A Comparative Analysis on Multiple Authorship Counting for Author Co-citation Analysis*

저자동시인용분석을 위한 복수저자 기여도 산정 방식의 비교 분석

Jae Yun Lee (이재윤)**
EunKyung Chung (정은경)***

ABSTRACT

As co-authorship has been prevalent within science communities, counting the credit of co-authors appropriately is an important consideration, particularly in the context of identifying the knowledge structure of fields with author-based analysis. The purpose of this study is to compare the characteristics of co-author credit counting methods by utilizing correlations, multidimensional scaling, and pathfinder networks. To achieve this purpose, this study analyzed a dataset of 2,014 journal articles and 3,892 cited authors from the *Journal of the Architectural Institute of Korea: Planning & Design* from 2003 to 2008 in the field of Architecture in Korea. In this study, six different methods of crediting co-authors are selected for comparative analyses. These methods are first-author counting (m1), straight full counting (m2), and fractional counting (m3), proportional counting with a total score of 1 (m4), proportional counting with a total score between 1 and 2 (m5), and first-author-weighted fractional counting (m6). As shown in the data analysis, m1 and m2 are found as extreme opposites, since m1 counts only first authors and m2 assigns all co-authors equally with a credit score of 1. With correlation and multidimensional scaling analyses, among five counting methods (from m2 to m6), a group of counting methods including m3, m4, and m5 are found to be relatively similar. When the knowledge structure is visualized with pathfinder network, the knowledge structure networks from different counting methods are differently presented due to the connections of individual links. In addition, the internal validity shows that first-author-weighted fractional counting (m6) might be considered a better method to author clustering. Findings demonstrate that different co-author counting methods influence the network results of knowledge structure and a better counting method is revealed for author clustering.

초 록

학문연구에서 공저가 빈번해짐에 따라서 저자 단위의 지적 구조 분석을 수행할 때 복수저자의 기여도 산정 방식이 중요한 고려사항이 되고 있다. 이 연구에서는 복수저자 기여도 산식에 따른 상관분석, 다차원척도법, 패스파인더 네트워크의 차이를 비교 분석해보았다. 〈한국건축학회지: 계획계〉에 2003년부터 2008년까지 발표된 2,014편의 논문을 대상으로 여섯 가지 복수저자 기여도 산식을 적용해보았다. 첫째는 제1저자만 고려하는 산식(m1), 둘째는 모든 공저자 대등 산식(m2), 셋째는 균등분할 산식(m3), 넷째는 합계 1이 되는 차등 산식(m4), 다섯째는 합계 1 이상 2 이하가 되는 차등 산식(m5), 여섯째는 제1저자 가중 산식(m6)이다. 이중에서 m1은 제1저자 이외의 공저자를 모두 무시하는 반면 m2는 제1저자와 다른 공저자를 동등하게 기여도가 1인 저자로 취급하므로 두 산식이 가장 양 극단의 방식인 것으로 분석되었다. 상관분석과 다차원척도분석을 수행할 때 m1을 제외한 다섯 가지 산식(m2~m6)의 결과를 비교해본 결과 m3, m4, m5는 상대적으로 유사한 결과를 도출하는 것으로 나타났다. 그러나 패스파인더 네트워크로 지적 구조를 시각화한 결과에서는 복수저자 기여도 산식을 달리함에 따라 변경되는 한 두 링크의 차이가 전체 네트워크 구조의 현저한 차이를 낳을 수 있는 것으로 나타났다. 저자 군집에 대한 내적 타당도 측정 결과에서는 제1저자 가중 산식(m6)이 좋은 성능을 보였다. 비교 분석 결과 여섯 가지 복수저자 기여도 산정 방식 중 유사한 방식들을 구분할 수 있었으며, 특히 지적 구조를 네트워크로 표현하는 경우에 산정 방식의 차이가 더 큰 영향을 끼치는 것으로 드러났다.

Keywords: author co-citation analysis, multiple authorship, co-authorship, intellectual structure, authorship counting, pathfinder network, multidimensional scaling
저자동시인용분석, 복수저자, 공저자, 지적구조, 저자 기여도 산정방식, 패스파인더 네트워크, 다차원척도법

---

* This paper is a revised version of 8th International Conference on WIS & 13th COLLNET Meeting presentation.

