Intertextual Du Fu: A Study of Citation Network Analysis

Kiho Sung, Jamie Jungmin Yoo, Changhee Lee
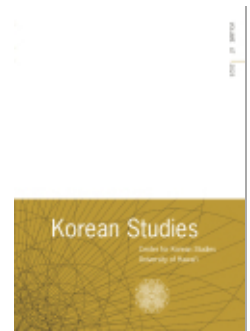
Intertextual Du Fu: A Study of Citation Network Analysis

Kiho Sung, Jamie Jungmin Yoo, and Changhee Lee

How does a reader give a text meaning? In what ways do we understand the operations of encounters between a text and a reader? Focusing on the materiality of reading, this study aims to understand how anonymous readers in late Chosŏn Korea read the poetry of Du Fu, a renowned literary canon from China. To identify the meaning of the texts as constituted by the readers, we look at both the texts and the readers' practices by identifying the readers' prior knowledge of the text that has been embedded and coded in their reading notes, and by analyzing the relationship between the notes and the main body of poetry. Through this analysis, this paper shows that the reading of texts was performed through constant interactions with interpretive traditions and cultural legacies. Through their practices of reading, consequently, they reveal which communities of interpretation they distinctively belong to. To identify the invisible patterns of the exegetical traditions in their reading practices, we particularly apply methods in digital humanities, such as citation network analysis, which is an effective tool to recognize the structure of relationships among the notes, poems, and many other factors of the texts.

Keywords: Du Fu, citation network analysis, intertextuality, reading and reception, late Chosŏn literature

Kiho Sung is a Ph.D. candidate in the Department of Sociology at Yonsei University, South Korea (kihosung@yonsei.ac.kr).
Jamie Jungmin Yoo is an assistant professor in the College of Liberal Arts at Yonsei University, South Korea (crimson.yoo@gmail.com).
Changhee Lee is a Ph.D. candidate in Korean literature at Korea University, South Korea (po77777@korea.ac.kr).

Introduction

" . . . readers are travelers; they move across lands belonging to someone else,
like nomads poaching their way across fields they did not write,
despoiling the wealth of Egypt to enjoy it themselves."
(Michel de Certeau, "Reading as poaching" in *The Practice of Everyday Life*)

How does a reader give a text meaning? In what ways do we understand the operations of encounters between a text and a reader? Reading is an active performance that governs ways of comprehending a text. Since the practices are not already inscribed in the text, the text acquires its meaning only through its readers. Therefore, there is inevitably a breach between the meaning assigned to the text and the interpretation readers make of it.[1] In this sense, studying or writing a literary history means that we trace the breaches between the piles of physical books and the meanings given to them. The meanings of texts have not been fixed, but the interactions between texts and readers shape the dynamic literary geography of history.

The act of reading itself is ephemeral, but readers often leave behind vestiges of their performances. They often take notes, reorganize the structure, or put a text into an anthology. These physical traces provide us with the forms and circumstances of how their readers received the texts. Focusing on the materiality of reading, this study aims to understand how anonymous readers in late Chosŏn Korea read the poetry of Du Fu, a renowned literary canon from China. To identify the meaning of the texts as constituted by the readers, we will look at both the texts and the readers' practices (1) by identifying the readers' prior knowledge of the text that has been embedded and coded in the headnotes (*Tuju*, 頭註), found in the margins of each page, and (2) by analyzing the relationship between the headnotes and the main body of poetry. Through such analysis, this paper aims to show that the reading of texts by individual readers is performed through constant interactions with interpretive traditions and cultural legacies. Noticed or unnoticed, readers have been influenced by certain traditions of knowledge, and through their practices of reading, consequently, they reveal which communities of interpretation they distinctively belong to. Each interpretive community is structured by shared sets of interests and manners of reading.

When readers in Chosŏn Korea encountered Du Fu and then published and institutionalized by the authorities, in what kinds of intellectual legacy were they participating? To answer this question, this research particularly experimented with methods in digital humanities, such as a citation network analysis. It is a useful analytical tool for

recognizing the structure of relationships among the notes and poems. More importantly, we expect these analytical methods to be particularly effective in delineating the "invisible" patterns of the exegetical traditions. Although the hidden clusters are rarely noticeable by descriptive research methods, the invisible interpretive communities are often proved to be essential agencies that configure the intellectual norms and conventions underlying the reading practices.

Shaping of a Text's Meaning: Intertextuality

A copy of Du Fu's poetry anthology, which was found at the rare book depository at Yonsei University (*Du Fu-Y*, hereafter), vividly demonstrates how the encounter between the "world of the text" and the "world of the reader" operated in late Chosŏn. Throughout the dynasty, the renowned literary canon from China was circulated in multiple shapes, depending on the publishers' editorial intentions.[2] According to the cover page, this particular version of *Du Fu-Y* was printed in woodblock, and the publisher was located in the Tae-gu area. The exact date of printing is unknown, but considering that the same title of the original version was initially compiled and printed with metal-movable type during king Chŏngjo's reign (正祖,
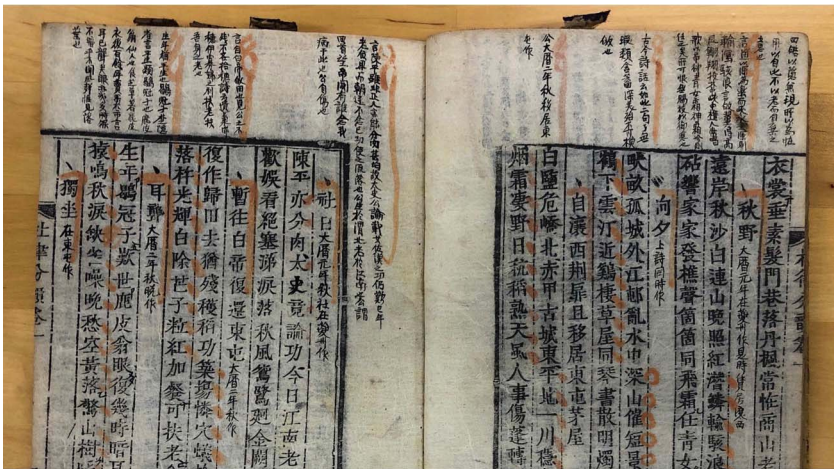

Fig. 1. An example of headnotes of Tuyul punun. Each page has this kind of hand-written headnotes. This copy is preserved in the library of Yonsei University. A color version of this figure is available in the online article.

r. 1776–1800), *Du Fu-Y* was undoubtedly printed after the eighteenth century.

What attracted me most in this copy was the "headnotes" in the margin, which were meticulously handwritten by anonymous readers. Each page shows distinctive marks and notes left by anonymous readers. At the top of all the pages are hand-copied notes, and side dots accompany selected lines of poems. Were they the readers' own personal notes? Did they consult someone else's references and selected some of their exegeses? To identify the nature of the notes, we conducted extensive research: (1) We typed all headnotes and translated them into Korean, (2) selected major commentaries circulated in Chosŏn, particularly the *Jiu jia ji zhu Du shi* (九家集注杜詩, hereafter *Jiujiaji*) from China, which provided the foundational exegesis tradition when the *Ch'anju Pullyu Tushi* (纂註分類杜詩, hereafter *Ch'anju*) was compiled during King Sejong's reign in Chosŏn (世宗, r. 1418–1450), (3) compared the headnotes of *Du Fu-Y* and the commentaries from *Jiujiaji* and *Ch'anju*, and (4) using multiple online searches, we also checked whether there were any quotes outside of *Du Fu-Y*.[3]

Our research reveals that the anonymous readers of *Du Fu-Y* had studied *Jiujiaji* or *Ch'anju*. Since the reprinted version of *Jiujiaji* was not well known to general readers in Chosŏn, a version of *Ch'anju*, either a copy directly from the fifteenth century or a reprint from a later period, would have been used for their reading references. The close associations between *Du Fu-Y* (18th C) and *Ch'anju* (15th C) indicate how the readers approached the poems of Du Fu. As seen from the physical shapes of *Du Fu-Y*, the primary texts for readers do not contain any commentaries. The only information they could get from the format was the "rhymes" of each poem since the poems were arranged by the rhyming categories (*unmok* 韻目) to which they belonged. To shape the meaning of each poem, they referred to the conventional references transmitted from the fifteenth century or earlier.

