the *Dharmaskandha* as the older text (Fukuhara Ryōgon (福原亮嚴) 1965, p. 110). Tanaka Kyōshō argues that the relative chronology of both texts cannot be established clearly (Cox 1998, p. 172). Yin Shun argues that the *Saṅgītiparyāya* was composed after the *Dharmaskandha*, but its content is more archaic and less developed (Yin Shun (印順) 1968, p. 135a4-14). The next group consists of the *Dhātukāya* and the *Prakaraṇapāda*, two texts that in scholarship have been observed to have a close relationship as well, with a general consensus that the *Dhātukāya* must have existed in some form before the *Prakaraṇapāda* (Cox 1998, p. 207). Yin Shun goes against this view and sees it as later than the *Prakaraṇapāda* (Yin Shun (印順) 1968, p. 162a5-9). The blue group dominated by the \**Mahāvibhāṣā* was the largest group in this visualization. The inclusion of *Jñānaprasthāna* in very close proximity to the \**Mahāvibhāṣā* is not very surprising, since the \**Mahāvibhāṣā* is a commentary on the *Jñānaprasthāna*, and a plethora of textual matches between both texts were detected. The proximity of the *Vijñānakāya* to the \**Mahāvibhāṣā* in this group was an outlier and contradicted common scholarship, which shows a general tendency to place it somewhat near the *Dhātukāya*.<sup>13</sup> The grouping of the green, orange, and blue communities with the exception of the placement of the *Vijñānakāya* also corresponded to Puguang 普光, a student of Xuanzang, who placed *Dharmaskandha* and *Saṅgītiparyāya* in the earliest layer composed of direct disciples of the Buddha, *Dhātukāya* and *Prakaraṇapāda* in the layer of works composed at the beginning of the third century after the Buddha, and the *Jñānaprasthāna* at the end of the third century after the Buddha (T1821 8b24-8c07).<sup>14</sup>

As these results show, the application of the MST based on an undirected network graph of quantified textual reuse could serve to elucidate the general doctrinal relationship and trends between the texts in a way that matched with observations made in the scholarship. However, it is important to note that these observations on doctrinal relationships did not necessarily correspond to temporal relationships. In the case of the *Vijñānakāya* for example, its close placement to the \**Mahāvibhāṣā* should not be taken as an indicator that

it necessarily postdates the texts of the green and orange groups. Rather, this placement can also be explained by the innovative character of the *Vijñānakāya*, which according to Frauwallner breaks from the descriptive character of the other early Abhidharma works and instead presents new problems and a new approach (Frauwallner 1995, p. 30). It is comparatively frequently quoted in the *Mahāvibhāṣa*,<sup>15</sup> while its content did not receive much attention in the other Sarvāstivāda treatises, which could explain its placement in this visualization.

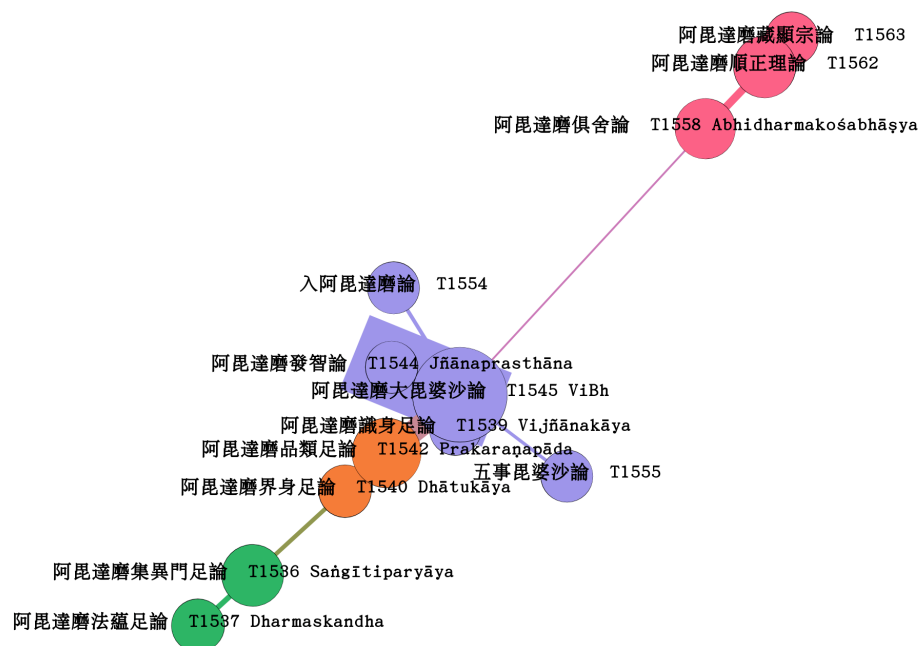


Figure 2. Maximum spanning tree (MST) of the Abhidharma literature in Xuanzang’s translation.

