圖書資訊學研究 15：1 (December 2020)：97-132

Differences Between Bibliographic Coupling and Co-Citation at the Article Level in Library and Information Science Publications

Tsung-Ming Hsiao
Ph.D. Candidate
Department of Library and Information Science
National Taiwan University

Kuang-hua Chen *
Professor
Department of Library and Information Science
National Taiwan University

【Abstract】

We investigated differences between bibliographic coupling (BC) and co-citation (CC) in article pairs and their possible effects. Although several studies have investigated these methods, most have focused on the most effective method for specific applications according to clustering results. We investigated the differences between BC strength (BCS) and CC strength (CCS) of library and information science (LIS) by using articles published from 2009 to 2018 among 44 LIS journals in Journal Citation Reports. Article pairs were based on 1,446 articles from 30 journals published in 2009. BCS was measured according to common references, and CCS was measured according to CCs in the 22,577 articles published in the 44 LIS journals from 2009 to 2018. Different relationship patterns were observed between BC and CC. The authors are usually in common

* Correspondence: Kuang-hua Chen Email: khchen@ntu.edu.tw
Received: 5 October 2020; Accepted: 24 February 2021

when an article pair has high BCS. It shows that authors' citation preferences may affect BCS largely. Although CCS was not affected by citation preferences, CC identified considerably fewer relationships and demonstrated weaknesses in the years following publication. Additionally, sufficient time lag was necessary to reveal the CC relationships between article pairs. In LIS, the time lag to reveal the majority of CC relationships was more than 2 years.

【Keyword】

Informetrics; Bibliographic coupling; Co-citation

Bibliographic coupling (BC) and co-citation (CC) were proposed in the 1960s and the 1970s, respectively, and have been widely applied in retrieving information and mapping science (Rousseau, 2010). BC measures the relationship between two articles by calculating the number of common references in both articles (Kessler, 1963a, 1963b). CC measures the relationship between two articles by calculating how frequently both articles are co-cited together by a later work (Marshakove, 1973; Small, 1973). Numerous scholars have applied BC and CC to various information carriers such as patents or journals (Lai & Wu, 2005; McCain, 1991) and distinct entities such as authors or words (Moya-Anegón et al., 2004; White & Griffith, 1981, 1982).

Although Small (1973) indicated that "Co-citation patterns are found to differ significantly from bibliographic coupling patterns" (p. 265), scholars have rarely used a large amount of data to investigate the differences between the relationships revealed by BC and CC at the article level. In the comparison of BC and CC, most studies have focused on which technique is more appropriate for a specific application (Ahlgren et al., 2020; Boyack & Klavans, 2010; Jarneving, 2005; Klavans & Boyack, 2017; Shibata et al., 2009; Zhao & Strotmann, 2008b). Although these studies have determined which techniques are advantageous in different circumstances, most have conducted analyses at the cluster level and did not explore the common features of article pairs with high BC strength (BCS) or CC strength (CCS). These features can demonstrate what types of article pairs are favored by BC or CC. That is, relevant studies suggest what applications clustering results are appropriate for, but the features of the article pairs identified by each technique remain unknown. These features are relevant for some applications, including information retrieval and scientific structure identification. Further investigation of these features can effectively improve the application of both techniques.

Accordingly, we investigated the features of BC and CC by scrutinizing the two citation relationships. After data collection and preprocessing, we established three analytic targets. First, we analyzed the distribution of BC and CC and re-examined BC and CC patterns to determine the differences between the two techniques. Second, we identified features of article pairs (e.g., common authors) with high BCS or CCS, examined the features, and discussed their effects and meanings. Finally, we considered the time lag of CC as a topic of concern. We scrutinized how the CCS of article pairs accumulates and investigated the effects of the resulting time lag. Our results can help scholars better understand these techniques and use them more appropriately.

The rest of this article describes the three citation relationships, their applications, and the studies that focused on investigating their advantages and weaknesses. After a review of the related literature, details of the research design are reported. Finally, we present the results of each analytic target.

Literature Review

Citation Relationships, Citation Entities, and Their Applications

The citation analysis was based on three types of citation relationships, namely direct citation (DC), BC, and CC. DC is "a relationship between a part or the whole of the cited document and a part or the whole of the citing document" (Smith, 1981, p. 83). Modified from the formulas described in Ahlgren et al. (2020) and Waltman et al. (2019), the formulas we used to calculate the three citation relationships are presented in the following. Let DC_{ik} represent the DC relationship between article *i* and *k*. DC_{ik} is defined as

$$DC_{ik} = \begin{cases} 1 \ if \ article \ i \ cites \ article \ k \\ 0 \ if \ otherwise \end{cases}$$

On the basis of DC, two other citation relationships, BC and CC, were proposed in 1963 and 1973, respectively (Kessler, 1963a; Kessler, 1963b; Marshokova, 1973; Small, 1973). Kessler (1963a, 1963b) proposed BC as a criterion for measuring the relationship between two articles and defined BC as the number of common references between them. That is, the BCS of two articles can be determined by their DC. Let BC_{ij} represents the BCS between article *i* and *j*. BC_{ij} is defined as

$$BC_{ij} = BC_{ji} = \sum_{k=1}^{m} DC_{ik} DC_{jk}$$

where *m* represents the number of distinct references cited by *i* or *j*.
Instead of measuring the relationship between two articles according to their common references, Small (1973) and Marchokova (1973) proposed a new citation relationship, namely CC. CC is determined by counting the number of times, which the articles are co-cited in later research. That is, the CCS of two articles is measured according to how the followings publications cited them. The CCS of two articles can be measured using DC. Let CC_{ij} represents the CCS between article *i* and *j*. CC_{ij} is defined as

$$CC_{ij} = CC_{ji} = \sum_{k=1}^{n} DC_{ki} DC_{kj}$$

where *n* represents the number of distinct materials that cite *i* or *j*.

The three citation relationships constitute the foundation of citation analysis. Initially, scholars proposed DC as an alternative information retrieval tool. The publication of *Shepard's Citations* in 1873 demonstrated the application of DC as a research tool for the legal profession (Garfield, 1955). In the 1950s, Garfield discussed "the possible utility of a citation index that offers a new approach to subject control of the literature of science" (Garfield, 1955, p. 108). In a comparison with subject indexing, Garfield claimed that citation relationships could reveal new topic-related materials because of the different construction approaches involved. After analyzing bibliographic references and constructing networks of scientific papers by using DC, Price (1965) argued that scholars could identify research fronts by applying DC. Instead of analyzing documents, Clark (1957) first applied DC to analyze authors. According to Clark, the number of journal citations counts of an expert moderately correlated with the number of experts chosen as highly visible people in psychology by a panel of experts. Researchers have also explored the information flow, or interrelation, between different subjects. For example, by analyzing the journal-to-journal citations among 275 journals of several disciplines, Narin et al. (1972) investigated the interrelationships among physics, chemistry, biochemistry, biology, and mathematics.

Similarly, BC and CC were employed to group papers when scholars first proposed both methods (Kessler, 1963a, 1963b; Marshakove, 1973; Small, 1973). Since then, several studies have demonstrated the use of BC and CC in measuring the interrelationships between authors (White & Griffith, 1981; Zhao & Strotmann, 2008a) or gauging the connection between subjects (Hsiao & Chen, 2019; Huang et al., 2018; Moya-Anegón et al., 2004). In general, researchers have used citation relationships to observe three citation entities, namely works, authors, and subjects.

Based on these citation relationships, researchers can analyze different citation entities for the following purposes:

1. **Evaluating Publications**. Gross and Gross (1927) applied citation counting to evaluate the importance of scientific publication, and Garfield (1972) proposed that citation analysis could be used for journal evaluation. Researchers have also used citation analysis to evaluate the research output of persons and institutions (Docampo & Bessoule, 2019; Moed, 2005).
2. **Retrieving Information**. Citation relationships reveal the connections between different citation entities and can be used to determine their relevance. Numerous studies have proposed approaches to improving information retrieval by using citation relationships (Gipp & Beel, 2009; Eto, 2018, 2019; Liu, 2017; Liu & Hsu, 2019; Tanner et al., 2019).
3. **Mapping Scientific Structures**. After citations have been accumulated over time, the resulting citation relationships can reveal the historical development of scientific research. Researchers can use these relationships to analyze scientific structures (Chang et al., 2015; González-Valiente et al., 2019; Olmeda-Gómez et al., 2019; Tang et al., 2017).
4. **Detecting Emerging Fields**. Instead of considering the longitudinal evolution of a discipline, certain studies have aimed to identify

emerging research problems (Hou et al., 2018; Zhao & Strotmann, 2008c).