These intertextual connections between the texts across the centuries are deliberate reading strategies of the readers. Using the readers' prior knowledge or understanding of the references, they made a continual dialogue between the poems and the commentaries. Therefore, the fragments of the headnotes can be understood as responses to both *Du Fu-Y* from the eighteenth century and *Ch'anju* from the fifteenth century. The meanings of Du Fu's poetry were not inscribed in the woodblocks; instead, they can be found somewhere in the transactions of reading between the texts and the references.

Invisible Communities: Analytical Methods

The headnotes of *Du Fu-Y* state that reading is not a solely abstract or solitary operation but is an interlocking form of relationship between the text and readers and between the perceptions in the eighteenth century and the accumulative interpretations from the fifteenth century. Table 1 summarizes how many times the annotators were cited in the headnotes, in descending order of frequency. As can be seen in Table 1, our preliminary research unfolds that the readers of *Du Fu-Y* depended on the annotators quoted in *Ch'anju*. Figure 2 is a bar graph visualization for those who cited 10 or more times in the headnotes.

Table 1. Full List of Annotators Cited in the Headnotes of *Du Fu-Y*

No.	Name (Chinese)	Name (English)	Frequency of Citations
1	蔡夢弼	Cai Mengbi (Cai MB)	385
2	趙次公	Zhao Cigong (Zhao CG)	344
3	王洙	Wang Zhu (Wang Z)	317
4	劉辰翁	Liu Chenweng (Liu CW)	103
5	師民瞻	Shi Minzhan (Shi MZ)	100
6	黃鶴	Huang He (Huang H)	71
7	黃希	Huang Xi (Huang X)	42
8	鄭卬	Zheng Ang (Zheng A)	17
9	杜修可	Du Xiuke (Du XK)	15
10	杜甫	Du Fu (Du F)	12
11	薛蒼舒	Xue Cangshu (Xue CS)	12
12	補註	Buzhu (Buzhu)	10
13	鮑彪	Bao Biao (Bao B)	7
14	蘇子瞻	Su Zizhan (Su ZZ)	6
15	杜田	Du Tian (Du T)	5
16	魯訔	Lu Yin (Lu Y)	4
17	杜定功	Du Dinggong (Du DG)	3
18	黃魯直	Huang Luzhi (Huang LZ)	3
19	沈括	Shen Kuo (Shen K)	3
20	王禹稱	Wang Yucheng (Wang YC)	3
21	張詠	Zhang Yong (Zhang Y)	3
22	洪芻	Hong Chu (Hong C)	2
23	呂本中	Lu Benzhong (Lu BZ)	2
24	秦觀	Qin Guan (Qin G)	2
25	孫	Sun (Sun)	2

(Continued)

Table 1. (*Continued*)

No.	Name (Chinese)	Name (English)	Frequency of Citations
26	唐庚	Tang Geng (Tang G)	2
27	王彥輔	Wang Yanfu (Wang YF)	2
28	謝逸	Xie Yi (Xie Y)	2
29	晏殊	Yan Shu (Yan S)	2
30	張耒	Zhang Lei (Zhang L)	2
31	張孝祥	Zhang Xiaoxiang (Zhang XX)	2
32	晁	Chao (Chao)	1
33	程	Cheng (Cheng)	1
34	洪覺範	Hong Juefan (Hong JF)	1
35	胡仔	Hu Zi (Hu Z)	1
36	華	Hua (Hua)	1
37	李泰伯	Li taibo (Li TB)	1
38	林敏功	Lin Mingong (Lin MG)	1
39	呂大防	Lu Dafang (Lu DF)	1
40	歐陽脩	Qu Yangxiu (Qu YX)	1
41	師古	Shi Gu (Shi G)	1
42	石敏若	Shi Minruo (Shi MR)	1
43	司馬光	Sima Guang (Sima G)	1
44	宋祁	Song Qi (Song Q)	1
45	孫莘老	Sun Shenlao (Sun SL)	1
46	王十朋	Wang Shipeng (Wang SP)	1
47	謝邁	Xie Mai (Xie M)	1
48	尹洙	Yin Zhu (Yin Z)	1
49	張子韶	Zhang Zishao (Zhang ZS)	1
50	趙傁	Zhao Sou (Zhao S)	1

The readers relied heavily on three annotators: Cai Mengbi (蔡夢弼, hereafter Cai MB), Zhao Cigong (趙次公, hereafter Zhao CG), and Wang Zhu (王洙, hereafter Wang Z). How can the similarities or differences in reading practices of choosing these people among others be described? What are the associations among the annotators or between the annotator and the poems? To identify such "invisible" clusters of interpretation from this quantitative result, this paper proposes the use of "citation network analysis." Traditionally, in the field of bibliometrics, which is the statistical analysis of books, articles, and other publications, citation network analysis is often used to identify the structure of the research community.[4] Since the act of "citing" is believed to be a reliable valid representation of communication in academia, the citation network is considered a vital indicator of how to measure associations among authors and references.[5]

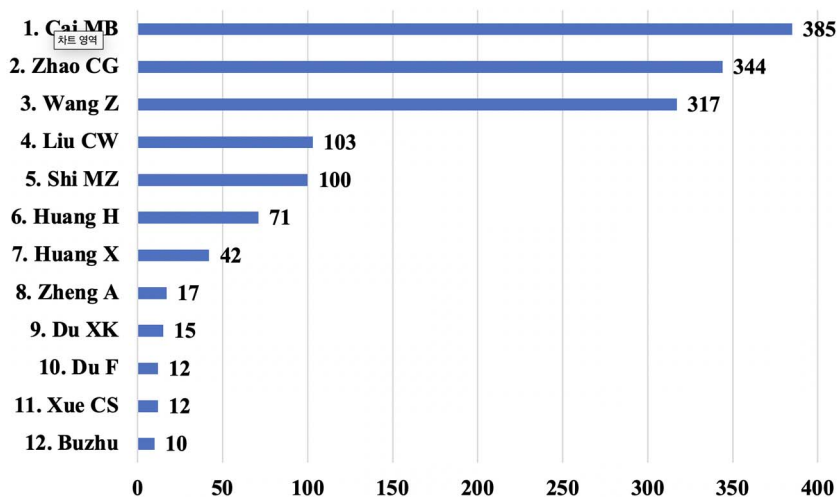|	
1. Cai MB	385
2. Zhao CG	344
3. Wang Z	317
4. Liu CW	103
5. Shi MZ	100
6. Huang H	71
7. Huang X	42
8. Zheng A	17
9. Du XK	15
10. Du F	12
11. Xue CS	12
12. Buzhu	10

Fig. 2. Distribution of citations among the selected annotators (cited 10 times or greater). A color version of this figure is available in the online article.