** 명지대학교 문헌정보학과 부교수(memexlee@mju.ac.kr) (제1저자)
*** 이화자대학교 사회과학대학 문헌정보전공 부교수(echung@ewha.ac.kr) (교신저자)

■ 논문접수일자 : 2014년 5월 20일 ■ 최초심사일자 : 2014년 5월 28일 ■ 게재확정일자 : 2014년 6월 17일
■ 정보관리학회지, 31(2), 57-77, 2014. 〔http://dx.doi.org/10.3743/KOSIM.2014.31.2.057〕

1. Introduction

As collaboration has been recognized as a major distinctive characteristic and a growing phenomenon in modern sciences (Garvey, 1979), co-authorship among researchers is a dominant practice in a wide variety of disciplines. Since co-authorship occurs not only within homogenous environments but also across different domains of knowledge, different geographical locations, and different communities, various fields of knowledge have been influenced by collaborative research activities (Sonnenwald, 2008). In this sense, there have been a considerable number of discussions on the appropriate credit sharing of multiple authorship for a single paper. The endeavors on appropriate credit counting of a paper have been classified into primarily two perspectives. One perspective is to evaluate individual researchers' research quality as individual researchers' academic careers have been dependent on bibliometric evaluations. The other is to utilize various counting methods in the context of identification of knowledge structure, especially in context of author co-citation analysis. First, in order to evaluate individual researchers' research performance with considering the multiple authorship appropriately, there have been considerable studies including novel proposals of counting methods and comparative analyses of difference counting methods (Chung & Lee, 2002; Egghe, Rousseau, & Van Hooydonk, 2000; Gauffriau & Larsen, 2005; Hagen, 2008; Howard, Cole, & Maxwell, 1987; Kwak & Chung, 2012; Lee & Yan, 2011; Trueba & Guerrero, 2004; Rousseau & Rousseau, 1998; Van Hooydonk, 1998). In order to credit appropriately the multiple authors of a single paper, a range of counting methods have been examined, investigated, and compared in various circumstances. The second perspective on counting the multiple authors is to understand the influences of different counting methods in the context of identification of knowledge structure, particularly in the author-based analysis. As a powerful approach towards the identification of the intellectual structure of knowledge domains, author co-citation analysis (White & Griffith, 1981) has been introduced and utilized to be one of the primary techniques to map and visualize the intellectual structure of various knowledge fields using information relating to how many times two authors are co-cited by others. Depending on the groups of authors with similar specialties, the structure of specific knowledge domains may be revealed and organized through the bases of those co-cited authors. While author co-citation analysis has been utilized to map specific knowledge domains, there has been a growing concern relating to the multiple authorship (Persson, 2001; Rousseau & Zuccala, 2004; Zhao, 2006). In particular, the influences on whether specific authors are included in the citation analysis depending on different counting methods have been conducted only in a few studies (Eom, 2008; Persson, 2001; Schneider, Larsen, & Ingwersen, 2009; Zhao, 2006).

Since only first-authors of multi-authored publications have been included for the analyses of author co-citation, mostly due to the limitations of the citation index database for a considerable period of time, the earliest challengers are to demonstrate the influences by primarily comparing all the authors versus first-author only. When identifying and presenting the differences and characteristics of different crediting counting methods, however, various approaches need to be examined and compared in the context of author co-citation analysis. In this sense, as research domains have various practices of crediting co-authors' contributions, a wide range of credit counting methods should be compared for analysis. More specifically, in order to present the intellectual structures of specific fields, which are rich in co-authorship, the ways of crediting co-authors play a crucial role in determining the results and interpretation of author co-citation analysis. The influences and the impacts on the author co-citation analysis must be identified depending on different ways of crediting co-authors. When analyzing the characteristics and influences by different credit counting methods, the examinations should be viewed in prevalent author co-citation analytical techniques, including multidimensional scaling and network analyses. In the line of Kwak and Chung (2012)'s study, the purpose of this study aims to demonstrate the differences and characteristics of different credit counting for co-authors in the context of author co-citation analysis. In particular, this study attempts to identify the characteristics of different credit counting methods for co-authors including correlation coefficients, multidimensional scaling, and pathfinder networks.