5. **Other**. The aforementioned applications are mentioned in several review articles, including Nicolaisen (2007), White (2009), and Tahamtan and Bronmann (2018). Several atypical applications have been proposed, such as detecting plagiarism (Gipp et al., 2014) and measuring the diffusion of innovation (Zhai et al., 2018).

Scholarly Networks Based on Multiple Relationships

According to Morris and Martens (2008), each citation relationship only reveals particular aspects when applied to mapping research specialties. Most early studies based on citation relationships have typically considered only one relationship (Griffith et al., 1974; Small & Griffith, 1974; White & Griffith, 1981, 1982; White & McCain, 1998). In the late 20th century, researchers began to use multiple citation relationships to analyze a domain from different perspectives and elicit a more comprehensive understanding.

According to Zhao and Strotmann (2008b), the results of author bibliographic coupling analysis (ABCA) and author co-citation analysis (ACA) represent two aspects of a domain. ABCA "provides an alternative and much more realistic view of the internal intellectual structure of a field and its current research activities" (Zhao & Strotmann, 2008b, p. 2081). By contrast, ACA reveals the structures of older works that significantly influence current studies and indirectly indicate how ongoing research activities develop (Zhao & Strotmann, 2008b). They concluded:

> We found that ABCA is an effective method for providing a realistic picture of current active research within a research field, whereas ACA studies the external and internal as well as recent and historical intellectual influences on the field. (Zhao & Strotmann, 2008b, p. 2084)

Their study indicated that both ACA and ABCA provide different perspectives that complement one another. Therefore, instead of using just one, using both ABCA and ACA can provide a comprehensive view of the intellectual structure.

Yan and Ding (2012) analyzed six scholarly networks and categorized them according to their similarities (Figure 1). They argued that the differences between these networks could be explained by two dimensions, namely network remoteness, which can be interpreted as noncitation-based versus citation-based, and social-versus-cognitive dimension. Both BC and CC are citation-based and very close to each other in network remoteness. However, in the social-versus-cognitive dimension, BC tends to indicate more social connections, and CC seems to reveal more similarities between lexical semantics (Yan & Ding, 2012). Overall, the aforementioned studies indicated various differences between BC and CC. BC analyzes the structure of scholarly activities from a social perspective, whereas CC reflects scholarly influences from a cognitive viewpoint. Applying both techniques together is more likely to elucidate intellectual structures.

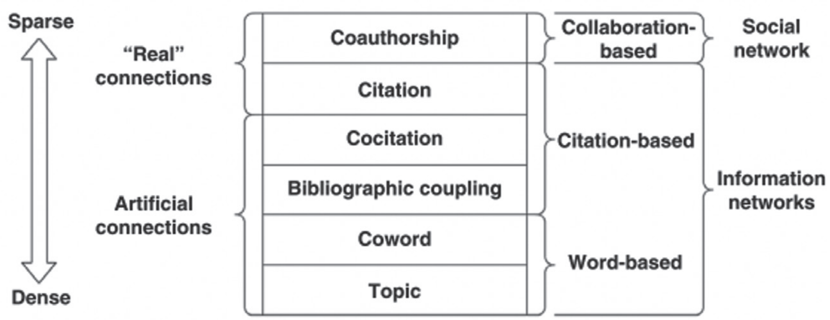

Figure 1. Types of Scholarly Networks from Different Perspectives
Note. From Yan and Ding (2012)

Several studies have also applied multiple methods, including both BC and CC, to identify intellectual structures. Consistent with their previous study, Zhao and Strotmann (2014) used both ACA and ABCA to examine research trends in information science from 2006 to 2010. Because each bibliometric method possesses specific characteristics and advantages for analyzing disciplines, Chang et al. (2015) analyzed the evolution of library and information science (LIS) subject matter between 1995 and 2014 by using three methods, namely BC, CC, and keywords. Tang et al. (2017) also used three methods, namely BC, CC, and coauthorship, to explore intellectual cohesion in digital humanities.

By adopting several methods simultaneously, these studies explored a single domain from various perspectives to offer a comprehensive observation. Instead of analyzing how BC and CC can complement each other, several other studies have focused on determining appropriate applications for them, particularly in mapping scientific structures and identifying research fronts. These studies are reviewed in the following section.

BC and CC in Mapping Science

Several studies have compared BC and CC to determine their advantages and weakness in mapping research fronts. An early attempt to compare scholarly networks by using BC and CC was by Sharabchiev (1989), who reported that "cluster analysis made by means of bibliographic coupling by Kessler and co-citation by Marshakova-Small present comparable results" (p. 127). Additional studies on the applicability of BC and CC in mapping science began in the early 21st century. Jarneving (2005) investigated the results of BC and CC on the basis of 7,239 articles published in the 50 most-cited environmental science journals listed in Journal Citation Reports (JCR). After categorizing these articles by BC and CC, the study compared the word profiles between the group pairs of BC and CC. Because most of the group pairs did not demonstrate high similarity between their word profiles, Jarneving concluded that the research fronts identified by BC and CC were quite different.

Although Jarneving (2005) identified differences between BC and CC, he did not conclude which one was more effective for identifying research fronts. In the following years, several studies investigated this topic. Shibata et al. (2009) compared the performance of BC and CC in detecting research fronts and compared three citation relationships based on normalized cluster size, average publication year, and cluster density. Under these criteria, DC is more effective than BC and CC, and the networks constructed by DC have the least risk of overlooking emerging research domains. However, on the basis of different criteria and datasets, Boyack and Klavans (2010) argued that BC outperformed CC and DC when mapping scientific structures and research fronts. Their study also reported that hybrid similarity, composed of reference and word similarity, exhibited the greatest performance. In another study, Klavans and Boyack (2017) compared the accuracy of three citation relationships in constructing topic-level taxonomies. They reported that DC more effectively concentrated references than did BC and CC. Ahlgren et al. (2020) evaluated multiple publication relatedness measurements on the basis of Medical Subject Headings and revealed that DC enhanced by additional indirect citation relationships exhibited the greatest performance. In addition, CC and DC were outperformed by other techniques, and CC exhibited the poorest performance.

Another concerning issue is the time lag of CC. As indicated by Hopcroft et al. (2004), "a certain time-lag is required in order for papers to build up a citation record" (p. 5250). They argued that a common reference set approach, namely BC, was more effective in the early detection of changes in research communities. Shibata et al. (2009) confirmed the findings of Hopcroft et al. (2004) when analyzing the scholarly networks of three domains. According to Shibata et al. (2009), the number of nodes in the CC network was the lowest in each domain. In addition, the divergence between the number of nodes and edges in the CC and BC networks increased as the time interval decreased. That is, the BC and BCS between two articles is determined immediately following publication. However, both the emergence of CC and the accumulation of CCS require time. Shibata et al. (2009) proposed that time lag is the reason for the low performance of CC in detecting research fronts.

Overall, the aforementioned studies revealed that BC and CC possess different characteristics. The results suggest that both methods complement each other (Yan & Ding, 2012; Zhao & Strotmann, 2008b). In addition, although several studies have also reported inconsistent findings for determining which method is more suitable for a particular purpose, most studies have concluded that BC outperforms CC in identifying research fronts (Ahlgren et al., 2020; Boyack & Klavans, 2010; Klavans & Boyack, 2017; Shibata et al., 2009). However, most relevant studies have investigated the differences at the cluster level but have not yet fully explored the different features between pairs of citation entities with high BCS or high CCS at the article level. Devarakonda et al. (2020) investigated the features and distribution patterns of CC at the article level, indicating that further investigation at the article level remains a worthy pursuit. These features may help researchers appropriately apply each method and interpret the results of both methods more effectively. In addition, exploring these features also contributes to the methodological development of citation analysis by improving the understanding of BC and CC.