Interestingly, citation is a powerful method of identifying "invisible colleges," which means informal research networks referring to one another in their academic publications without formal organizational ties that link scholars.[6] Since the 1960s, the notion of invisible colleges has been the foundation of citation network analysis, and subsequent studies have developed a mathematical basis for it.[7]

When research is concerned with citation networks, three bibliometric methods are commonly used: (1) citation analysis, (2) co-citation analysis, and (3) bibliographic coupling.[8] As shown in Table 1 and Figure 2, citation analysis is a set of descriptive techniques, such as a summary of citation frequency. Co-citation analysis measures similarities between authors, journals, or documents. A co-citation exists if the writers list two authors or documents in the same bibliography.[9] Bibliographic coupling uses the number of common references in two documents to measure similarity. This paper particularly applies two methods, co-citation analysis and bibliographic coupling, since they are useful for identifying the structure of the research community based on similarities between objects of analysis. In this process, the choice of the similarity measure and how it is assessed are central issues in the analytical dimension.[10] Studies aiming to delineate the structure of the research community widely use co-citation analysis, and as a dominant analytical strategy to date, it is more popular than bibliographic coupling.[11] In addition, there have been methodological

reasons which made the bibliographic coupling relatively less popular. As citation practices change over time, bibliographic coupling should be performed within a limited time span.[12] For this reason, the bibliographic coupling technique has advantages over the co-citation analysis method when it comes to the short period of time span.[13] Acknowledging the strengths and weaknesses of each method, this research applies both co-citation and bibliographic coupling in order to identify the structure of the relationship between the annotators and between the poems.

In the field of digital humanities, there are a few remarkable examples of the application of general citation network analytical methods. In his seminal work, for example, Brughmans attempted to reveal how network analysis has been implemented in archaeology. The citation network of 33,556 publications and 42,993 citations was redefined and split into co-citation and bibliographic coupling networks. These two networks are represented by chronological visualization and main path analysis to track the use of network analysis among archaeologists.[14] Second, in order to understand the citation patterns of ancient texts, Romanello takes advantage of natural language processing and the construction of citation networks.[15] Similarly, Nicoll-Johnson conducted an impressive study using annotations to ancient titles that appeared in the chapters of *Sanguozhi* (三國志) and *Shishuo Xinyu* (世說新語).[16]

Our research is expected to contribute to the digital humanities by applying citation network analysis to classical texts from traditional Korea. Particularly, we define "similarity measures" based on co-citation and bibliographic coupling analysis. By implementing it, we will map to find the structure of invisible communities in the context of the defined data. These analyses assume that if two items are likely to be cited (or likely to cite each other), their contents are expected to be related. In the next section, the construction of citation network data and choice of analytical strategy will be further explained.

Transformations: Text to Data

The headnotes in *Du Fu-Y* were the materials used to construct the citation network. There are 777 poems in *Du Fu-Y*, all classified by the rhymes. The headnotes in *Du Fu-Y* were written by unknown readers, and they cite important annotators and annotations before each text so as to help correctly interpret the Du Fu poems. In this study, *Ch'anju* is used to compare the poem classification systems. *Ch'anju* contains 1,470 poems in

total. One distinct feature of *Ch'anju* is its classification system of the poems which was based on the themes. There are 773 poems in both *Du Fu-Y* and *Ch'anju*, so it was possible to match the theme classification in the latter to the rhyme classification of the former.[17]

Once the poems and annotators were matched, we defined two sets of nodes, one for poems and the other for annotators, so that we made a two-mode network structure. The links were defined between different sets of nodes. Based on the two-mode network of poems and annotators, we simultaneously defined the co-citation of annotators and the bibliographic coupling of poems, as illustrated in Figure 3. First, we entered the matching pairs between each poem's headnote and cited annotator(s) in a table of 777 rows (poems) and 50 columns (annotators). In Figure 3(a), the cells filled with circles indicate the patterns how the annotators were cited in the
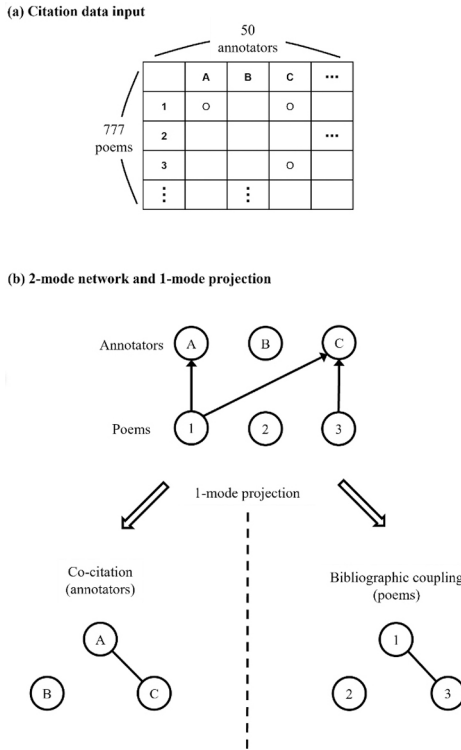

Fig. 3. Citation data input and network data construction. A color version of this figure is available in the online article.

headnotes. The illustration represents, for example, how annotators A and B were cited in the headnote of poem 1, while annotator C was only cited in the headnote of poem 3. Second, the table was converted into two-mode network data. One set of nodes representing the annotators, and the other consists of the poems. Their edges link the poems and the annotators by sharing the same headnotes.

Figure 3(b) shows the network data built from the table in Figure 3(a). Poem 1 has two links (citations) to annotators A and B, while poem 3 has just one link to annotator C. Third, we projected the two-mode network to a one-mode, one mode for the annotators and for the poems, respectively. If two annotators were cited together in a single headnote, for instance, they were considered to have one co-citation. This procedure has been repeated whenever two annotators were co-cited in the headnotes. In Figure 3(b), for instance, annotators A and B were co-cited in the poem 1's headnote, so they have a link in the co-citation network, which was projected from the two-mode network. This step has been repeated for all the possible annotator pairs, yielding the 50 × 50 annotator matrix and the corresponding network data.

In order to build a bibliographic coupling network, this study has implemented a similar procedure for the poems' one-mode projection. As seen in Figure 3(b), for example, the headnotes of poems 2 and 3 have a common citation from annotator C, which means that the poems share a link in the bibliographic coupling network. As a result, we constructed a 777 × 777 poem matrix and the corresponding network data. Finally, we produced two distinct matrices and their corresponding networks, which are co-citation of annotators and bibliographic coupling of poems. Our analytical methods are as follows.

Similarity Measure and Hierarchical Clustering

After building the two distinct one-mode networks, we conducted similarity measures based on the co-citations and the bibliographic coupling so as to identify "latent clusters" based on network structure. As discussed earlier, the assessment of similarity measures would be the key element in citation network analysis. For the similarity measures, Pearson's r (correlation coefficient), the Jaccard Index, and Salton's cosine has been widely accepted.[18] This research adopted the CoCit-Score which has been proposed by Gmür (2003), for our calculations of co-citations and bibliographic couplings[19]:

$$\text{CoCit}_{AB} = \frac{(\text{co-citation}_{AB})^2}{\text{minimum}(\text{citation}_A ; \text{citation}_B) \times \text{mean}(\text{citation}_A ; \text{citation}_B)}$$

$$= \frac{\text{co-citation}_{AB}}{\text{minimum}(\text{citation}_A ; \text{citation}_B)} \times \frac{\text{co-citation}_{AB}}{\text{mean}(\text{citation}_A ; \text{citation}_B)}$$

Gmür's CoCit-Score is a product of two ratios. The first is the ratio of co-citation counts to minimum number of citation counts in the documents A and B. The second is the ratio of co-citation counts to average citation counts in the documents A and B. Each ratio takes any value between 0 and 1, and their product also takes any value between 0 and 1 as a result. Each ratio could be examined as an independent co-citation index, but the first tends to emphasize links around highly cited core references, which leads to overrating less-cited references as closer to frequently cited ones. The second part is known to be sensitive to the citation count variance in the given data, resulting in underestimation of important co-citations between references of different citation counts and over-estimation of commonly cited references. Consequently, it was suggested to use the product of the two ratios to reduce the influence of the citation between two co-cited documents. Compared to other co-citation analysis methods, such as factor analysis or Pearson's *r*, Gmür's CoCit-Score has no obvious restrictions as such.[20]

The resulting score indicates that if the value is closer to 0, the similarity between two documents is low, while if the value is closer to 1, their similarity is high. The calculation of the CoCit-Score was equally implemented for both the annotator and poem networks.