2. Related Studies

As co-authorship is prevalent in research communities, crediting co-authors is one of the crucial factors in the context of identifying the intellectual structure of fields. The way of crediting co-authors in multiple authorship affects the results of various bibliometric analyses from *h*-index to the domain analysis. When counting all authors with full credit, there is an inflationary factor on the authors who simply co-authored with other researchers, but were not directly involved with the core work of the publication. On the other hand, when dividing the credit of one publication equally among all co-authors, there is an equalizing bias among the authors who contributed more to the publications (Hagen, 2008). Since some authors might contribute more to the publications than others, those who contribute more need to be considered with higher credit scores. Consequently, there have been considerable endeavors to demonstrate the differences and influences of different credit counting methods in the context of multiple authorship (Egghe, Rousseau, & Van Hooydonk, 2000; Gauffriau & Larsen, 2005). Since various bibliometric indicators can be affected by the ways of counting co-authors (Egghe, Rousseau, & Hoydoonk, 2000; Price, 1981), novel ways of counting co-authorship have been proposed and examined. In addition to the studies on various counting methods, there have been approaches to investigate the influences by different ways of counting, primarily first-only versus all-authors. In this sense, the related studies can be categorized into two primary groups: (1) the introductions on novel credit counting methods, and (2) the investigations on the influences of author-based bibliometric analyses depending on different credit counting methods.

In terms of introducing novel counting methods for crediting co-authors in multiple authorship, various efforts have been conducted (Chung & Lee, 2002; Hagen, 2008; Howard, Cole, & Maxwell, 1987; Trueba & Guerrero, 2004). In terms of assessing the productivity of researchers in the field of psychology, Howard, Cole, and Maxwell (1987) proposed a proportional counting method reflecting the order of author names in the byline of a paper. According to Howard *et al*., the credit score of an author is calculated by a formula, credit $= (1.5^{n-i})/(\sum_{i=1}^{n} 1.5^{i-1})$, where n is the total number of authors and i is the particular author's ordinal position. The sum of all credit scores given to co-authors is equal to 1. The findings show that there is a much higher correlation between productivity and reputational ranking, when the credit scores are calculated by the proportional counting method. In a similar line of research to Howard *et al*., Chung and Lee (2002) proposed another proportional counting method, credit $= 3n + 1 - 2i / n(n+1)$, where n is the total number of authors and i is the particular author's ordinal position in the byline of a paper. The difference from Howard *et al*.'s proportional counting method is that the sum of all scores given to co-authors is equal to the value between 1 and 2. The reason for a higher sum of credit scores, which is between 1 and 2, is to reflect the higher impact of co-authored papers compared to single authored papers (Glanzel, 2002). Chung and Lee compared the differences between all author co-citation and first author co-citation in the domain of Information Science and Computer Science with using the proposed proportional counting method. They point out that certain areas of intellectual structures are mapped differently, depending on whether all authors were included or only first authors are included. Hagen (2008) proposed a harmonic authorship crediting system allocation, which is i^{th} author credit $= (1/i) / [1 + (1/2) + \cdots + (1/N)]$ for the ith author of a publication with N co-authors. The idea of the harmonic authorship crediting system is based on three aspects: (1) all authors can share one publication credit, (2) the hierarchical positions are important and (3) more authors in one publication should result in less credit per author. In the fields of Medicine, Psychology, and Chemistry, this study shows that a harmonic authorship credit allocation is close to the data results which are surveyed from those three fields. Park and Jeong (2013) showed the domain of Library and Information Science field in Korea using this method. On the other hand, Tscharntke et al. (2007) discussed the domain practices in the field of the ecological and environmental science, the last authors often receive as much credit as the first author, in recent years. It is because they are assumed to be the driving forces, both intellectually and financially, for specific research projects. First or last authors, however, mistakenly benefit frequently since interpretation of author sequence in the byline of a paper can be arbitrary. The authors of this study proposed to utilize the acknowledgement section for specifying four different ways of crediting in order to reflect the contributions of co-authors: (1) the sequence-determined-credit approach, (2) the equal contribution approach, (3) the first-last-author emphasis approach and (4) the percent-contribution-indicated approach. On the other hand, although *h*-index has been introduced and accepted in various fields in order to capture the quality and quantity of individual researchers, it does not take into account the multiple authorship, which is dominant in diverse domain areas. Prathap (2011) proposed an index, fractional and harmonic *p*-indices, which are extended for research assessment to reflect the multiple authorship in terms of *h*-index. The fractional and harmonic *p*-indices can be applied without the inflationary and equalizing bias of full credits or fractional credits for co-authors, respectively.