### 5. Visualization of the Relationships between Individual Sections of Selected Abhidharma Works

In order to learn about the relationships between individual sections of these texts, a method of text segmentation first needed to be applied, since not all of the texts are divided into chapters or other units that would make a detailed analysis feasible. The method used in this study was the unsupervised text segmentation method employing a semantic relatedness graph (Glavaš et al. 2016).<sup>16</sup> The results of the unsupervised text segmentation are intrinsically interesting, as they revealed the degree of semantic coherence in each text and whether it is composed of larger unified blocks or smaller independent units. As this method employed the average of the vector representations of extended text sections, it facilitated the comparison of texts from different translation teams. This approach did not solely rely on literal matches, but could also identify more distant semantic similarities. For this reason, the bipartite network graph analysis also included three texts not translated by the team of Xuanzang: the *\*Abhidharmahṛdaya* (T1550), the *\*Abhidharmahṛdayaśāstra* (1551), and the *Miśrakābhidharmahṛdaya* (1552).

For the visualization of this data, bipartite network graphs were constructed between the individual texts based on the cosine similarity of the previously detected sections of each text. These bipartite network graphs only showed those sections of a text that also had a connection with a section in another text, either precedent or subsequent. Figure 3 shows the bipartite network graphs of *Saṅgītiparyāya* / *Dharmaskandha*, *Dhātukāya*, *Prakaraṇapāda*, *Jñānaprasthāna*, and the *\*Abhidharmahṛdaya*. The order of these texts followed the results of the previous MST analysis, and the *\*Abhidharmahṛdaya* was placed at the end based on its general assessment in the research literature. This graph followed the hypothesis that the *Jñānaprasthāna* postdates the *Prakaraṇapāda*, which was in conformity with the MST analysis



but contradicted the research of Yin Shun.<sup>17</sup> In each bipartite network graph, every vertex in section *U* was connected greedily with one vertex in section *V*.