Therefore, we explored the features and distinctions between pairs of research articles with high BCS or high CCS. Our study investigated how these features affect the results of scholarly networks revealed by BC and CC. These features may help define appropriate applications for citation relationships. The present study also explored how time lag affects the results of CC networks, the questions posed by Hopocroft et al. (2004) and Shibata et al. (2009). Investigating these questions can help improve the understanding of BC and CC, define more appropriate applications for them, and more accurately explain the scholarly networks revealed by the two methods.

Research Design

We investigated the differences between BC and CC by analyzing LIS publications. Although forming a general conclusion on the basis of a single domain presents challenges, such an analysis may still elucidate research issues and assist in future research. In addition, we can appropriately define this domain based on our familiarity with it. To define LIS, we used the Information Science and Library Science classification (IS-LS) JCR category, which was used to define LIS in other studies, namely Åström (2007). The journals listed in JCR during 2009-2018 were considered. In addition, because this category also includes journals related to management information science (MIS), we excluded MIS journals in accordance with Abrizah et al. (2015) and Huang et al. (2019). Both studies investigated IS-LS subcategories by reviewing questionnaires and recategorized the IS-LS journals into several subcategories. MIS journals were defined as journals not classified as LIS journals in both aforementioned studies. Accordingly, only journals classified as LIS journals were included.

Figure 2 presents our research design. According to the aforementioned criteria, 44 journals were defined as LIS-related journals and used in our study. These journals are presented in Appendix A. The entire dataset comprises articles published in these journals from 2009 to 2018. We downloaded the article data from Web of Science (WoS) on March 19, 2020.

Only records whose document type was "article" were downloaded. To ensure sufficient time for observing the possible time lag effects, we selected articles published in 2009 to construct article pairs. The 10-year span is longer than the majority of the 44 LIS journals' cited half-life. In addition, we used Digital Object Identifier (DOI) to determine how frequently an article was cited after publication and calculated the CCS of each article pair. Therefore, when constructing article pairs, we excluded articles published in several journals because of incomplete DOI records. The article pairs were based on 1,446 articles from 30 journals, a subset of the 44 LIS-related journals. In total, 1,044,735 article pairs were constructed.

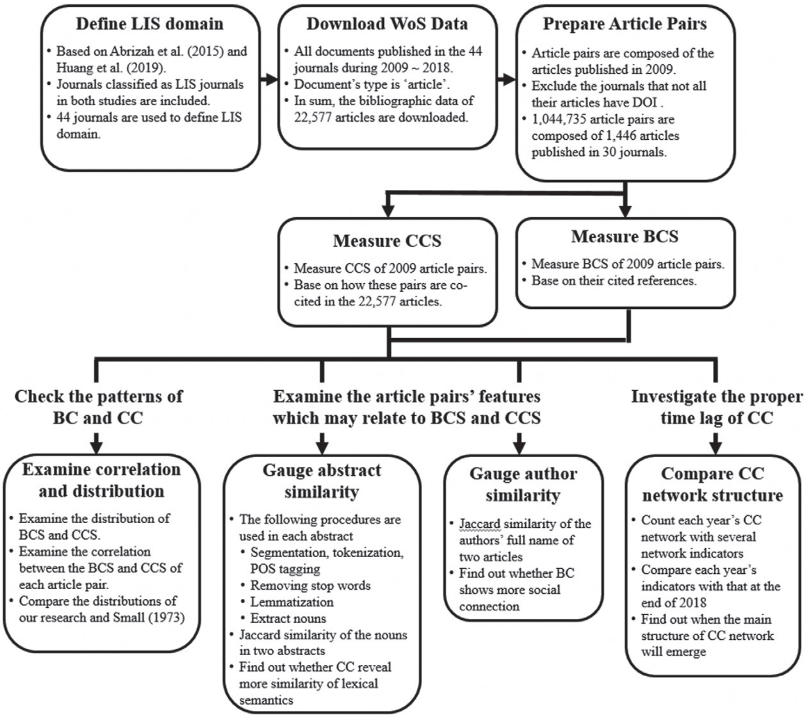

Figure 2. Research Design

First, we analyzed the distribution of BC and CC as well as re-examined BC and CC patterns to determine the differences between the two techniques. We measured the BCS and CCS of each article pair according to the aforementioned formulas. The BCS of an article pair was based on the number of common references in an article pair. CCS was defined as how an article pair are co-cited are co-cited by the 22,577 articles, all of which were published in the 44 LIS-related journals from 2009 to 2018. After calculation, we examined the distribution of BCS and CCS as well as the correlation between the BCS and CCS of each article pair. We also compared our results with those of Small (1973) to determine whether two studies revealed a similar pattern in the BCS and CCS of each article. The re-examination was conducted to ensure that BC and CC revealed different patterns at the article level when a large amount of data was analyzed.

Second, we examined the features of article pairs to determine whether BCS or CCS correlated with certain features. According to Yan and Ding (2012), BC and CC tend to reveal social connections and cognitive relationships, respectively. That is, BC tends to reflect scientific structure from collaborative relationship perspectives, whereas CC tends to identify scientific structure based on the similarities in lexical semantic similarity. Therefore, we examined two features, namely author similarity and abstract similarity, for each article pair. We measured the author similarity of each article pair on the basis of Jaccard similarity. Specifically, our study defined the author similarity between two articles as follows:

$$AuthorSimilarity_{m,n} = \frac{Group_m \cap Group_n}{Group_m \cup Group_n}$$

$$Group_o = \{Author_{o1}, Author_{o2}, \dots, Author_{ox}\}$$

m *and* n represent two articles.

$Group_o$ is composed of all authors of an article.

The authors' full names, extracted from the bibliographic data downloaded from WoS, were used to determine author similarity. Author similarity determines whether BC reveals social connections between the researchers. Abstract similarity verifies whether CC reveals cognitive relationships. An open-source library for natural language processing in Python, spaCy, was applied to several preprocessing procedures, including segmentation, tokenization, part-of-speech tagging, removing stop words, and lemmatization. Following these procedures, we used the Jaccard similarity of nouns in the abstracts as the indicator for whether an article pair with high CC revealed high lexical similarity. The result determined whether CC revealed cognitive relationships between two articles.

Time lag effect was also analyzed. Hopcroft et al. (2004) and Shibata et al. (2009) indicated that time lag is inevitable in the accumulation of CCS and may reduce its efficiency. We investigated the severity of the time lag effect and how much time lag may be required for accumulating CCS. First, for each article pair, we identified when the two articles were co-cited for the first time in LIS to reveal the length of time required to construct the initial CCS between the two articles since publication. In addition, several network indicators were employed to determine the coverage of the CC network each year. We compared the CC network of each year with that at the end of 2018 to determine the extent to which the network grew each year. Furthermore, we analyzed how CCS accumulates each year to identify its growth pattern. We explored whether the article pairs' CCS increases drastically after a given period. Accordingly, we attempted to determine how much time lag was required and identify the possible effects of insufficient time lag.

Results and Discussion

Distributions of BC and CC

Among the 1,446 articles, the average number of references per article was 32.48, and the distribution of references was right-skewed (Figure 3). More than half of the articles (51.73%) cited 13-37 references; approximately 10% of articles cited more than 60 references and approximately 4.5% of articles cited no more than five references. Regarding the distribution of citations, Figure 4 indicates that it was extremely right-skewed. In total, 308 articles (21.3%) were not cited in LIS, and most other articles were cited fewer than 20 times. Only less than 1% of articles were cited more than 50 times.