As author co-citation analyses have been performed with the data sets from various fields, considerable studies investigated the different results of knowledge structure depending on different counting methods (Eom, 2008; Persson, 2001; Schneider, Larsen, & Ingwersen, 2009; Zhao, 2006; Zhao & Strotmann, 2011). According to Schneider et al. (2009), Persson (2001), one of the earliest challengers on first-author only practices, analyzed the differences in terms of the multidimensional scaling analysis between straight full counting and first-author counting for the author co-citation analysis of a data set of information science from 1986 to 1996. Although only a few differences were distinguished in the maps of author co-citation depending on the different counting methods, large numbers of authors, co-authors, were ignored since only first authors were included for analysis. Similarly, in the field of a decision support system from 1971 to 1990, Eom (2008) compared the differences between all co-authors and only first-authors using a data set with a total of 692 articles. He pointed out that the knowledge structure including all co-authors showed all influential authors and presented the detailed knowledge structure maps reflecting sub-domains. Zhao (2006) analyzed three different ways of including co-authors in terms of author co-citation analysis. From a data set of 312 publications and 4,578 references in the XML research field, Zhao pointed out that all-author counting showed more coherent groups of authors, while fewer specialties than traditional first-author only counting were found in the research field. On the other hand, in a data set of IEEE collection, Schneider, Larsen, and Ingwersen (2009) compared the results between "inclusive all-authors" (Zhao, 2006) and first authors based author co-citation analyses. By utilizing different analytical methodologies, which include the multidimensional scaling and factor analysis, the findings showed that all authors can present coherent intellectual structures of the fields; yet, they are not able to support the notion that all authors can present more detailed sub-domains of the fields. Recently, Zhao and Strotmann (2011) proposed a last-author-weighted counting method by presenting the mapping results compared to the traditional first-author counting and all-author counting. Analyzing a dataset of stem cell research from PubMed and Scopus, which is a highly collaborative field, Zhao and Strotmann demonstrated that the mapping results based on different counting methods showed similar overall knowledge structures. They found that there were different detailed structures and specialties depending on different counting methods to various degrees.

It is a crucial issue regarding how to count co-authors in context of author co-citation analysis since co-authoring has been one of the primary practices in diverse research communities. In this sense, credit counting for co-authors has been studied in a wide range from research evaluation to knowledge structure identification. As the related studies demonstrate, while it has been relatively well explored to propose appropriate co-author credit counting methods in a way of evaluating researchers' productivity (Zhao & Strotmann, 2011), there have been mixed and unclear findings regarding the influences of different co-author counting methods on author co-citation analysis for knowledge structure. In addition, when the influences on knowledge structure were discussed depending on different counting methods, the discussions have been mostly limited in first-author versus all-author. Analysis tools, on the other hand, have been limited in traditional multidimensional scaling and factor analysis, although network-based analysis such as pathfinder network has been recently introduced and considerably used to identify the knowledge structure of field. In this sense, the purpose of this study is to compare different counting methods for co-authors and examine the differences of knowledge structure in terms of pathfinder network depending on different counting methods. For comparative analysis on author co-citation, counting methods for co-authors are selected to reflect various co-author contribution practices in research communities. In addition, the pathfinder networks of author co-citation analysis are comparatively analyzed to examine the differences of the knowledge structure depending on different counting methods.