Using CoCit-Scores as a similarity measure, we conducted hierarchical clustering, one of the clustering algorithms. We chose "Ward's linkage," one of the most commonly used methods in hierarchical cluster analysis. Ward's minimum variance method computes the error sum of squares by merging two clusters into one until it minimizes the total within-cluster variance.[21] Although hierarchical clustering provides distance between clusters and merges the sequences, the analysis itself does not provide the "optimum" number of clusters.[22] After assessing different numbers of clusters for the annotator and poem networks, respectively, we chose only the annotator network to present the results of the hierarchical clustering analysis. The Scikit-network package, developed for graph analysis in Python, was used for our hierarchical clustering implementation.[23]

The general rules for our network visualization were as follows. First, the size of each node corresponds to the number of citations if the poem network was drawn or to the number of being cited if the annotator network was drawn. Second, the edge thickness and the color intensity are correlated to the similarity results. Third, the Fruchterman-Reingold layout, one of the spring-embedding (or, equivalently, force-directed) layouts, is applied to both the annotator and the poem networks. This layout tends to place nodes with a higher degree and/or edge weight in the center and then sets pair of nodes with higher edge weight closer each other. In the poem network, we represent two node attributes as colors: the rhyme classification from *Du Fu-Y* and the clusters from the hierarchical clustering. In the annotator network, we assigned only one type of node attribute a color-clusters from the hierarchical clustering.[24]

The annotator network's hierarchical clustering are visualized as a dendrogram. Dendrogram is a standard option for visualizing hierarchical clustering results, as they effectively draw both distance between nodes and differentiated clusters. A dendrogram was thus only useful when displaying the results of the annotator network since the number of nodes was relatively small enough to be plotted. In turn, the number of nodes in the poem network was 777, and dendrograms lose their usefulness when drawn from relatively large networks.[25]

Analytical Results

Table 2 shows the descriptive statistics of the network data. Originally, the number of nodes in the two-mode network was the sum of nodes, which was 827 (=777 + 50). There were 1,504 headnote citations that contain the information of the annotators that equal the number of edges in the two-mode network. The annotators are only be "cited" which correspond to "indegree." On the other hand, the poems could have citations in the form of headnotes, which make both indegree and outdegree. On average, there are 1.94 annotators per each poem, while each annotator is cited 30.10 times. Note that variance of citation is relatively low (range: 0–5), while that of being cited is very large (range: 1–385). After projecting to two distinct one-mode networks, the annotator network has 50 nodes and 145 edges, while the poem network has 777 nodes and 153,426 edges. The CoCit-Score calculation results in these two networks also show clear differences in the mean statistics. The mean CoCit-Score of the annotator network is 0.04, which is close to 0, and the result implies that similarity between

Table 2. Descriptive Statistics on Network Data

Networks	Number of Nodes	Number of Edges	Frequency of Citations				Edge Weight (CoCit-Score)			
			Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
2-mode network	827	1,504	–	–	–	–	–	–	–	–
Annotators	50	–	30.10	84.56	1	385	–	–	–	–
Poems	777	–	1.94	1.08	0	5	–	–	–	–
1-mode network										
Annotator network (co-citation)	50	145	–	–	–	–	0.04	0.13	<0.001	1
Poem network (bibliographic coupling)	777	153,426	–	–	–	–	0.45	0.26	0.04	1

annotators is very low. In turn, the mean CoCit-Score of the poem network is 0.45, which is far greater than that of the annotator network.

Figures 4(a) and (b) are the network visualizations of the poem network. In both figures, node size corresponds to number of annotator(s) cited in each headnote, and the edge thickness and color intensity increase according to the greater CoCit-Score between nodes. In Figures 4(a) and (b), the bigger nodes are placed in the center, and the nodes connected by higher CoCit-Score edges are closer together, having thicker and intense edge colors. The difference between Figures 4(a) and (b) is in node color. In Figure 4(a), for instance, the node colors represent the rhyme classification of the poems in *Du Fu-Y*, with 28 different rhymes in total. In contrast, the node colors in Figure 4(b) are differentiated by the cluster derived from the hierarchical clustering. The visual assessment clearly shows that the clusters identified from the similarity measure are "not" related to the rhyme classification of each poem. In Figure 4(b), the nodes in the same cluster are close to each other, forming distinct locations in the network visualization. However, the nodes in the same rhyme classification are not close to each other but scattered without order as in Figure 4(a). Thus, we can conclude that the "informal and invisible clusters" presented in Figure 4(b) would be identified based on the bibliographic coupling of the headnotes.[26]

Figures 5(a) and (b) form zoomed-in snapshots of Figure 4(b). The numbers represent the listing orders of the poems in *Du Fu-Y*. The poem number 754 and the 763, for example, show that the poems cited similar annotators. Figures 6(a) and (b) are made by selecting sets of nodes and edges from Figure 4. Note that the node sizes, colors, edge thicknesses, and colors are all presented in the same way as in the Figure 4(a). Figure 6(a) is a visualization of the poems classified to theme, "Expressing my thoughts (述懷)," in *Ch'anju*. Figure 6(b) is for the poems classified into the theme, "Traveling (紀行)," in *Ch'anju*. For these implementations, we can discover there is no noticeable pattern among the rhyme families. It would mean that the compilers of *Du Fu-Y* in the eighteenth century did not consider the thematic arrangements, which had been a major compiling rule for *Ch'anju* in the fifteenth century.

Figure 7 is a dendrogram visualization for the hierarchical clustering of the annotator co-citation network. Four clusters are selected and then presented in the different colors and different patterns of lines. The names of the annotators are on the right side of Figure 7. The numbers in front of the names correspond to the citation frequency. There are two large clusters, which are in blue (dotted lines) and red (long-dashed lines),
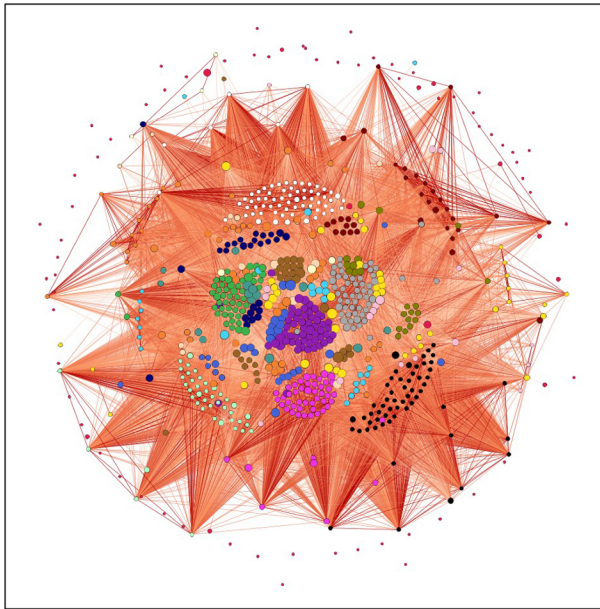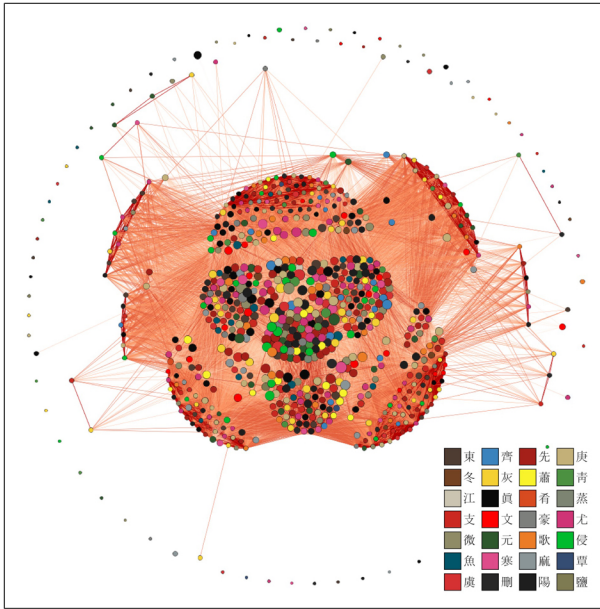
東 齊 先 庚
冬 灰 蕭 青
江 眞 肴 蒸
支 文 豪 尤
微 元 歌 侵
魚 寒 麻 覃
虞 刪 陽 鹽