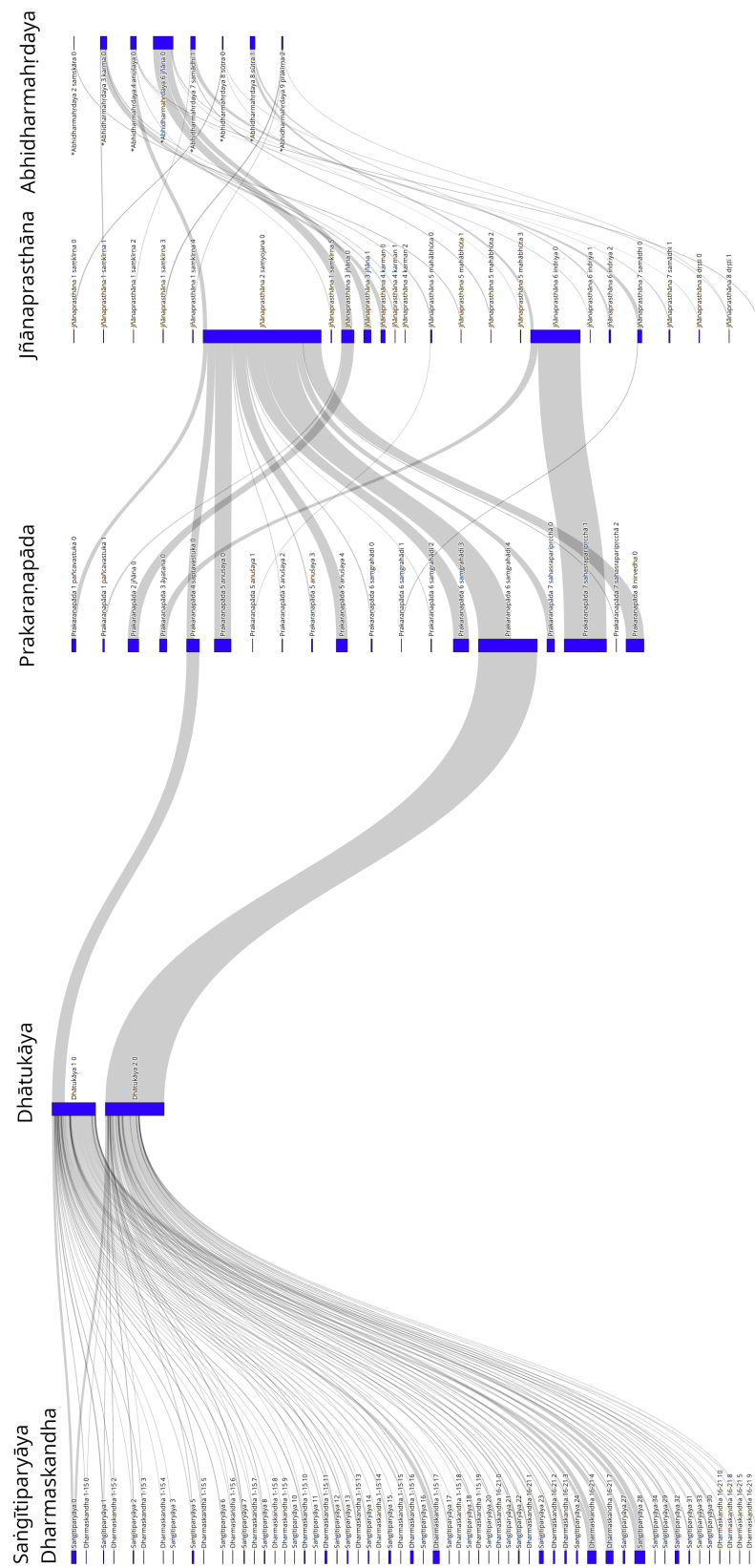


Figure 3. Bipartite network graphs of the intertextual relationships between the *Saṅgītiparyāya/Dharmaskandha*, *Dhātukāya*, *Prakaraṇapāda*, *Jñānaprasthāna*, and the *\*Abhidharmahrdaya*.

Since the relationship between the *Saṅgītiparyāya* and the *Dharmaskandha* is debated in philological research and the MST analysis placed them in the same group and in close proximity to each other, they are shown in the same part of the first bipartite network graph here, leaving the question of their temporal and doctrinal relationship open. The unsupervised text segmentation divided both texts into a comparatively large number of small sections. This aligned well with the fact that the *Saṅgītiparyāya* contains unsystematic collections of commentaries on doctrinal concepts that are scattered throughout a large number of Sūtras and have not yet been arranged into thematic blocks (Frauwallner 1995, pp. 14–15). The *Dharmaskandha*, while being more systematic than the *Saṅgītiparyāya*, is still an early example of Sarvāstivāda scholasticism in which the topical categories of later Abhidharma doctrinal discussions are present but not yet developed (Cox 1998, p. 187).

The next text, the *Dhātukāya*, is in philological research usually divided into two parts (Cox 1998, p. 208). Neither part was further divided by the text segmentation algorithm. This reflected the fact that the first part simply lists the 14 categories of 91 mental factors, while the second part gives an analysis of these categories of mental factors. Additionally, as they comprise lists of mental factors or the analysis of mental factors, both parts form complete semantic units. The edges branching out from the sections of the *Saṅgītiparyāya*/*Dharmaskandha* were roughly evenly distributed among the two parts of the *Dhātukāya*.

Upon the examination of the bipartite network graph correlating *Dhātukāya* and *Prakaraṇapāda*, it was notable that the fourth chapter of the *Prakaraṇapāda* shared an edge with the first part of the *Dhātukāya*.<sup>18</sup> Another strong edge was shown between the second part of the *Dhātukāya* and a segment of the sixth chapter of the *Prakaraṇapāda*. This could be explained by the fact that the sixth chapter of the *Prakaraṇapāda* consists of a numerical collection of factors in a similar fashion to the *Saṅgītiparyāya* and *Dharmaskandha*, including those mentioned in the second part of the *Dhātukāya*, and their recursive nature contributed to a high match score.

The structure of the *Prakaraṇapāda* stands in marked contrast to that of the preceding texts. Instead of being a mere collection of doctrinal concepts or factors, this text is an important step towards a compendium wherein the different achievements of the preceding works are arranged into coherent blocks. The unsupervised text segmentation divided the eight chapters into comparatively large semantically coherent blocks. This was not as pronounced and refined as in the later works, but clear developments can be seen when looking closely at the bipartite network graph of the *Prakaraṇapāda* and the *Jñānaprasthāna*. It becomes obvious that thematic blocks were forming in dedicated chapters such as chapter two (*jñāna*) and chapter five (*anūsaya*), which were connected with the corresponding chapters of the *Jñānaprasthāna*. Similar to the *Prakaraṇapāda*, the *Jñānaprasthāna* compiled material into eight chapters around different topics. The bipartite network graph showed how the second chapter (*jñāna*) of the *Prakaraṇapāda* connected with the third chapter (on *jnāna*) of the *Jñānaprasthāna*, as well as the fifth chapter (on *anūsaya*) with the second chapter (on *saṃyojana*). The *Jñānaprasthāna* also shows the emergence of other topical chapters such as chapter five (on *mahābhūta*), chapter six (on *indriya*), chapter seven (on *samādhi*), and chapter eight (on *dr̥ṣṭi*). The *\*Abhidharmahṛdaya* was shown as the last text in this visualization.