3. Methodology

In order to achieve research objectives of this study, a set of data is prepared from the study of Ryoo & Choi (2011). In the field of Architecture in Korea, *Journal of the Architectural Institute of Korea: Planning & Design* is selected from 2003 to 2008 since it contains comprehensive topics in the domain of Architecture. A total of 2,014 articles are collected; the total number of cited authors is 3,892. According to the numbers of citation frequency, 47 most frequently cited authors are selected for analyses. When authors appeared more than once in the reference list of a single article, the minimum numbers of citation are selected because the entire citation scores for an author should not be over-valued. Six methods of crediting co-authors are shown in Table 1 and the examples from six methods are presented in Table 2. The first method is called first-author counting (m1), since only the first author of the n authors of a paper receives a credit equal to one and the other authors do not receive any credit (Cole & Cole, 1973). This method has been substantially used mainly because of the citation index such as the Science Citation Index. The second method is straight full counting (m2) and not affected by the order of author names in the byline of a paper, since each of the *n* authors of a paper receives one credit. This method is also called normal or standard counting (Egghe, Rousseau, Van Hooydonk, 2000). Since all co-authors are credited with one credit score without a maximum credit score, the straight full counting (m2) has been criticized with inflated co-author credits (Lindsay, 1980). For the third method, as fractional counting (m3), each of the *n* authors receives a credit score equal to 1/n (Price, 1981), which can maintain a maximum credit score equal to one. The fractional count (m3) is also called adjusted counting and not affected by the order of author names in the byline of a paper either. In a way of taking into consideration the order of author names in the byline of a paper, several counting methods have been introduced. As one of those counting methods, proportional counting (m4) is proposed by Howard et al. (1987). Proportional counting reflects the order of co-author names in the byline of a paper, and the sum of all credit scores to each of all co-authors is equal to 1. According to the particular author's ordinal position, each of co-authors receives proportional credit scores. For instance, in a three-author paper, first author receives 0.47; second author receives 0.32; and third author receives 0.21. The sum of all credit scores for this example is equal to 1. In a similar line of the fourth method as proportional counting, the fifth method is proposed by Chung and Lee (2002). Proportional counting (m5) not only takes into consideration the particular author's ordinal position in the byline of a paper, but assigns the total credit score of a single paper to be between 1 and 2. It is based on the fact that a paper with multiple authors is likely to be more influential than a paper with a single author (Glanzel, 2002). For instance, in a three-author paper, according to m5, first author receives 0.67; second author receives 0.5; and third author receives 0.33. The sum of all credit scores is equal to 1.5, which is greater than the sum of all credit scores from m4. For the sixth method (m6), first-author-weighted fractional counting is sort of a combination between first-author counting and fractional counting. The first author is given a full credit score, which is equal to 1, and the rest of co-authors receives credit scores with *1/n*, where *n* is the total number of authors of a paper. For instance, in a three-author paper, first author receives 1 as full credit score; second author receives 0.33; and third author receives 0.33. The sum of all credit scores for this example is equal to 1.66. The sum of all credit scores from m6 increases without any maximum value as the number of co-authors increases.

For this study, six different counting methods are selected and comparatively analyzed with three phases of analysis. In the first, as an overall comparative analysis, two comparison analyses are performed; Pearson's and Spearman's rank correlation and multidimensional scaling analysis. The Pearson's and Spearman's rank correlations are computed among author citation counts from six different methods. In addition, six different counting methods are located in a two-dimensional space from multidimensional scaling analysis. After conducting the first phase of analysis, the method of first-author counting (m1)

〈Table 1〉 Six methods of counting multiple co-authors

Notation	Description	Counting method
m1	First author counting	Credit 1 only to first authors
m2	Straight full counting	Credit 1 to all co-authors
m3	Fractional counting	Credit 1/n to all *n* authors, where *n* is the total number of authors
m4	Proportional counting with a total score of 1	Credit $1.5^{n-i}/\Sigma 1.5^{i-1}$ to all co-authors, where *n* is the total number of authors and *i* is the particular author's ordinal position in the byline of a paper (Howard, Cole & Maxwell, 1987)
m5	Proportional counting with a total score between 1 and 2	Credit $(3n+1-2i)/n(n+1)$, where *n* is the total number of authors and *i* is the particular author's ordinal position in the byline of a paper (Chung & Lee, 2002)
m6	First-author-weighted fractional counting	Credit 1 to first author and credit *1/n* to the remaining co-authors, where *n* is the total number of authors

〈Table 2〉 Credit counting example with a three-author paper

Notation	Description	1st author	2nd author	3rd author	Total credit score	Maximum credit score
m1	First author counting	1	0	0	1	1
m2	Straight full counting	1	1	1	3	No Maximum
m3	Fractional counting	0.33	0.33	0.33	1	1
m4	Proportional counting with a total score of 1	0.47	0.32	0.21	1	1
m5	Proportional counting with a total score between 1 and 2	0.67	0.5	0.33	1.5	Between 1 and 2
m6	First-author-weighted fractional counting	1	0.33	0.33	1.66	No Maximum

is excluded because the list of 47 authors is different from those of the other five methods. As pointed out by Zhao and Strotmann (2011), author selection has a great effect on the mapping results of the author co-citation analysis. Specifically for this study, two authors are replaced from the list of 47 authors produced by first author counting method (m1). One author is found as co-authored only, implying that the author does not publish any papers as first author in this data set. The other author is replaced because the author is not co-cited with other authors in this data set. In the second phase, the knowledge structures using multidimensional scaling analysis are visualized with 47 authors. Then, the distances of selected authors from each of the multidimensional scaling maps are comparatively analyzed using correlation and dispersion analysis. In the third phase, pathfinder network analyses are performed with different counting methods. The pathfinder networks are manually examined and compared in terms of the degree of link duplicate and shortest path distance.