Fig. 4. Poem 1-mode network with node size corresponding number of citations, edge weighted by CoCit-Score, and node colors differentiated by (a) rhyme classification from Tuyul punun and (b) result of hierarchical clustering analysis (20 clusters). A color version of this figure is available in the online article.
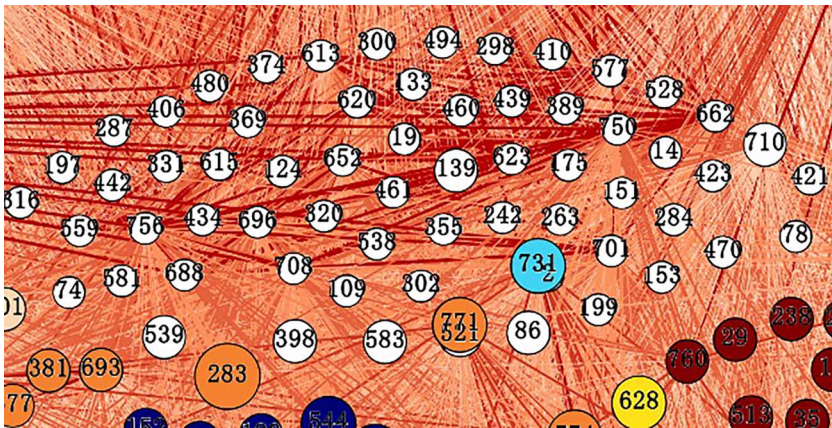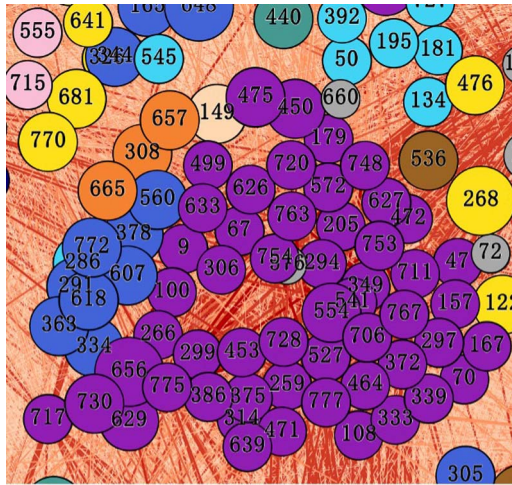
Fig. 5. Poem 1-mode network with node size corresponding number of citations, edge weighted by CoCit-Score, node colors differentiated by result of hierarchical clustering analysis (20 clusters), labeled with poem ID numbers. Zoomed in the (a) center of Figure 4 (b) and (b) upper area of Figure 4(b). A color version of this figure is available in the online article.

respectively. The three most frequently cited annotators (*Cai MB*, *Zhao CG*, and *Wang Z*) appear together in the red cluster (long-dashed lines). As the length of the tree between different nodes indicates distance, it also shows that *Cai MB*, *Zhao CG*, and *Wang Z* are likely to be cited together in the headnotes.[27]
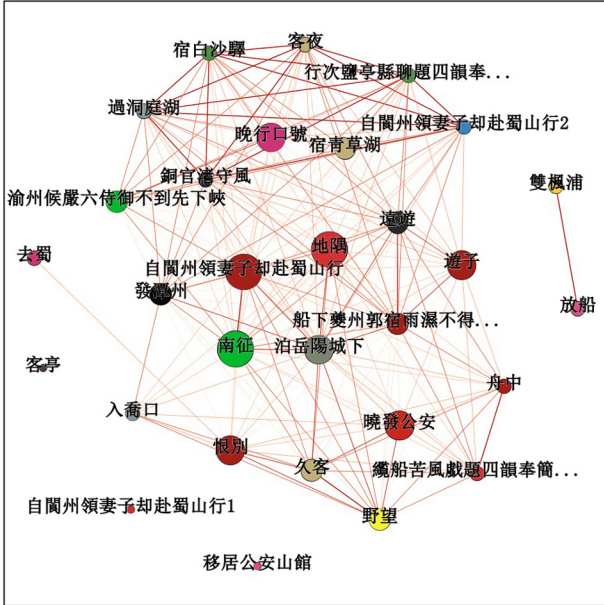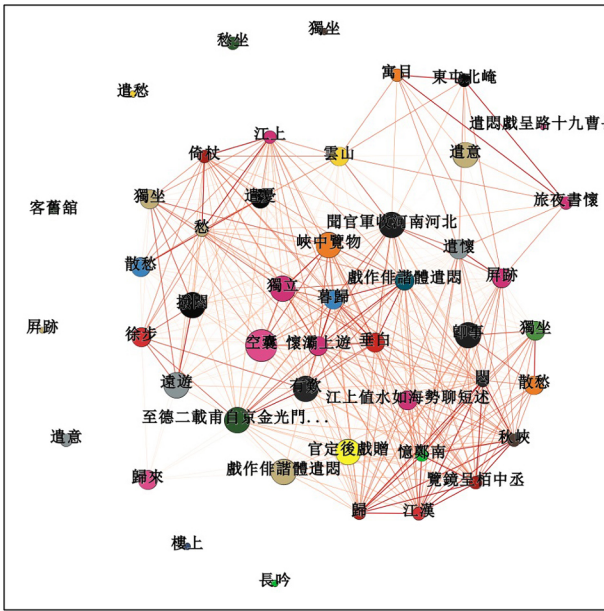
Fig. 6. Poem 1-mode network with node size corresponding number of citations, edge weighted by CoCit-Score, and node colors differentiated by rhyme classification from Tuyul punun. Poems classified as theme (a) 述懷 and (b) 紀行 in Ch'anju pullyudusi are presented with titles. A color version of this figure is available in the online article.