The temporal relationship between the *Jñānaprasthāna* and the *\*Abhidharmahṛdaya* is debated. According to Frauwallner, the *\*Abhidharmahṛdaya* is the older text,<sup>19</sup> while Fukuhara argues for the *Jñānaprasthāna* as the older text (Fukuhara Ryōgon (福原亮嚴) 1965, p. 395). Yin Shun also sees the *Jñānaprasthāna* as the older text.<sup>20</sup> Considering that the structure of the *\*Abhidharmahṛdaya* was adopted by almost all subsequent major Abhidharma works up to those of Vasubandhu and Saṅgabhadra—as is clearly illustrated in Figure 4—there are compelling reasons to position the *\*Abhidharmahṛdaya* after the *Jñānaprasthāna* in this arrangement. The unsupervised text segmentation of the *\*Abhidharmahṛdaya* showed the strong semantic coherence of the individual chapters, which were generally not broken down into more than two parts, with the exception of the ninth chapter.

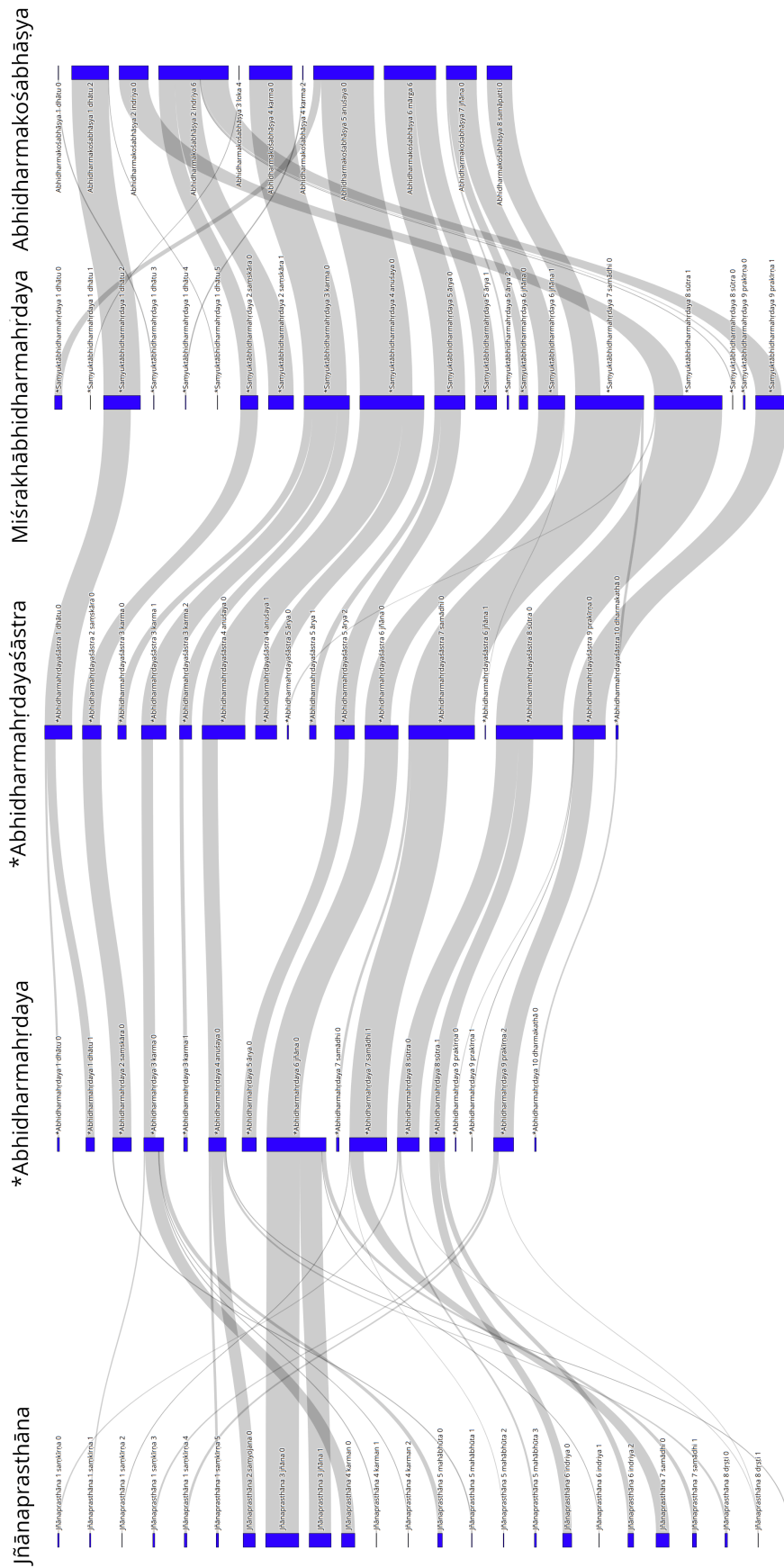


Figure 4. Bipartite network graphs of the intertextual relationships between the *Jñānaprasthāna*, *\*Abhidharmahṛdaya*, *\*Abhidharmahṛdayaśāstra*, *Miśrakābhīdharmahṛdaya*, and the *Abhidharmakośabhāṣya*.

The bipartite network graph between the *Jñānaprasthāna* and the *\*Abhidharmahṛdaya* showed how certain topical chapters of the *Jñānaprasthāna* also appear in the *\*Abhidharmahṛdaya*, while other new topical chapters were created. Noteworthy here are the edges between *Jñānaprasthāna* chapter two (*saṃyojana*) and *\*Abhidharmahṛdaya* chapter four (*anuśaya*), as well as *Jñānaprasthāna* chapter three (*jñāna*) and *\*Abhidharmahṛdaya* chapter six (on *jñāna*); *Jñānaprasthāna* chapter four (on *karman*) and *\*Abhidharmahṛdaya* chapter three (on *karman*); and *Jñānaprasthāna* chapter seven (on *samādhi*) and *\*Abhidharmahṛdaya* chapter seven (on *samādhi*).