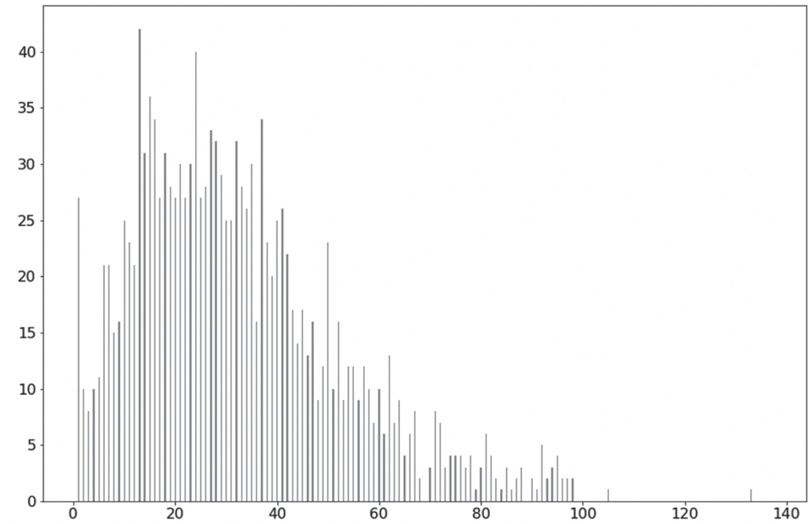

Figure 3. Distribution of References

Note. The x-axis represents the number of references, and the y-axis represents the number of articles.

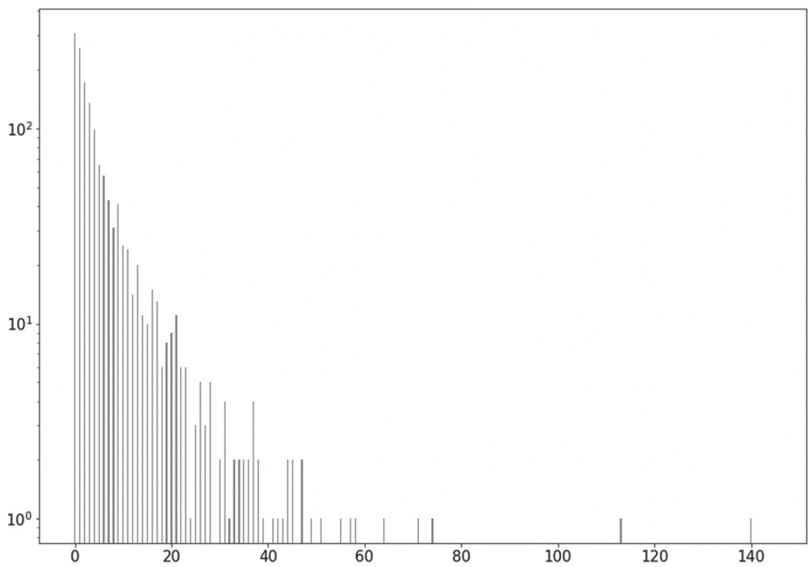

Figure 4. Distribution of Citations

Note. The x-axis represents the number of citations, and the y-axis represents the number of articles.

Figures 5 and 6 present the distribution of article pairs whose BCS or CCS was higher than zero. Among all 1,044,735 possible article pairs, 14,587 and 3,637 pairs exhibited BC and CC relationships, respectively. The incidence of BC and CC was 1.41% and 0.35%, respectively. The incidence of both relationships was rare, but BC revealed more relationships than CC did, indicating that BC connects more articles and constructs a denser network. More than 85% of articles had at least one BC relationship with another article, but CC relationships only connected 61.07% of articles. Although the number of article pairs with CC relationships was much lower than that of those with BC relationships, the full range of CCS was higher than BCS. The distribution of CC may indicate a greater capability to discriminate between strong and weak relationships. That is, CC has poorer coverage than BC does because many uncited articles were initially excluded. However, CC exhibits a greater capability to discriminate articles and relationships if the criterion, cited or not, is of an appropriate standard for judging relatedness.

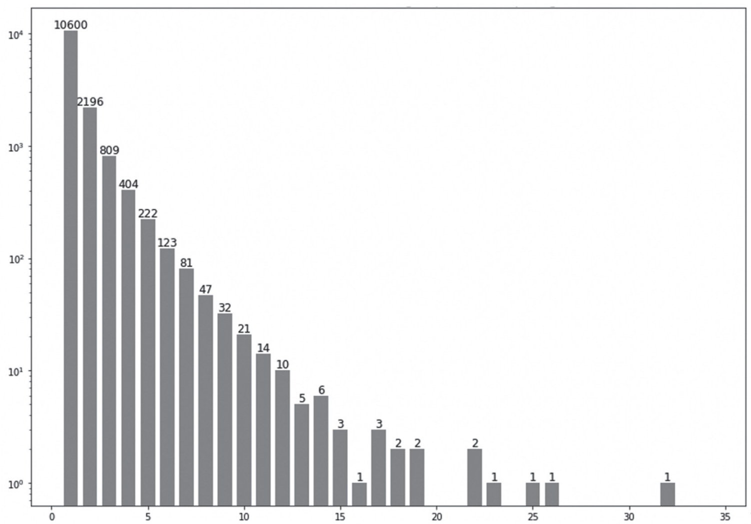

Figure 5. Distribution of BCS

Note. The x-axis denotes BCS, and the y-axis denotes the number of pairs on a log scale.

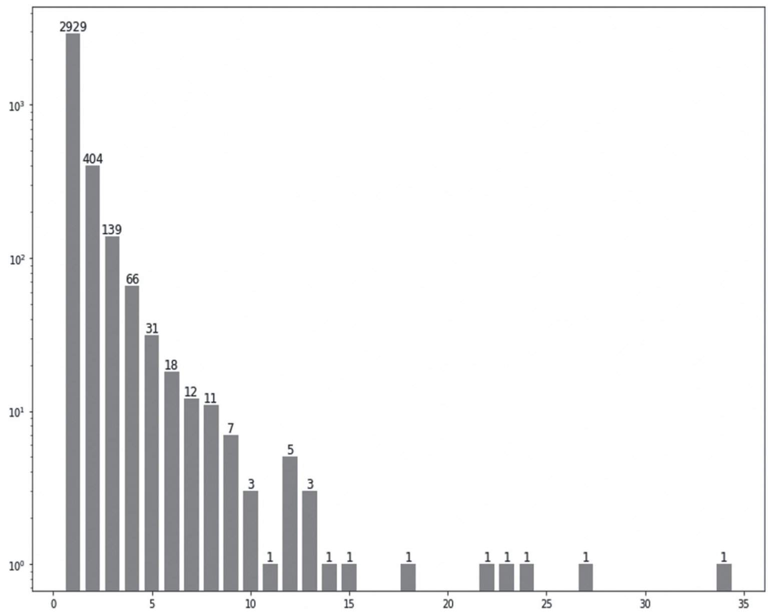

Figure 6. Distribution of CC Strength

Note. The x-axis denotes CC strength, and the y-axis denotes the number of pairs on a log scale.

Relationships and Features of BC and CC

The 2D histogram in Figure 7 represents the relationship between the BCS and CCS of an article pair. The left and right images in Figure 7 present data from Small (1973) and our study, respectively. Data from Small (1973) included 10 articles and 45 article pairs, whereas the present study used 1,446 articles and 1,044,735 article pairs. Note that our study excluded pairs without BC and CC, as shown in the right image of Figure 7. In addition, the two figures differ in scale. Although the two studies used different procedures for data collection, the results of both indicate that BCS did not correlate with CCS.

In the present study, the number of article pairs with high BCS and low CCS largely surpassed that of those with high CCS and low BCS. Given the distribution of BCS and CCS, the results are not surprising. Small (1973) indicated that some pairs have high BC but low CC, which is consistent with our research. Only one pair had high BC and high CC, and the Spearman Correlation Test indicated that the correlation between BC and CC was negative (N = 16,797, $r_s = -0.3558$, $p < .0001$) when pairs with no BC and CC were excluded. This indicates that the pair with high BC was more likely to have low CC. Among the article pairs from 2009, only one pair had very high BCS and CCS, and few pairs had both high BCS and CCS. This polarity further confirms Small (1973), who suggested that the patterns of the two techniques differ significantly. In addition, among the 3,637 CC pairs, 2,210 pairs had zero BCS, further verifying the differences in BC and CC patterns. Overall, the results confirm the different patterns between BC and CC at the article level on the basis of the large amount data analyzed.