4. Results

4.1 Comparison with citation scores

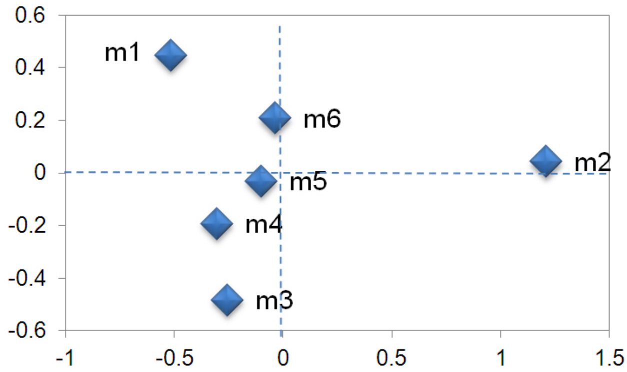
In the first phase of analysis, the Pearson's and Spearman's correlation coefficients from the co-citation scores of 47 authors are shown in Table 3 in order to compare six different counting methods. As shown in the correlation coefficients, three aspects are pointed out as follows. First, the traditional first author only counting method (m1) is highly correlated with three methods such as first-author-weighted fractional counting (m6), proportional counting with a total score between 1 and 2 (m5), and proportional counting with a total score of 1 (m4). Since the methods of m6, m5, and m4 commonly weigh more on first authors, strong correlations are found with m1. Second, the method of crediting 1 to all co-authors (straight full counting, m2), is highly correlated with the methods of proportional counting with a total score between 1 and 2 (m5) and first-author-weighted fractional counting (m6). One of possible reasons for this strong correlation among three methods is that they can reach to the total credit score, which is greater than 1. Third, the method of fractional counting (m3) is highly correlated with both proportional counting methods (m4 and m5).

On the other hand, six different counting methods are presented in a two-dimensional space using multidimensional scaling analysis as depicted in Figure 1. The methods of m1 and m2 are located at extreme opposite positions from each other. On the other hand, the methods of m3, m4, m5, and m6 are located closely.

From the results on correlation coefficients and multidimensional scaling, four methods, fractional counting (m3), proportional counting with a total score of 1 (m4), proportional counting with a total score between 1 and 2 (m5), and first-author-weighted fractional counting (m6), are considerably similar.

〈Table 3〉 Pearson's correlation coefficient (upper half) and Spearman's rank correlation coefficient (lower half) among the six methods

	m1	m2	m3	m4	m5	m6
m1		.674	.817	.908	.928	.955
m2	.672		.751	.743	.836	.846
m3	.758	.785		.982	.950	.875
m4	.864	.772	.977		.981	.934
m5	.882	.872	.939	.966		.979
m6	.927	.853	.859	.914	.968.	



<Figure 1> Multidimensional scaling (PROXSCAL) map among the six different counting methods