Figure 4 depicts the relationships between the *Jñānaprasthāna*, *\*Abhidharmahṛdaya*, *\*Abhidharmahṛdayaśāstra*, *Miśrakābhīdharmahṛdaya*, and the *Abhidharmakośabhāṣya*. With the exception of the *Jñānaprasthāna*, there is broad scholarly consensus regarding the temporal order of these texts, and the depiction here followed this consensus. What is immediately clear is the shared general layout of chapters 1–7 in all works following the *\*Abhidharmahṛdaya*, with only minor variations. This is a testimony to the innovative power and influence that the *\*Abhidharmahṛdaya* exercised on Abhidharma scholasticism, an observation that has already been made by Frauwallner (Frauwallner 1995, pp. 151–52).<sup>21</sup> At this point, it is important to clarify that the *\*Abhidharmahṛdaya*'s influence on later Abhidharma works does not automatically imply that it is a younger text than the *Jñānaprasthāna*, despite the temptation to draw such a conclusion. The *\*Abhidharmahṛdaya* also shows the emergence of a thematic block centered around the fifth chapter (on *ārya*), which has no precedent in the *Jñānaprasthāna*. Erich Frauwallner argued that the fourth and fifth chapters of the *\*Abhidharmahṛdaya* contain the *Abhisamayavāda*, a new doctrine that was “the unique creation of a remarkable man” (Frauwallner 1995, p. 183). The visualization showed how this thematic block evolved in the *\*Abhidharmahṛdayaśāstra* and *Miśrakābhīdharmahṛdaya* and finally led to the sixth chapter (*mārgapudgalanirdeśa*) of the *Abhidharmakośabhāṣya*. The ninth chapter (*ātmavādapratīṣedha*) was missing in the bipartite network graph showing the *Miśrakābhīdharmahṛdaya* and the *Abhidharmakośabhāṣya*, since it was a new thematic block that had no section to connect to in the previous text. The visualization also demonstrated the incorporation of the additional content from chapter 8 and beyond of the *Miśrakābhīdharmahṛdaya* into the seven primary chapters of the *Abhidharmakośabhāṣya* (Dessein 1998, p. 273).

## 6. Conclusions

This study demonstrated that the application of network graph theory methods to textual match data produced by BuddhaNexus can provide valuable insights into the translation corpus of Xuanzang. Moreover, it can elucidate the doctrinal relationships of individual works, aligning well with observations noted in philological scholarship.