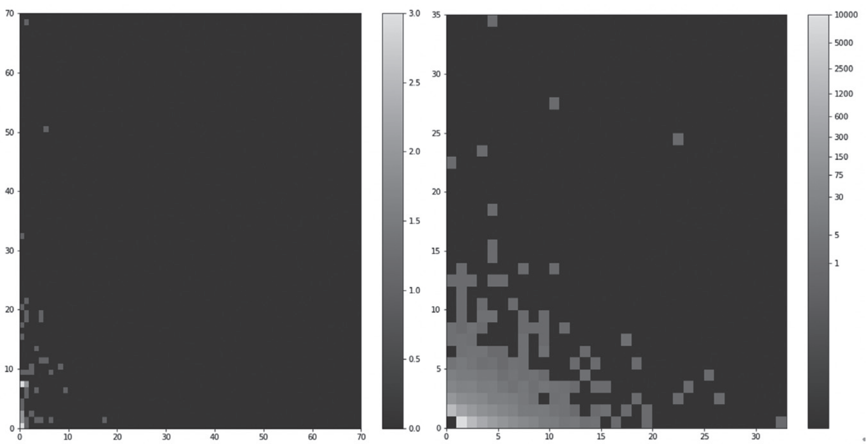

Figure 7. Relationships Between BC and CC

Note. The x-axis and y-axis denote BCS and CCS, respectively. The color of each square represents the number of pairs corresponding to a combination of BCS and CCS. The left image is based on Small (1973), and the right figure is based on our data.

Table 1 presents details of the top 10 BCS pairs, including author similarity, abstract similarity, reference number, and citation counts. Similarly, Table 2 presents the details of the top 11 CCS pairs. Between the two tables, the Jaccard similarity of authors differs considerably. Two articles appear to be written by the same or several common authors when

BCS is very high. Table 1 indicates that 8 of the top 10 BCS pairs shared the same authors and that one pair had numerous common authors. The only article pair with no repeated authors was related to h-index, and one article from this pair was a review article. This tendency was the opposite in the top CCS pairs. Among all 11 pairs, only 2 shared the same authors. Five pairs had no authors in common, and the remaining four pairs exhibited a low author similarity. A possible explanation for this difference is that authors prefer certain citations for a given research topic. Accordingly, high BCS may be due to authors' citation preferences and article type. These results are consistent with those of Yan and Ding (2012), who suggested that BC tends to identify a higher number of social connections between two articles than CC does.

Table 1
Top 10 BC Article Pairs

DOI(a)	DOI(b)	Refs	Cit	BC	CC	AuS	AbS
10.1177/0961000608096717	10.1016/j.lisr.2008.06.004	128	24	32	0	1.00	0.24
10.1108/00012530911005535	10.1016/j.lisr.2009.02.001	105	23	26	2	1.00	0.21
10.1108/00220410910970249	10.1002/asi.21030	92	29	25	4	1.00	0.17
10.1016/j.joi.2009.03.003	10.1002/asi.21199	147	19	23	3	1.00	0.12
10.1002/asi.20967	**10.1002/asi.21086**	**106**	**211**	**22**	**24**	**1.00**	**0.27**
10.1007/s11192-008-2174-9	10.1007/s11192-009-0415-1	127	52	22	2	0.00	0.07
10.1177/0961000609345088	10.1016/j.lisr.2008.06.004	124	24	19	1	1.00	0.36
10.1002/asi.21098	10.1108/14684520910944382	98	6	19	0	0.67	0.34
10.1016/j.ipm.2008.05.005	10.1002/asi.21021	91	17	18	5	1.00	0.15
10.1007/s11192-009-2126-z	10.1007/s11192-008-2074-z	72	52	18	2	1.00	0.26

Note. The article pair in bold also appears in Table 2. "Refs" is the total number of references between the two articles, "Cit" is the total number of citations between the two articles, "AuS" is the Jaccard similarity of authors, and "AbS" is the Jaccard similarity of abstracts.

Table 2
Top 10 BC Article Pairs

DOI(a)	DOI(b)	Refs	Cit	BC	CC	AuS	AbS
10.1002/asi.20967	10.1002/asi.20991	87	214	4	34	0.00	0.06
10.1002/asi.20967	10.1007/s11192-008-2197-2	95	253	10	27	0.33	0.06
10.1002/asi.20967	**10.1002/asi.21086**	**106**	**211**	**22**	**24**	**1.00**	**0.27**
10.1002/asi.20991	10.1002/asi.21086	71	145	3	23	0.00	0.05
10.1016/j.joi.2008.10.002	10.1002/asi.21045	71	66	0	22	0.00	0.12
10.1002/asi.21085	10.1002/asi.21045	99	76	4	18	0.00	0.11
10.1002/asi.20991	10.1007/s11192-008-2197-2	60	187	4	15	0.00	0.06
10.1002/asi.21059	10.1007/s11192-008-2075-y	19	57	4	14	1.00	0.09
10.1108/02640470910934669	10.1108/02640470910966916	34	30	1	13	0.25	0.36
10.1002/asi.21086	10.1002/asi.21020	109	94	7	13	0.33	0.17
10.1002/asi.21086	10.1007/s11192-008-2197-2	79	184	10	13	0.33	0.08

Note. The article pair in bold also appears in Table 1. “Refs” is the total number of references between the two articles, “Cit” is the total number of citations between the two articles, “AuS” is the Jaccard similarity of authors, and “AbS” is the Jaccard similarity of abstracts.

In addition, we used spaCy to identify nouns from each article’s abstract. These nouns were used to determine whether CC tends to reveal similarity in lexical semantics, as Yan and Ding (2012) proposed. Surprisingly, pairs with high BCS typically have an equal or higher average similarity at the article level compared with pairs with high CCS. That is, the capability of BC to reveal lexical similarity may be not lower than that of CC at the article level. However, given the complexity of semantic meaning and the different granularities of analysis, this only indicates that further investigation is warranted and does not contradict the findings of Yan and Ding (2012).

Time Lag and Its Effects on CC

Time lag, as mentioned in Hopocroft et al. (2004) and Shibata et al. (2009), was also examined in the current study. Among the 3,637 CCS pairs, we identified when relationships emerged, how CCS accumulated, and how CC networks changed in four nonconsecutive years:

- 2009: The year both articles were published.
- 2011: The last year that the articles published in 2009 were included when counting the journal's impact factor.
- 2014: The last year that the articles published in 2009 were included when counting the journal's 5-year impact factor.
- 2018: The final year included in our research.

Table 3 and Figure 8 present the proportion of articles published in 2009 and the number of article pairs co-cited in each year compared with the number at the end of 2018. Although coverage grew rapidly from 2009 to 2011, numerous articles and pairs were not included in the CC network by the end of 2011. Until 2011, 61.72% of articles were co-cited, and only 45.7% of pairs emerged. After 2011, the growth rate slowed and remained steady in the second half of the period. At the end of 2014, 88.79% of articles and 77.98% of article pairs were included in the CC network. According to these figures, sufficient time lag should be at least 2 years, which explains why the aforementioned studies reported that CC rarely outperformed other methods in identifying research fronts.

Table 3
Number of Included Articles and Identified Article Pairs Per Year

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Included 2009 Articles	37	290	545	655	740	784	819	848	864	883
	4.19%	32.84%	61.72%	74.18%	83.81%	88.79%	92.75%	96.04%	97.85%	100.00%
Identified Articles Pairs	37	529	1662	2156	2486	2836	3090	3310	3479	3637
	1.02%	14.54%	45.70%	59.28%	68.35%	77.98%	84.96%	91.01%	95.66%	100.00%

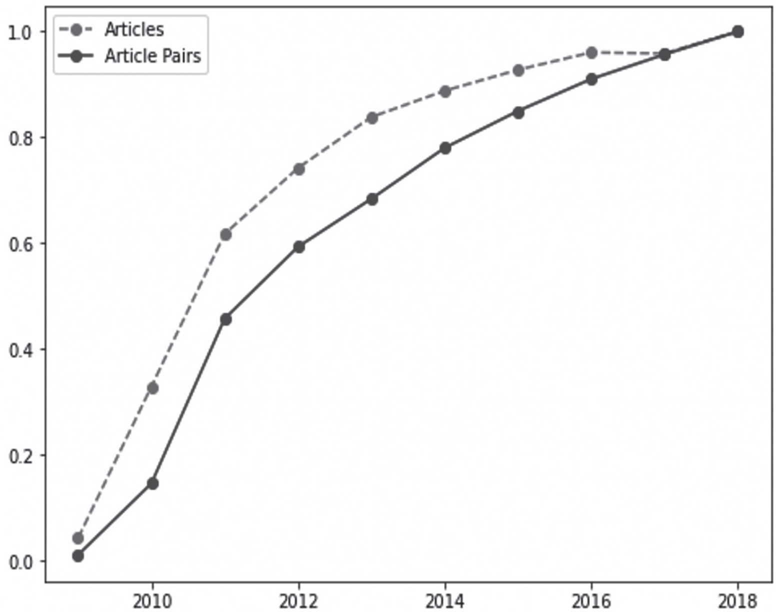

Figure 8. Growth of Coverage for Individual Articles and Article Pairs Revealed by CC

Note. The y-axis is the cumulative percentage.