Figure 8 is a network visualization of the annotation network. Node size corresponds to number of "being cited." The edge thickness and color intensity increase if the CoCit-Score becomes greater. The node colors indicate different clusters found in the hierarchical clustering analysis.
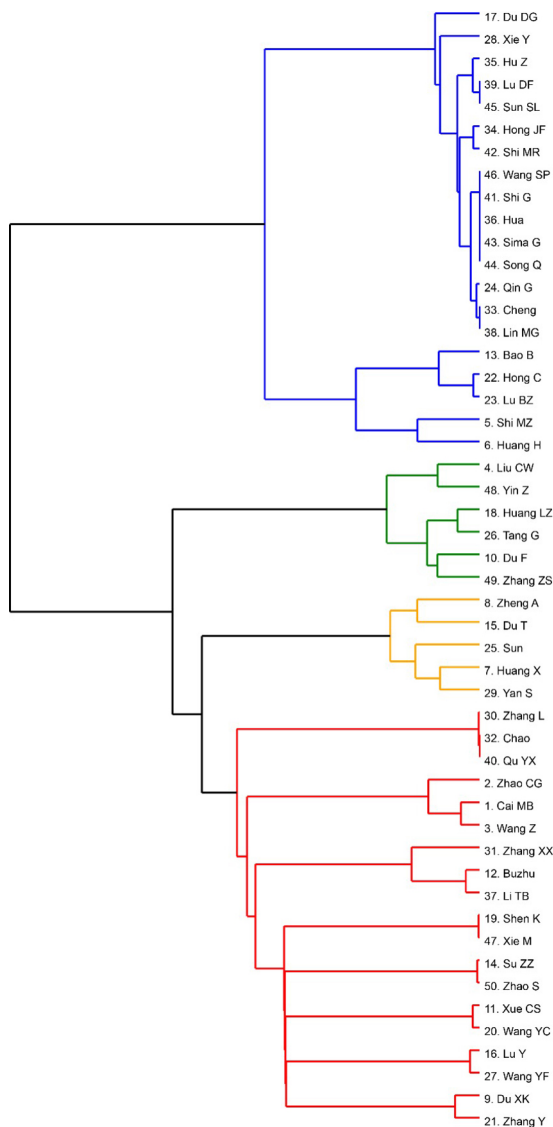

Fig. 7. Dendrogram of hierarchical clustering result on annotator co-citation network. Annotators are grouped into 4 clusters that drawn in different colors. A color version of this figure is available in the online article.

What is evident in Figure 8 is that the three small-sized nodes for *Zhang L*, *Chao*, and *Qu YX* have very thick edges between them. This occurs because these three annotators are cited only once or twice, but they are cited in common poems and the resultant CoCit-Scores among them are very high. As they are not as frequently cited as the most frequently cited annotators, this feature may overrepresent the importance of these three nodes. For this reason, we also present Figure 9, which shows all annotators with at least 3 citations only.

The four clusters derived from the hierarchical clustering of the annotator co-citation network would indicate the latent "knowledge base" groups identified by similarity. It is unclear whether the writers of the headnotes were aware of this classification. Regardless of the writers' consciousness, the result provides an "informal" classification of annotators. In this regard, the results of the co-citation networks' hierarchical clustering provide an analytical model to identify the knowledge bases of persons who tried to understand a canonical literary anthology of the time.
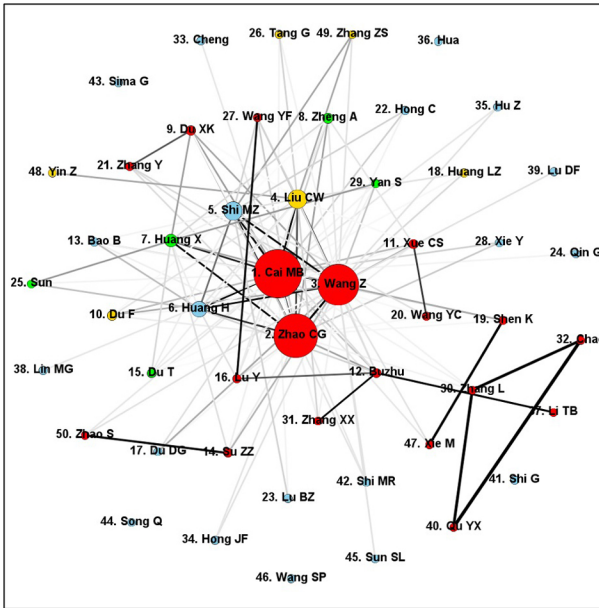
Fig. 8. Annotator 1-mode network with node size corresponding number of citations, node colors differentiated by result of hierarchical clustering analysis (4 clusters), edge weighted by CoCit-Score, and node labeled as annotators' names. A color version of this figure is available in the online article.
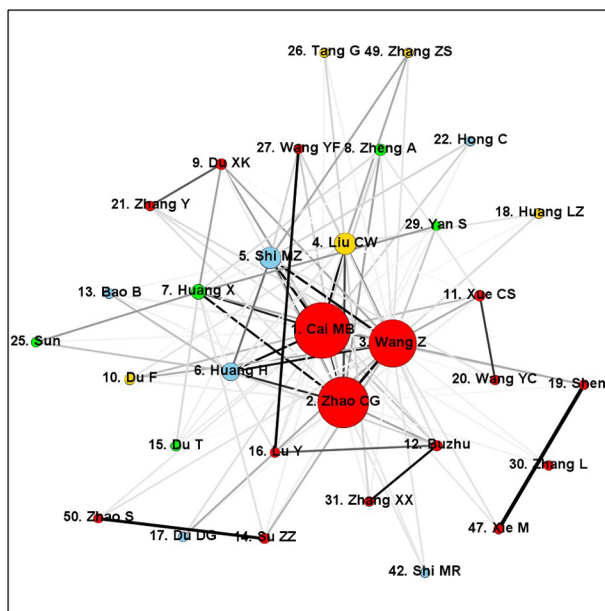
Fig. 9. Annotator 1-mode network with node size corresponding number of citations, node colors differentiated by result of hierarchical clustering analysis (4 clusters), edge weighted by CoCit-Score, and node labeled as annotators' names (only nodes of at least 3 degrees are displayed). A color version of this figure is available in the online article.

Conclusion

In *Desire in Language: A Semiotic Approach to Literature and Art*, Julia Kristeva argues that the meaning of a text does not reside in the text itself but is produced by the reader. Du Fu is a historical figure from Tang dynasty China, and his poetic works have been hailed as one of the greatest literary canons by many literary historians around the world. However, a copy of Du Fu's poetry found in a corner of a rare book archive seems to support the intertextual view of literature proposed by Kristeva. The significance of Du Fu's poetry has been made through complex networks of texts, and its meaning was not directly transferred from the Chinese poet and his works to readers in Korea. Instead, his poetry was mediated through the author, readers, and numerous other texts. Local readers engaged in specific acts and in different places in Korea imposed new meanings on the text.

This paper explored the process through which Du Fu's poetry was contextualized by the local in eighteenth-century Korea. The notes from

anonymous readers in *Du Fu-Y* were originally written as a result of personal reading activities, but they also functioned as an interlocking form of sociability. Once the readers received a copy printed on woodblocks, they actively made use of their prior knowledge and understanding of other references to read the book. Our citation analysis of the headnotes in *Du Fu-Y* showed that their reading was highly overshadowed by the intellectual legacy of the fifteenth century, particularly the exegetical tradition adopted in *Ch'anju*. Moreover, the readers of *Du Fu-Y* seemed to be clearly conscious of the differences in the editorial intentions of *Du Fu-Y* and *Ch'anju*. The rhyming patterns, which are the governing system of *Du Fu-Y*, do not show meaningful associations with the thematic categories used in *Ch'anju*. To identify hidden invisible communities of interpretation, co-citation network analysis was conducted in this study. The results of visualization and quantitative analysis will be enhanced by content analysis in our next planned project.

Notes

1. For the idea of the relationship between texts and readers, I am inspired by Guglielmo Cavallo and Roger Chartier, eds., *A History of Reading* (1999), 1–33.