The provided visualization of Xuanzang's translation corpus offers a comprehensive, bird's-eye view of the doctrinal relationships between the works contained within. It could also act as a starting point for a more thorough investigation of specific text subsets. Several major findings emerged from this study. The arrangement of the texts in the visualization generally followed a pattern of doctrinal progression, consistent with previous research, albeit with some deviations.

The fact that the so-called “Sautrāntika” texts attributed to Vasubandhu were clustered in one community, while the Maitreya commentaries attributed to Vasubandhu appeared in a very different community, reflected previous observations about their doctrinal relationship made by Lambert Schmithausen. Moreover, the placement of two texts of the Abhidharma genre in the Sūtra community of the visualization underscored the challenge of distinguishing between Sūtra and early Abhidharma literature.

The use of the MST algorithm on a smaller subset of Abhidharmic treatises in Xuanzang's translation corpus produced a compelling depiction of their doctrinal progression. This image aligned closely with the assessments by Xuanzang's student Puguang and modern research. The unusual positioning of the *Vijñānakāya*, which at first glance appears to contradict the general consensus, can be attributed to the text's innovative nature.



In terms of applying unsupervised text segmentation for the analysis of individual sections within selected Abhidharma texts, a key observation was that earlier texts were divided into relatively more numerous, smaller sections, while later texts contained longer, coherent text sections. Corresponding to this observation was the emergence of certain topical blocks such as those centering around the concepts of knowledge (*jñāna*) and latent tendencies (*anuśaya*) in the *Prakaraṇapāda* and the *Jñānaprasthāna*. Another notable observation was the immense influence of the *Abhidharmahṛdaya* in shaping the thematic blocks. Figure 4 clearly demonstrates how later Abhidharma treatises adopted and further expanded its layout.

A critical question for future research is the quality of semantic text embedding, which impacts the effectiveness of unsupervised text segmentation and the identification of semantically similar units within a corpus. While the averaging of fastText embeddings used in this study provides a good foundation, anticipated improvements should come with the recent development of sentence representations based on deep learning models like BERT (Devlin et al. 2019). The combination of expressive sentence representations, unsupervised text segmentation, and the network-graph-based methods used in this study promises a novel approach to tracing the emergence and formation of canons and the history of ideas in vast text collections preserved in primary Buddhist source languages.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data presented in this study is accessible at <https://github.com/sebastian-nehrdich/abhidharma-analysis> (accessed on 10 July 2023).

**Conflicts of Interest:** The author declares no conflict of interest.

## Notes

<sup>1</sup> <http://buddhanexus.net> (accessed on 10 July 2023).

<sup>2</sup> See Bronkhorst (2016), especially p. 38 ff. for a discussion on the importance of Gandhāran Abhidharma for the intellectual history of South Asia.

<sup>3</sup> For a detailed evaluation of Xuanzang's translation effort from an Indological perspective see Delhey (2016).

<sup>4</sup> This assumption is shared by Robert Kritzer (Kritzer 2005, p. XXXI).

<sup>5</sup> For a full publication on the methodology of identifying the matches see Nehrdich (2020).

<sup>6</sup> For this possible objection see (Cox 1998, p. 167).

<sup>7</sup> Version number 0.9.2. Accessible at: <https://gephi.org/users/download/> (accessed on 10 July 2023).

<sup>8</sup> See (Kimura Taiken (木村泰賢) 1937, p. 36), (Cox 1998, pp. 171–72), and (Frauwallner 1995, pp. 14–21).

<sup>9</sup> As an example of this difficulty within Xuanzang's translation corpus the relationship between the *Dharmaskandha* (T1537) and the *Yuánqǐ jīng* 緣起經 (T124) can be singled out. a plethora of common passages between these two texts shows how entangled their relationship is.

<sup>10</sup> In Xuanzang's translation these are the *\*Mahāyānasaṃgrahabhāṣya* (T1597) and the *Madhyāntavibhāṅgabhāṣya* (T1600).

<sup>11</sup> Lambert Schmithausen has expressed his view of the relationship of the Maitreya commentaries and the Sautrāntika-Yogācāra works attributed to Vasubandhu at (Schmithausen 1987, p. 262) and reiterated this position at (Schmithausen 2014, p. 27). It's important to note that the methodology of this study isn't designed to identify authorship, and definitive answers to this question can't be solely derived from network graph representations of textual reuse. However, the pattern that becomes visible at this point is that these two groups are indeed clearly separated. The likely reason for this is that the texts of the "Sautrāntika" Vasubandhu are part of a closely connected web of textual reuse that connects them with the texts of the Vaibhāṣika Abhidharma tradition, especially the *\*Mahāvibhāṣā*, a situation that is not found to the same extent with regard to the Maitreya commentaries attributed to Vasubandhu. This could indicate a different educational background of the author(s) of the Sautrāntika-influenced works, who might have been more familiar with Vaibhāṣika positions than the author(s) of the Maitreya commentaries.