We also used Gephi 0.92 to calculate four network indicators, as shown in Table 4. These indicators were used to identify features in CC networks at different times. Both the average degree and average weighted degree grew rapidly until 2011. After 2011, the growth level became relatively low. Similarly, the number of connected components peaked in 2010, decreasing thereafter to 36 in 2011. In the remaining years, the number of connected components varied between 33 and 38. Graph density only declined slightly at the beginning and remained stable after 2013. Overall, the indicators initially fluctuated and stabilized thereafter. The results suggest that the primary section of a CC network is revealed after a sufficiently long time lag, and the structure remains relatively stable thereafter.

Table 4
Average Degree, Network Density, and Number of Components Per Year

	Average Degree	Avg Weighted Degree	Graph Density	Connected Components
2009	2	2.216	0.056	12
2010	3.648	4.207	0.013	43
2011	6.099	7.27	0.011	36
2012	6.583	8.26	0.01	37
2013	6.719	8.708	0.009	38
2014	7.235	9.714	0.009	35
2015	7.546	10.361	0.009	35
2016	7.807	11.017	0.009	36
2017	8.053	11.5	0.009	36
2018	8.238	11.853	0.009	33

The following discussion concerns how CCS increased over the study period. Figure 9 presents the number of article pairs whose weight increased at different levels from 2009 to 2018. For example, the number in the third column at the bottom row indicates that the CCS value of 1,138 pairs increased one in 2011, and the number in the fourth column at the top row indicates that the CCS value of one pair increased eight in 2012. These pairs were co-cited 3,808 times from 2009 to 2014. After 2015, they were co-cited 1,425 times. That is, nearly 30% of CCS pairs appeared after 2015. In addition, several article pairs were frequently co-cited in the second half of the study period. For example, in 2016, the CCS of one and six pairs increased by seven and four, respectively. Among these seven pairs, one pair was first co-cited in 2016, and its CCS increased to four in the same year; in addition, the other five pairs were co-cited more than 10 times before 2016. The first increase represents a late-occurring CC relationship, and the second increase reflects a phenomenon similar to the Matthew Effect.

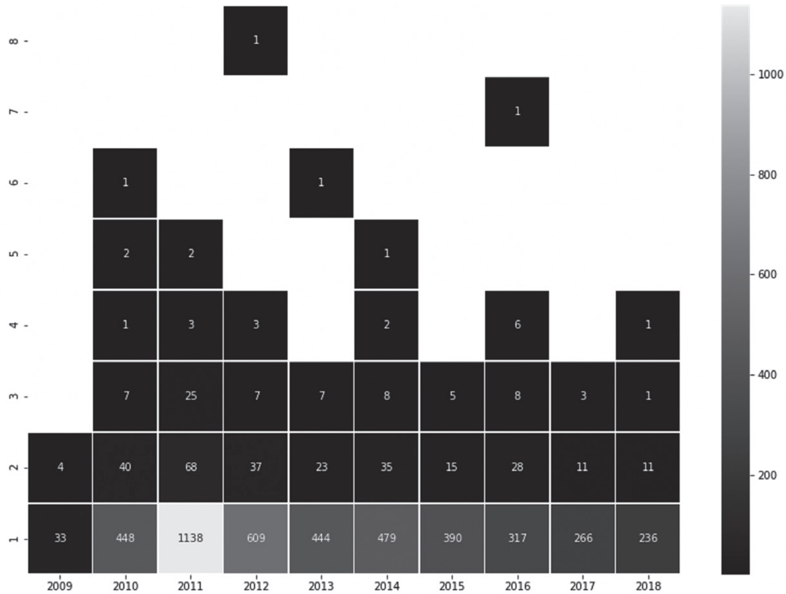

Figure 9. Number of Article Pairs Whose Edge Weight Increased from 2009 to 2018

Overall, our study confirms that BC and CC, as argued by Small (1973), differ significantly at the article level in the context of large-scale data. Although BC and CC are both indirect citation relationships based on DC, they provide distinct perspectives. In our study, article pairs with high BCS rarely had high CCS, and the correlation coefficient between BCS and CCS was weakly negative.

We also observed that article pairs with high BCS were more likely to have common authors and use similar nouns. That is, they were likely written by the same authors and may have focused on similar topics. Because BC is based on authors' cited references, BC likely reflects not only the relatedness between two articles but also the similarity of their authors' citation preferences within a specific research topic. Although certain core works may be common to a research topic, the use of non-core but related works depends on authors' preferences. Therefore, authors' preferences play critical roles in generating the BC structures.

However, the effects of authors' preferences are not apparent when identifying relationships by using CC. When calculating the BCS of an article pair, only the preferences of the two articles' authors are considered. For CC, the preferences of all citing articles' authors are relevant. Therefore, high CCS is more likely to reveal a relationship that conveys common acknowledgment. Our results also indicate that high CCS pairs do not have common features such as repeated authors or similar abstracts. By examining how the later published works cite the previous articles, CC provides a method of identifying a relationship between two articles based on the later works. Compared with BC, in which BCS remains stable after their publication of an article pair, CC is very dynamic and reveals how scholars regard the relationships among the articles.

Our results support the arguments in Zhao and Strotmann (2008b) and Yan and Ding (2012). The results indicate that BC and CC provide different perspectives and possess a particular type of focus. As Zhao and Strotmann (2008b) proposed, BC and CC results complement one another, and both techniques are necessary for determining scientific structures and elucidating other applications. In addition, as Yan and Ding (2012) reported, BC tends to reveal social connections, and CC is likely to reveal cognitive connections. Our study indicates that the BCS of two articles may relate to their authors' citation preferences, particularly when their BCS is high. Regarding CC, the reason why an article pair has high CCS is not obvious. However, high CCS may indicate common acknowledgment by multiple citing authors. Thus, CC likely represents the cognitive connections between different articles.

The capability to reveal researchers' general cognition regarding a group of articles is the primary advantage of CC. However, this advantage also damages the immediacy of CC because of the amount of time required for the confluence of general cognition. Our results indicate that only a minority of articles are co-cited immediately following publication. According to our results, a time lag of more than 2 years following the publication of both articles may be sufficient for their CC relationships to accumulate in LIS. Therefore, CC may not be an appropriate method for analyzing current research activities.

Another advantage of CC is its ability to reveal dynamic relationships between articles. Although BC network structures vary slightly as new articles enter the network, the BC relationship between two articles stabilizes following publication. The relationships identified by CC are more dynamic and change more over time than those identified by BC. This makes CC a more effective tool for measuring the relevance of two articles from different perspectives in different periods. BC does not possess this capability. Liu and Hsu (2019) measured different articles' relatedness according to category-based CC to address BC's weakness in identifying that different references in two articles may relate to one another. The researchers reported that their method "performs significantly better than state-of-the-art variants of BC in identifying highly related articles" and "provides helpful information to further improve a biomedical search engine" (Liu & Hsu, 2019, p. 5176). This suggests that both BC and CC possess advantages, and any suitable method combining them may likely outperform current techniques among different applications. Although BC is static and leans toward authors' perspectives, it provides a quick way to measure articles' relatedness. CC reflects the relevance of two articles' according to the citing authors in different periods, but a considerable amount of time is required for CCS accumulation.