2. For the history of publication of Du Fu in traditional Korea, see, Kyŏng-ho Sim, *Chosŏn sidae Hanmunhak kwa sigyŏngnon* [A Study of the Classics and the Book of Poetry in Chosŏn] (Sŏul T'ŭkpyŏlsi: Ilchisa, 1999), and Jong-mook Lee, "Tusiŭi ŏnhaeyangsang" [Vernacular Translations of Du Fu in Chosŏn], *Tusiwa tusŏnhae yŏn'gu* [A Study of the Poetry of Du Fu and its Vernacular Translations in Chosŏn] (Sŏul-si: T'aehaksa, 1998).

3. For the compilation history of Ch'anju, see, No Yohan, *Chosŏnch'ogi kwanch'an chuhaesŏŭi munhŏnhakchŏk yŏn'gu* [A Bibliographical Study of Exegetical Traditions Made by the Government During Late Chosŏn Korea] (PhD diss., Korea University, Korea, 2019).

4. The term "research community" refers to a group of scholars, the interactions among them, and their formal or informal associations in disciplines such as information science, sociology of knowledge, sociology of science and technology, and other related studies. Although "scientific community" is also common because the natural and medical sciences have been the primary subject of inquiry in the disciplines mentioned above, interest in the humanities and social studies is growing, and so research community is considered more comprehensive. A similar use of terms is adopted throughout this paper.

5. Henry G. Small, "Cited Documents as Concept Symbols," *Social Studies of Science* 8, no. 3 (1978); Eugene Garfield, "Is Citation Analysis a Legitimate Evaluation Tool?" *Scientometrics* 1, no. 4 (1979): 359–75.

6.   Derek J. De Solla Price, "Networks of Scientific Papers: The Pattern of Bibliographic References Indicates the Nature of the Scientific Research Front," *Science* 149, no. 3683 (1965): 510–15; Derek J. de Solla Price and Donald Beaver, "Collaboration in an Invisible College," *American Psychologist* 21, no. 11 (1966): 1011; Diana Crane, "Social Structure in a Group of Scientists: A Test of the 'Invisible College' Hypothesis," *American Sociological Review* 34, no. 3 (1977): 335–52; Leah A. Lievrouw, "The Invisible College Reconsidered: Bibliometrics and the Development of Scientific Communication Theory," *Communication Research* 16, no. 5 (1989): 615–28.

7.   The most pioneering work is Eugene Garfield, Irving H. Sher, and Richard J. Torpie, *The Use of Citation Data in Writing the History of Science* (Philadelphia: Institute for Scientific Information, 1964). Subsquently, Ralph Garner examines whether graph theory, a subdiscipline of mathematics, is applicable to measure features of a citation network. Ralph Garner, "Computer-Oriented Graph Theoretic Analysis of Citation Index Structures," in *Three Drexel Information Science Research Studies*, ed. Barbara Flood (Philadelphia: Drexel Press, 2004).

8.   Zupic and Čater introduce five bibliometric methods to map research fields. Two have little relevance to citation network analysis and thus are not discussed in this paper: co-author and co-word analyses. For a detailed description of each method, see Ivan Zupic and Tomaž Čater, "Bibliometric Methods in Management and Organization," *Organizational Research Methods* 18, no. 3 (2015): 429–72.

9.   Depending on the unit being analyzed, the method is called author, journal, or document co-citation analysis.

10.   Because the similarity measure is a key part of the analysis, recent studies combine multiple bibliometric methods to maximize their strengths and reduce their limitations. For example, Jeong, Song, and Ding argue that a similarity measure based on co-citation counting could be improved when authors' citation content is also considered. Yoo Kyung Jeong, Min Song, and Ying Ding, "Content-Based Author Co-citation Analysis," *Journal of Informetrics* 8, no. 1 (2014): 197–211.

11.   This trend emerged in part from historical circumstances; the inventor of co-citation analysis, Henry Small, played an important role in the development of bibliometrics. See Zupic and Čater, "Bibliometric Methods," 434–5.

12.   Wolfgang Glänzel and Bart Thijs, "Using 'Core Documents' for Detecting and Labelling New Emerging Topics," *Scientometrics* 91, no. 2 (2012): 399–416.

13.   Wolfgang Glänzel and H. Czerwon, "A New Methodological Approach to Bibliographic Coupling and Its Application to the National, Regional and Institutional Level," *Scientometrics* 37 (1996): 195–221.

14.   Tom Brughmans, "Networks of Networks: A Citation Network Analysis of the Adoption, Use, and Adaptation of Formal Network Techniques in Archaeology," *Literary and Linguistic Computing* 28, no. 4 (2013): 538–62.

15.   Using *L'Année philologique*, a comprehensive international resource for works created in ancient Greece and Rome, three levels of citation networks are constructed and visualized: macro (citation of an ancient author), meso (citation of an ancient work), and micro (citation of a text passage). Romanello calls canonical citation (or canonical reference) "references to passages of the ancient texts," and at first glance, this indicates canonical citation could encompass various studies on citations and references to original ancient texts in digital humanities. Matteo Romanello, "Exploring Citation Networks to Study Intertextuality in Classics," *Digital Humanities Quarterly* 10, no. 2 (2016). However, Romanello and Pasin (2011) illustrate that canonical citations are expressed by means of an abridged canonical format—such as "Hom. *Il*. XII 1," which references the 12th book of Homer's *Iliad*. As such, canonical citation is considered a specific citation to an ancient text, and the term is not adopted in this paper. Matteo Romanello and Michele Pasin, "An Ontological View of Canonical Citations," CIDOC-CRM, 2011, https://cidoc-crm.org/Resources/humanities-citation-ontology-hucit.

16.   Evan Nicoll-Johnson, "Drawing out the Essentials: Historiographic Annotation as a Textual Network," *Journal of Chinese Literature and Culture* 5, no. 2 (2018): 214–49.

17.   We found that four of the 777 poems in *Du Fu-Y* did not match the poems in *Ch'anju*.

18.   Per Ahlgren, Bo Jarneving, and Ronald Rousseau, "Requirements for a Cocitation Similarity Measure, with Special Reference to Pearson's Correlation Coefficient," *Journal of the American Society for Information Science and Technology* 54, no. 6 (2003): 550–60; "Author Cocitation Analysis and Pearson's R," *Journal of the Association for Information Science and Technology* 55, no. 9 (2004): 843; Loet Leydesdorff, "On the Normalization and Visualization of Author Co-Citation Data: Salton's Cosine Versus the Jaccard Index," *Journal of the American Society for Information Science and Technology* 59, no. 1 (2008): 77–85; Loet Leydesdorff and Liwen Vaughan, "Co-occurrence Matrices and Their Applications in Information Science: Extending ACA to the Web Environment," *Journal of the American Society for Information Science and Technology* 57, no. 12 (2006): 1616–28; Howard D. White, "Author Cocitation Analysis and Pearson's R," *Journal of the American Society for Information Science and Technology* 54, no. 13 (2003): 1250–9.

19.   Markus Gmür, "Co-citation Analysis and the Search for Invisible Colleges: A Methodological Evaluation," *Scientometrics* 57, no. 1 (2003): 27–57.

20.   Gmür, "Co-Citation Analysis and the Search for Invisible Colleges," 37–45.

21.   Joe H. Ward, Jr., "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association* 58, no. 301 (1963): 236–44.

22.   On the fundamental problems and debates of clustering, see Maciej Eder, "Visualization in Stylometry: Cluster Analysis Using Networks," *Digital Scholarship in the Humanities* 32, no. 1 (2017): 50–64.

23.   Thomas Bonald et al., "Scikit-Network: Graph Analysis in Python," *Journal of Machine Learning Research* 21 (2020): 1–6.