<sup>12</sup> The relative arrangement of the Abhidharma texts up to the *Abhidharmakośabhāṣya* and their scholarly assessment will be discussed in this study below. For the relative chronology of the Yogācāra works see (Deleanu 2006, p. 195). The composition and compilation of the *\*Mahāvibhāṣā* and the *Yogācārabhūmi* likely took many decades, maybe even centuries, and can therefore only be approximated in a rather simplistic manner here.

<sup>13</sup> Frauwallner places it after the *Dhātukāya*, see (Frauwallner 1995, p. 28). Collet Cox describes their relative chronology as uncertain (Cox 1998, p. 206). Yin Shun places it after the *Prakaraṇapāda* and even the *Jñānaprasthāna* (Yin Shun (印順) 1968, p. 170a6-11).

- <sup>14</sup> This general grouping of the texts is widely discussed in modern research, see for example (Frauwallner 1995, p. 13), (Cox 1998, p. 171) and (Eltschinger and Honjō 2015, p. 95).
- <sup>15</sup> See (Eltschinger and Honjō 2015, p. 97).
- <sup>16</sup> In the first step of this method, a semantic relatedness graph is constructed based on the cosine similarity of the averages of the vector representations of each character of an individual sentence in each text. In the second step, maximum cliques are detected, and overlapping maximum cliques are merged together as representations of semantically coherent sections of text. In addition to this, the texts have been pre-segmented into chapters/parts if that information was available in the research literature.
- <sup>17</sup> This is broadly discussed in (Yin Shun (印順) 1968, p. 148a9 ff.). Yin Shun's theory of placing the *Prakaraṇapāda* after the *Jñānaprasthāna* cannot be discussed in detail here. The MST analysis indicates general trends in the development of the literature and is not suited to give clear answers to such particular questions.
- <sup>18</sup> See (Frauwallner 1995, p. 33) and (Cox 1998, p. 218).
- <sup>19</sup> Frauwallner seems to base his assessment solely on the description in Tao-yen's foreword to Buddhavarman's *Vibhāṣa* and his application of the principle of *lectio difficilior* (Frauwallner 1995, p. 152).
- <sup>20</sup> In fact, Yin Shun argues that the \**Abhidharmahr̥daya* is based on the *Abhidharmāmṛtarasāśāstra* (T1553), which he sees as younger than the *Jñānaprasthāna* and \**Mahāvibhāṣā* (Yin Shun (印順) 1968, p. 479a9 ff.).
- <sup>21</sup> Yin Shun also observed the structural influence of the \**Abhidharmahr̥daya* on later treatises such as the *Abhidharmakośabhāṣya* but attributes the innovation to the \**Abhidharmāmṛtarasa* (甘露味論), which he claims to be older than the \**Abhidharmahr̥daya* (Yin Shun (印順) 1968, p. 493a6). If his theory should be right, the true innovation here has been achieved by the \**Abhidharmāmṛtarasa*. The question of whether the \**Abhidharmāmṛtarasa* pre- or postdates the \**Abhidharmahr̥daya*, while being very interesting, lies outside the scope of this article.