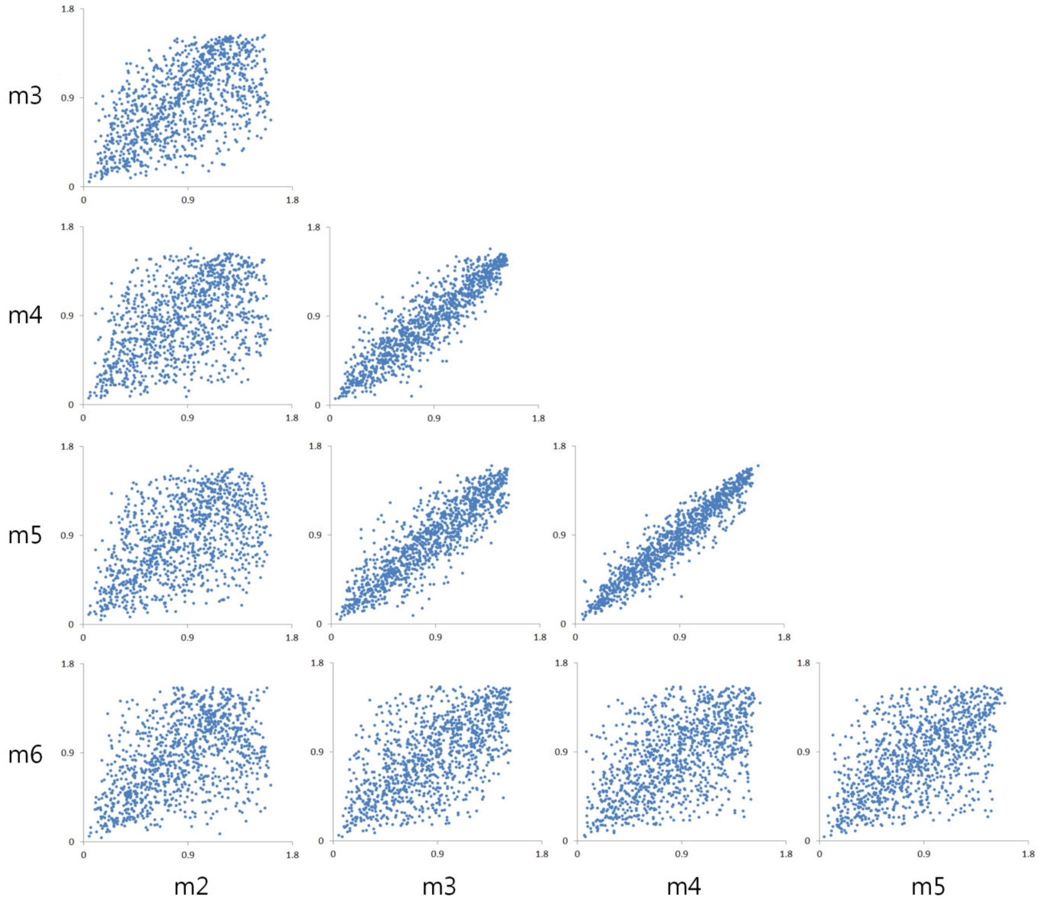
4.2 Comparison with multidimensional scaling maps

For the second phase of analysis, five methods except m1 are compared since the list of 47 authors for m1 is different from the lists of the rest five methods. For the five different counting methods, the maps of multidimensional scaling (PROXSCAL) are analyzed. The distances of the 47 authors on the five multidimensional scaling maps are calculated and computed as Pearson's and Spearman's correlation coefficient as presented in Table 4. Similar to the previous results, the fractional counting method (m3) is highly correlated with both proportional counting methods (m4 and m5).

In addition, the dispersions among the distances of authors on the five multidimensional scaling maps are analyzed as shown in Figure 2. Clearly, three pairs of methods are shown with relatively high correlations. The method of fractional counting (m3) is highly correlated with proportional counting with a total score of 1 (m4) and proportional counting with a total score between 1 and 2. In addition, both proportional counting methods (m4 and m5) are high

<Table 4> Pearson's correlation coefficient (upper half) and Spearman's rank correlation coefficient (lower half) among five multidimensional scaling maps

	m2	m3	m4	m5	m6
m2		.595	.506	.503	.498
m3	.593		.907	.884	.613
m4	.501	.908		.949	.556
m5	.495	.884	.949		.571
m6	.493	.609	.549	.562	



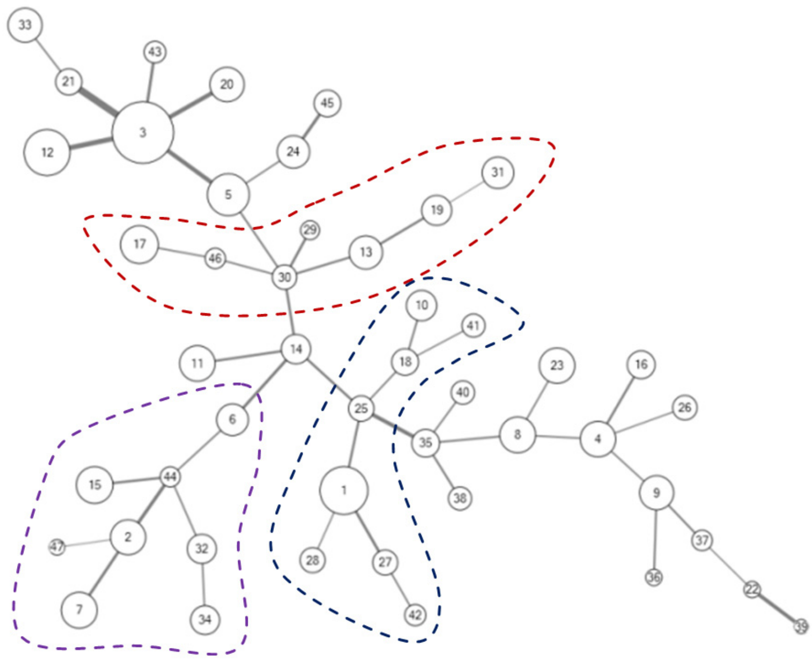
<Figure 2> The dispersions among the distances on five multidimensional scaling maps

ly correlated with each other.