Conclusion

Our study compared the BCS and CCS of article pairs. At the article level, our results confirm the differences and demonstrate the degree of divergence. The features of article pairs with high BCS indicate that both topic relatedness and authors' citation preferences influence BCS. Although CC is more capable of identifying commonly acknowledged relationships than BC is, further examination revealed that CC exhibits apparent weakness in immediacy compared with BC. In LIS, the sufficient time lag is at least 2 years.

Because our dataset comprised only articles published in LIS journals from 2009 to 2018, further investigation is required to determine whether these conclusions are applicable to other disciplines. Future studies may include data from various disciplines and periods to ensure comprehensive findings. In addition, our results reveal that BCS is more related to lexical similarity than CCS is. Because it does not fully support those of Yan and Ding (2012), who suggested that CC reveals relationships with more lexical semantic similarity, additional studies are required to validate our limited sample size and the complicated nature of semantic meaning.

Overall, our results revealed the possible features of article pairs with high BCS or CCS, identified how time lag affects CC, and improved the understanding of citation relationships of article pairs at the article level. Exploring the differences between BC and CC helps researchers improve their understanding of citation relationships and determine suitable applications for them. BC is advantageous in immediately and directly identifying current activities in a domain, but it may largely be affected by personal citation preferences. Conversely, CC reveals common acknowledge relationships dynamically but requires time to form an acceptable result. Both methods have advantages and disadvantages, and additional studies are necessary to determine a suitable technique for applying both methods.

References

Abrizah, A., Noorhidawati, A., & Zainab, A. N. (2015). LIS journals categorization in the Journal Citation Report: A stated preference study. *Scientometrics, 102*, 1083-1099. https://doi.org/10.1007/s11192-014-1492-3

Ahlgren, P., Chen, Y., Colliander, C., & van Eck, N. J. (2020). Enhancing direct citations: A comparison of relatedness measures for community detection in a large set of PubMed publications. *Quantitative Science Studies,1*(2), 714-729. https://doi.org/10.1162/qss_a_00027

Åström, F. (2007). Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990-2004. *Journal of the American Society for Information Science and Technology, 58*(7), 947–957. https://doi.org/10.1002/asi.20567

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology, 61*(12), 2389-2404. https://doi.org/10.1002/asi.21419

Chang, Y.-W., Huang, M.-H., & Lin, C.-W. (2015). Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics, 105*(3), 2071-2087.

Clark, K. E. (1957). Indices of eminence. *America's psychologists: A survey* *of a growing profession* (pp. 26-61). Washington, D. C.: American Psychological Association.

Devarakonda, S., Bradley, J. R., Korobskiy, D., Warnow, T., & Chacko, G. (2020). Frequently cocited publications: Features and kinetics. *Quantitative Science Studies, 1*(3), 1223-1241. https://doi.org/10.1162/qss_a_00075

Docampo, D., & Bessoule, J.-J. (2019). A new approach to the analysis and evaluation of the research output of countries and institutions. *Scientometrics, 119*(2), 1207-1225. https://doi.org/10.1007/s11192-019-03089-w

Eto, M. (2018). Can rough co-citation improve the performance of co-citation clustering? *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, 329-330. https://doi.org/10.1145/3197026.3203908

Eto, M. (2019). Extended co-citation search: Graph-based document retrieval on a co-citation network containing citation context information. *Information Processing & Management, 56*(6), 102046. https://doi.org/10.1016/j.ipm.2019.05.007

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science, 122*(3159), 108-111.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science, 178*(4060), 471-479.

Gipp, B., & Beel, J. (2009). Citation Proximity Analysis (CPA) - A new approach for identifying related work based on co-citation analysis. *Proceedings of ISSI 2009*.

Gipp, B., Meuschke, N., & Breitinger, C. (2014). Citation-based plagiarism detection: Practicability on a large-scale scientific corpus. *Journal of the Association for Information Science and Technology, 65*(8), 1527-1540. https://doi.org/10.1002/asi.23228

González-Valiente, C. L., Santos, M. L., & Arencibia-Jorge, R. (2019). Evolution of the socio-cognitive structure of knowledge management (1986–2015): An author co-citation analysis. *Journal of Data and Information Science, 4*(2), 36-55. https://doi.org/10.2478/jdis-2019-0008

Griffith, B. C., Small, H. G., Stonehill, J. A., & Dey, S. (1974). The structure of scientific literatures II: Toward a macro- and microstructure for science. *Science Studies, 4*(4), 339-365.

Gross, P. L. K., & Gross, E. M. (1927). College libraries and chemical education. *Science, 66*(1713), 385-389. https://doi.org/10.1126/science.66.1713.385

Hopcroft, J., Khan, O., Kulis, B., & Selman, B. (2004). Tracking evolving communities in large linked networks. *Proceedings of the National Academy of Sciences of the United States of America, 101*(suppl. 1), 5249-5253. https://doi.org/10.1073/pnas.0307750100

Hou, J., Yang, X., & Chen C. (2018). Emerging trends and new developments in information science: A document co-citation analysis (2009-2016). *Scientometrics, 115*(2), 869-892. https://doi.org/10.1007/s11192-018-2695-9

Hsiao, T.-M., & Chen, K. (2019). Word bibliographic coupling: Another way to map science field and identify core references. *Proceedings of the Association for Information Science and Technology, 56*(1), 107-116.

Huang, M., Shaw, W.-C., & Lin, C.-S. (2019). One category, two communities: Subfield differences in "Information Science and Library Science" in Journal Citation Reports. *Scientometrics, 119*, 1059-1079. https://doi.org/10.1007/s11192-019-03074-3

Huang, W., Wang, B., Bu, Y., & Min, C. (2018). A study on scientometrics of co-citation analysis of keywords. *Information and Documentation Services, 2018*(2), 37-42.

Jarneving, B. (2005). A comparison of two bibliometric methods for mapping of the research front. *Scientometrics, 65*(2), 245-263. https://doi.org/10.1007/s11192-005-0270-7

Kessler, M. (1963a). An experimental study of bibliographic coupling between technical papers. *IEEE Transactions on Information Theory, 9*(1), 49-51.

Kessler, M. (1963b). Bibliographic coupling between scientific papers. *American Documentation, 14*(1), 10-25.

Klavans, R., & Boyack, K. W. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology, 68*(4), 984-998.

Lai, K.-K., & Wu, S.-J. (2005). Using the patent co-citation approach to establish a new patent classification system. *Information Processing & Management, 41*(2), 313-330. https://doi.org/10.1016/j.ipm.2003.11.004.

Liu, R.-L. (2017). A new bibliographic coupling measure with descriptive capability. *Scientometrics, 110*(2), 915-935. https://doi.org/10.1007/s11192-016-2196-7

Liu, R.-L., & Hsu, C.-K. (2019). Improving bibliographic coupling with category-based cocitation. *Applied Sciences, 9*(23), 5176-5193. https://doi:10.3390/app9235176

Marshakove, I. (1973). System of document connections based on references. *Nauchno-Tekhnicheskaya Informatsiya, 2*(6), 3-8.

McCain, K. W. (1991). Core journal networks and cocitation maps: New bibliometric tools for serials research and management. *The Library Quarterly: Information, Community, Policy, 61*(3), 311-336.

Moed, H. F. (2005). *Citation analysis in research evaluation*. Springer.

Morris, S. A., & Martens, B. V. der V. (2008). Mapping research specialties. *Annual Review of Information Science and Technology, 42*(1), 213-295. https://doi.org/10.1002/aris.2008.1440420113

Moya-Anegón, F., Vargas-Quesada, B., Herrero-Solana, V., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., & Munoz-Fernández, F. J. (2004). A new technique for building maps of large scientific domains based on the cocitation of classes and categories. *Scientometrics, 61*(1), 129-145.

Narin, F., Carpenter, M., & Berlt, N. C. (1972). Interrelationships of scientific journals. *Journal of the American Society for Information Science, 23*(5), 323-331.

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology, 41*(1), 609-641.

Olmeda-Gómez, C., Romá-Mateo, C., & Ovalle-Perandones, M.-A. (2019). Overview of trends in global epigenetic research (2009-2017). *Scientometrics, 119*(3), 1545-1574. https://doi.org/10.1007/s11192-019-03095-y

Price, D. (1965). Networks of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front. *Science, 149*(3683), 157-162.