## References

- Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics* 5: 135–46.
- Brin, Sergey, and Lawrence Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems* 30: 107–17. [CrossRef]
- Bronkhorst, Johannes. 2016. Abhidharma and Indian Thinking. In *Text, History, and Philosophy: Abhidharma across Buddhist Scholastic Traditions*. Leiden: Brill, pp. 29–46. [CrossRef]
- Cox, Collett. 1998. Kaśmira: Vaibhāṣika Orthodoxy. In *Sarvāstivāda Buddhist Scholasticism*. Leiden: Brill, pp. 138–254.
- Dessein, Bart. 1998. Bactria and Gandhāra. In *Sarvāstivāda Buddhist Scholasticism*. Leiden: Brill, pp. 255–85.
- Deleanu, Florin. 2006. *The Chapter on the Mundane Path (Laukikamārga) in the Śrāvākabhūmi*. Studia Philologica Buddhica Monograph Series. Tokyo: The International Institute for Buddhist Studies.
- Delhey, Martin. 2016. From Sanskrit to Chinese and Back Again. In *Cross-Cultural Transmission of Buddhist Texts*. Edited by Dorji Wangchuk. Hamburg: Indian and Tibetan Studies, vol. 5.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for language understanding. Paper presented at the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MI, USA, June 2–7. pp. 4171–86.
- Eltschinger, Vincent, and Yoshifumi Honjō. 2015. Abhidharma. In *Brill's Encyclopedia of Buddhism. Volume One*; Leiden: Brill, pp. 88–102.
- Frauwallner, Erich. 1995. *Abhidharma Literature and the Origins of Buddhist Philosophical Systems*. New York: State University of New York Press.
- Fukuhara Ryōgon (福原亮嚴). 1965. *The Development of the Abhidharmaśāstras of the Sarvāstivāda School (有部阿毘達磨論書の発達)*. Kyoto: Nagata Bunshōdō (永田文昌堂).
- Glavaš, Goran, Federico Nanni, and Simone Paolo Ponzetto. 2016. Unsupervised Text Segmentation Using Semantic Relatedness Graphs. Paper presented at the Fifth Joint Conference on Lexical and Computational Semantics, Berlin, Germany, August 11–12; Berlin: Association for Computational Linguistics, pp. 125–30.
- Hellwig, Oliver. 2013. Googling the Rishi–Graph Based Analysis of Parallel Passages in Sanskrit Literature. Paper presented at Recent Researches in Sanskrit Computational Linguistics: Fifth International Symposium IIT Mumbai, Mumbai, India, January 4–6.
- Kimura Taiken (木村泰賢). 1937. *A Study of the Abhidharma Treatises (阿毘達磨論の研究)*. Tokyo: Meiji Shoin (明治書院).
- Kramer, Jowita. 2014. Innovation and the Role of Intertextuality in the *Pañcaskandhaka* Related Yogācāra Work. *Journal of the International Association of Buddhist Studies* 37: 281–352.
- Kritzer, Robert. 2005. *Vasubandhu and the Yogācārabhūmi*. Studia Philologica Buddhica Monograph Series. Tokyo: The International Institute for Buddhist Studies.
- Kruskal, Joseph B. 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society* 7: 48–50.
- Nehrdich, Sebastian. 2020. A Method for the Calculation of Parallel Passages for Buddhist Chinese Sources Based on Million-scale Nearest Neighbor Search. *Journal of the Japanese Association for Digital Humanities* 5: 132–53. [CrossRef]

- Nicoll-Johnson, Evan. 2018. Drawing Out the Essentials: Historiographic Annotation as a Textual Network. *Journal of Chinese Literature and Culture* 5: 214–49. [CrossRef] [PubMed]
- Sakuma, Hidenori, and Ulrich Timme Kragh. 2013. Remarks on the Lineage of Indian Masters of the Yogācāra School. In *The Foundation for Yoga Practitioners*. Edited by Ulrich Timme Kragh. Harvard Oriental Series. Cambridge, MA: Harvard University Press, vol. 75. [CrossRef]
- Schmithausen, Lambert. 1967. Sautrāntika-Voraussetzungen in Viṃśatikā und Trimśikā. *Wiener Zeitschrift für die Kunde Süd- und Ostasiens* XI: 109–250.
- Schmithausen, Lambert. 1987. *Ālayavijñāna*. Tokyo: The International Institute for Buddhist Studies.
- Schmithausen, Lambert. 2014. *The Genesis of Yogācāra-Vijñānavāda*. Tokyo: The International Institute for Buddhist Studies of the ICPBS.
- Tharsen, Jeffrey, and Clovis Gladstone. 2020. Using Philologic For Digital Textual and Intertextual Analyses of the Twenty-Four Chinese Histories 二十四史. *Journal of Chinese History* 中國史學刊 4: 558–63.
- Vierthaler, Paul, and Mees Gelein. 2019. A BLAST-based, Language-agnostic Text Reuse Algorithm with a MARKUS Implementation and Sequence Alignment Optimized for Large Chinese Corpora. *Journal of Cultural Analytics* 4: 1–25. [CrossRef]
- Yin Shun (印順). 1968. *A Study on the Main Works and Masters of Abhidharma, with Special Focus on the Sarvāstivāda School* (一切有部為主的論書與論師之研究). CBETA 2023.Q1, Y36, No. 34. Available online: <https://cbetaonline.dila.edu.tw/Y0034> (accessed on 10 July 2023). [CrossRef] [PubMed]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.