24. All network visualization was implemented using Gephi 0.9.5.

25. The Scikit-network package was imported into Python to draw the dendrogram of the hierarchical clustering results.

26. It is highly possible that even the anonymous readers, who had written may have been unaware of this pattern.

27. However, two annotators who were also cited frequently, *Liu CW* and *Shi MZ*, are not in the same cluster. This would indicate that the number of citations is associated with similarity, but it does not determine resultant clusters by itself.

References Cited

Du Fu 杜甫. *Ch'anju pullyu tusi* 纂註分類杜詩. Asami Collection, University of Berkeley.

Du Fu 杜甫 and Kŏ-in Sŏ 徐居仁, eds. *Ji qian jia zhu fen lei Du Gongbu shi* 集千家註分類杜工部詩. Naikai Bunko 内閣文庫, Japan (集002-0004).

Du Fu 杜甫 and Yu Yun-gyŏm 柳允謙, eds. *Pullyu Tu Kongbu si ŏnhae* 分類杜工部詩諺解, Korea: Kyŏngsang Kamsa, 16–.

Du Fu 杜甫 and Zhao Pang 趙汸 eds. *Du lü Zhao zhu* 杜律趙註. Taibei: Da tong shu ju 大通書局, 1974.

Du Fu 杜甫 and Pu Qilong 浦起龍 eds. *Du Du xin jie* 讀杜心解. Taibei: Da tong shu ju, 1974.

Ahlgren, Per, Bo Jarneving, and Ronald Rousseau. "Requirements for a Cocitation Similarity Measure, with Special Reference to Pearson's Correlation Coefficient." *Journal of the American Society for Information Science and Technology* 54, no. 6 (2003): 550–60.

Ahlgren, Per, Bo Jarneving, and Ronald Rousseau. "Author Cocitation Analysis and Pearson's R." *Journal of the Association for Information Science and Technology* 55, no. 9 (2004): 843.

Bonald, Thomas, Nathan de Lara, Quentin Lutz, and Bertrand Charpentier. "Scikit-Network: Graph Analysis in Python." *Journal of Machine Learning Research* 21 (2020): 1–6.

Brughmans, Tom. "Networks of Networks: A Citation Network Analysis of the Adoption, Use, and Adaptation of Formal Network Techniques in Archaeology." *Literary and Linguistic Computing* 28, no. 4 (2013): 538–62.

Crane, Diana. "Social Structure in a Group of Scientists: A Test of the 'Invisible College' Hypothesis." *American Sociological Review* 34, no. 3 (1977): 335–52.

Eder, Maciej. "Visualization in Stylometry: Cluster Analysis Using Networks." *Digital Scholarship in the Humanities* 32, no. 1 (2017): 50–64.

Garfield, Eugene. "Is Citation Analysis a Legitimate Evaluation Tool?" *Scientometrics* 1, no. 4 (1979): 359–75.

Garfield, Eugene, Irving H. Sher, and Richard J. Torpie. "The Use of Citation Data in Writing the History of Science." Philadelphia, PA: Institute for Scientific Information Inc, 1964.

Garner, Ralph. "Computer-Oriented Graph Theoretic Analysis of Citation Index Structures." In *Three Drexel Information Science Research Studies*, edited by Barbara Flood, 3–46. Philadelphia: Drexel Press, 2004.

Glänzel, Wolfgang, and H. Czerwon. "A New Methodological Approach to Bibliographic Coupling and Its Application to the National, Regional and Institutional Level." *Scientometrics* 37, no. 2 (1996): 195–221.

Glänzel, Wolfgang, and Bart Thijs. "Using 'Core Documents' for Detecting and Labelling New Emerging Topics." *Scientometrics* 91, no. 2 (2012): 399–416.

Gmür, Markus. "Co-Citation Analysis and the Search for Invisible Colleges: A Methodological Evaluation." *Scientometrics* 57, no. 1 (2003): 27–57.

Hummon, Norman P, and Patrick Dereian. "Connectivity in a Citation Network: The Development of DNA Theory." *Social Networks* 11, no. 1 (1989): 39–63.

Jeong, Yoo Kyung, Min Song, and Ying Ding. "Content-Based Author Co-Citation Analysis." *Journal of Informetrics* 8, no. 1 (2014): 197–211.

Leydesdorff, Loet. "On the Normalization and Visualization of Author Co-Citation Data: Salton's Cosine Versus the Jaccard Index." *Journal of the American Society for Information Science and Technology* 59, no. 1 (2008): 77–85.

Leydesdorff, Loet, and Liwen Vaughan. "Co-Occurrence Matrices and Their Applications in Information Science: Extending ACA to the Web Environment." *Journal of the American Society for Information Science and Technology* 57, no. 12 (2006): 1616–28.

Lievrouw, Leah A. "The Invisible College Reconsidered: Bibliometrics and the Development of Scientific Communication Theory." *Communication Research* 16, no. 5 (1989): 615–28.

Liu, John S., Louis Y.Y. Lu, and Mei Hsiu-Ching Ho. "A Few Notes on Main Path Analysis." *Scientometrics* 119, no. 1 (2019): 379–91.

Nicoll-Johnson, Evan. "Drawing out the Essentials: Historiographic Annotation as a Textual Network." *Journal of Chinese Literature and Culture* 5, no. 2 (2018): 214–49.

Noyons, Ed C.M., Henk F. Moed, and Marc Luwel. "Combining Mapping and Citation Analysis for Evaluative Bibliometric Purposes: A Bibliometric Study." *Journal of the American Society for Information Science* 50, no. 2 (1999): 115–31.

Price, Derek J. De Solla. "Networks of Scientific Papers: The Pattern of Bibliographic References Indicates the Nature of the Scientific Research Front." *Science* 149, no. 3683 (1965): 510–15.

Price, Derek J. de Solla, and Donald Beaver. "Collaboration in an Invisible College." *American Psychologist* 21, no. 11 (1966): 1011.

Romanello, Matteo. "Exploring Citation Networks to Study Intertextuality in Classics." *DHQ: Digital Humanities Quarterly* 10, no. 2 (2016).

Small, Henry G. "Cited Documents as Concept Symbols." *Social Studies of Science* 8, no. 3 (1978): 327–40.

Van Raan, Anthony F.J. "Measuring Science." In *Handbook of Quantitative Science and Technology Research*, 19–50. Springer, 2004.

Ward Jr., Joe H. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58, no. 301 (1963): 236–44.

White, Howard D. "Author Cocitation Analysis and Pearson's R." *Journal of the American Society for Information Science and Technology* 54, no. 13 (2003): 1250–59.

Zupic, Ivan, and Tomaž Čater. "Bibliometric Methods in Management and Organization." *Organizational Research Methods* 18, no. 3 (2015): 429–72.

Kim Mun-sik. *Chŏngjo ŭi kyŏnghak kwa chujahak* [The Study of Classics and Zhu Xi by King Chŏngjo]. Sŏul-si: Munhŏn kwa Haesŏk, 2000.

Lee, Jong-mook. "Tusiŭi ŏnhaeyangsang" [Vernacular Translations of Du Fu in Chosŏn]. *Tusiwa tusŏnhae yŏn'gu* [A Study of the Poetry of Du Fu and its Vernacular Translations in Chosŏn]. Sŏul-si: T'aehaksa, 1998.

No Yohan. *Chosŏnch'ogi kwanch'an chuhaesŏŭi munhŏnhakchŏk yŏn'gu* [A Bibliograohical Study of Exegetical Traditions Made by the Government During Late Chosŏn Korea]. PhD diss., Korea University, Korea, 2019.

Sim, Kyŏng-ho. *Chosŏn sidae Hanmunhak kwa sigyŏngnon* [A Study of the Classics and the Book of Poetry in Chosŏn]. Sŏul T'ŭkpyŏlsi: Ilchisa, 1999.