Rousseau, R. (2010). Bibliographic coupling and co-citation as dual notions.

In L. Birger (Ed.), *The Janus Faced Scholar: A Festschrift in Honour of Peter Ingwersen* (pp. 173-183).

Sharabchiev, J. T. (1989). Cluster analysis of bibliographic references as a scientometric method. *Scientometrics, 15*(1), 127–137. https://doi.org/10.1007/BF02021804

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology, 60*(3), 571-580.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265-269. https://doi.org/10.1002/asi.4630240406

Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies, 4*(1), 17-40.

Smith, L. C. (1981). Citation analysis. *Library Trends, 30*(1), 83-106.

Tahamtan, I., & Bornmann, L. (2018). Core elements in the process of citing publications: Conceptual overview of the literature. *Journal of Informetrics, 12*(1), 203-216.

Tang, M.-C., Cheng, Y. J., & Chen, K. H. (2017). A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses. *Scientometrics, 113*(2), 985-1008.

Tanner, W., Akbas, E., & Hasan, M. (2019). Paper recommendation based on citation relation. *2019 IEEE International Conference on Big Data* (Big Data), 3053-3059. https://doi.org/10.1109/BigData47090.2019.9006200

Waltman, L., Boyack, W. K., Colavizza, G., & van Eck, N. J. (2019). A principled methodology for comparing relatedness measures for clustering publications. arXiv. Retrieved from https://arxiv.org/abs/1901.06815

White, H. D. (2009). Citation analysis. *In Encyclopedia of Library and Information Sciences, Third Edition* (pp. 1012-1026). Taylor & Francis. http://www.tandfonline.com/doi/abs/10.1081/E-ELIS3-120043464

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for*

Information Science, 32(3), 163-171.

White, H. D., & Griffith, B. C. (1982). Authors as markers of intellectual space: Co-citation in studies of science, technology and society. *Journal of Documentation, 38*(4), 255-272. https://doi.org/10.1108/eb026731

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science, 49*(4), 327-355. https://doi.org/10.1002/(SICI)1097-4571(19980401)49:4<327::AID-ASI4>3.0.CO;2-4

Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, co-authorship networks, and coword networks relate to each other. *Journal of the American Society for Information Science and Technology, 63*(7), 1313-1326. https://doi.org/10.1002/asi.22680

Zhai, Y., Ding, Y., & Wang, F. (2018). Measuring the diffusion of an innovation: A citation analysis. *Journal of the Association for Information Science and Technology, 69*(3), 368-379. https://doi.org/10.1002/asi.23898

Zhao, D., & Strotmann, A. (2008a). Author bibliographic coupling: Another approach to citation-based author knowledge network analysis. *Proceedings of the American Society for Information Science and Technology, 45*(1), 1-10.

Zhao, D., & Strotmann, A. (2008b). Evolution of research activities and intellectual influences in information science 1996-2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology, 59*(13), 2070-2086.

Zhao, D., & Strotmann, A. (2008c). Information science during the first decade of the web: An enriched author cocitation analysis. *Journal of the American Society for Information Science and Technology, 59*(6), 916-937. https://doi.org/10.1002/asi.20799

Zhao, D., & Strotmann, A. (2014). The knowledge base and research front of information science 2006-2010: An author cocitation and bibliographic coupling analysis. *Journal of the Association for Information Science and Technology, 65*(5), 995-1006. https://doi.org/10.1002/asi.23027

Appendix A

Journals List

1. AFRICAN JOURNAL OF LIBRARY ARCHIVES AND INFORMATION SCIENCE
2. **ASLIB JOURNAL OF INFORMATION MANAGEMENT/ASLIB PROCEEDINGS**
3. CANADIAN JOURNAL OF INFORMATION AND LIBRARY SCIENCE-REVUE CANADIENNE DES SCIENCES DE L
4. **COLLEGE & RESEARCH LIBRARIES**
5. ECONTENT
6. **ELECTRONIC LIBRARY**
7. **GOVERNMENT INFORMATION QUARTERLY**
8. **HEALTH INFORMATION AND LIBRARIES JOURNAL**
9. **INFORMATION PROCESSING & MANAGEMENT**
10. INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL
11. **INFORMATION SOCIETY**
12. **INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE**
13. INVESTIGACION BIBLIOTECOLOGICA
14. JOURNAL OF ACADEMIC LIBRARIANSHIP
15. **JOURNAL OF DOCUMENTATION**
16. **JOURNAL OF HEALTH COMMUNICATION**
17. **JOURNAL OF INFORMATION SCIENCE**
18. **JOURNAL OF INFORMETRICS**
19. **JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE**
20. **JOURNAL OF SCHOLARLY PUBLISHING**
21. **JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION**
22. **JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY/JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY**

23. JOURNAL OF THE MEDICAL LIBRARY ASSOCIATION
24. LEARNED PUBLISHING
25. LIBRARY & INFORMATION SCIENCE RESEARCH
26. LIBRARY AND INFORMATION SCIENCE
27. LIBRARY COLLECTIONS ACQUISITIONS & TECHNICAL SERVICES
28. LIBRARY HI TECH
29. LIBRARY QUARTERLY
30. LIBRARY RESOURCES & TECHNICAL SERVICES
31. LIBRARY TRENDS
32. LIBRI
33. MALAYSIAN JOURNAL OF LIBRARY & INFORMATION SCIENCE
34. ONLINE INFORMATION REVIEW
35. PORTAL-LIBRARIES AND THE ACADEMY
36. PROFESIONAL DE LA INFORMACION
37. DATA TECHNOLOGIES AND APPLICATIONS/PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS
38. RESEARCH EVALUATION
39. RESTAURATOR-INTERNATIONAL JOURNAL FOR THE PRESERVATION OF LIBRARY AND ARCHIVAL MATERIAL
40. SCIENTIST
41. SCIENTOMETRICS
42. SERIALS REVIEW
43. SOCIAL SCIENCE INFORMATION SUR LES SCIENCES SOCIALES
44. ZEITSCHRIFT FUR BIBLIOTHEKSWESEN UND BIBLIOGRAPHIE

Note. The articles published in 30 journals whose names are in bold were used to construct the article pairs. The BCS of an article pair was based on their common references. The CCS of an article pair was based on the number of times the two articles were co-cited by the articles published in all 44 journals from 2009 to 2018.

圖書資訊學研究 15：1 (December 2020)

文章層級之書目耦合關係與共被引關係差異分析：以圖書資訊學文獻爲例

Differences Between Bibliographic Coupling and Co-Citation at the Article Level in Library and Information Science Publications

蕭宗銘

Tsung-Ming Hsiao

國立臺灣大學圖書資訊學研究所博士候選人
Ph.D. Candidate
Department of Library and Information Science
National Taiwan University

陳光華*

Kuang-hua Chen

國立臺灣大學圖書資訊學研究所教授
Professor
Department of Library and Information Science
National Taiwan University

【摘要】

本研究分析比較書目耦合與共被引方法，以探討文章層級之書目關係的差異及其影響。先前相關研究主要是依網絡分群結

*通訊作者：陳光華 khchen@ntu.edu.tw
投稿日期：2020 年 10 月 05 日；接受日期：2021 年 2 月 24 日

果，探討不同方法所適合之應用，本研究則關注於同對文章之書目耦合強度與共被引強度異同的比較。分析資料之範圍為 44 種在 2009 至 2018 年間被完整收錄於 Journal Citation Report 中的圖書資訊學期刊。本研究從中選定於 2009 年在 30 種具完整 DOI 記錄之期刊上出版的 1,446 篇文章，組成分析用之文章對，並計算其書目耦合強度。復以 2009 至 2018 年間出版於 44 種期刊上之 22,577 篇文獻，計算各文章對之共被引強度。研究發現兩個方法在文章層級上具不同模式。高書目耦合強度之文章對，其兩篇文章之作者組合相似度高，顯示書目耦合強度可能受作者引用習慣影響。此因素對於共被引強度無明顯影響，但共被引方法亦有辨識關係較少及需較長累積時間的弱點。進一步分析顯示，圖書資訊學領域的共被引關係之累積時間需兩年以上。

【關鍵字】

資訊計量學；書目耦合；共被引