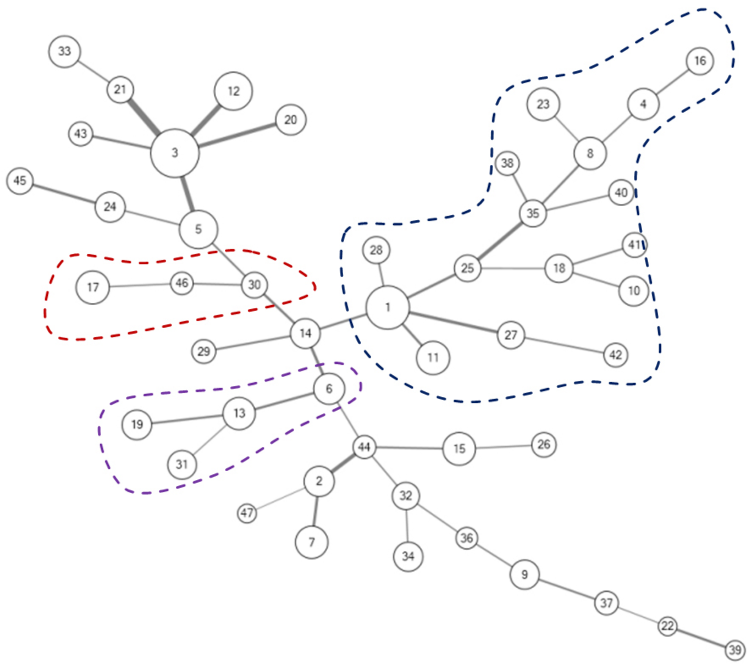
Based on the analyses of correlation of authors' citation scores and the distances of authors in the maps of multidimensional scaling, a group of counting methods is found to be similar. The methods of fractional counting (m3), proportional counting with a total score of 1 (m4), and proportional counting with a total score between 1 and 2 (m5) show a strong correlation.

4.3 Comparison with pathfinder networks

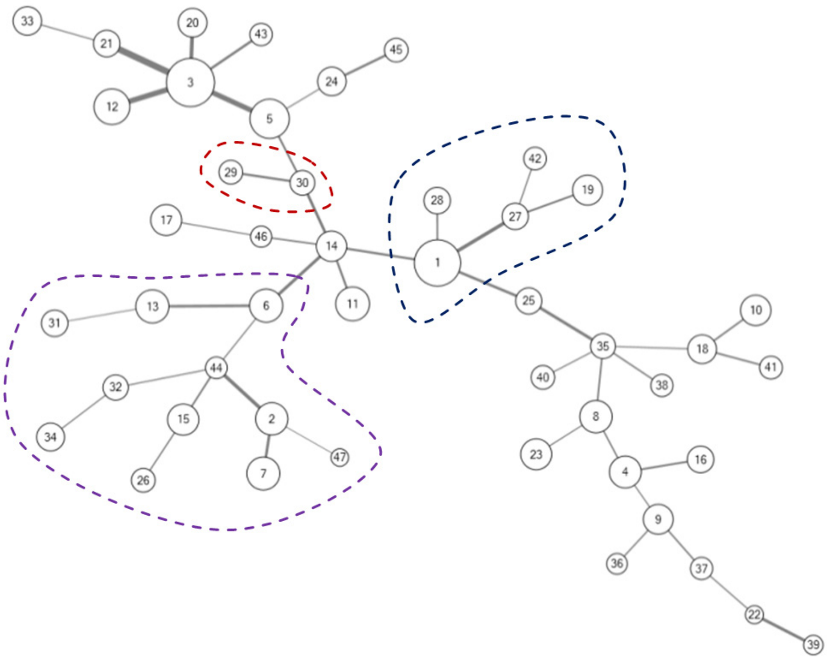
In the third phase of analysis, the selected 47 authors are presented with the pathfinder networks to visualize the knowledge structure depending on five different counting methods. The individual pathfinder networks are shown in Figure 3, Figure 4, Figure 5, Figure 6, and Figure 7, respectively. In these networks, the size of nodes represents the frequency



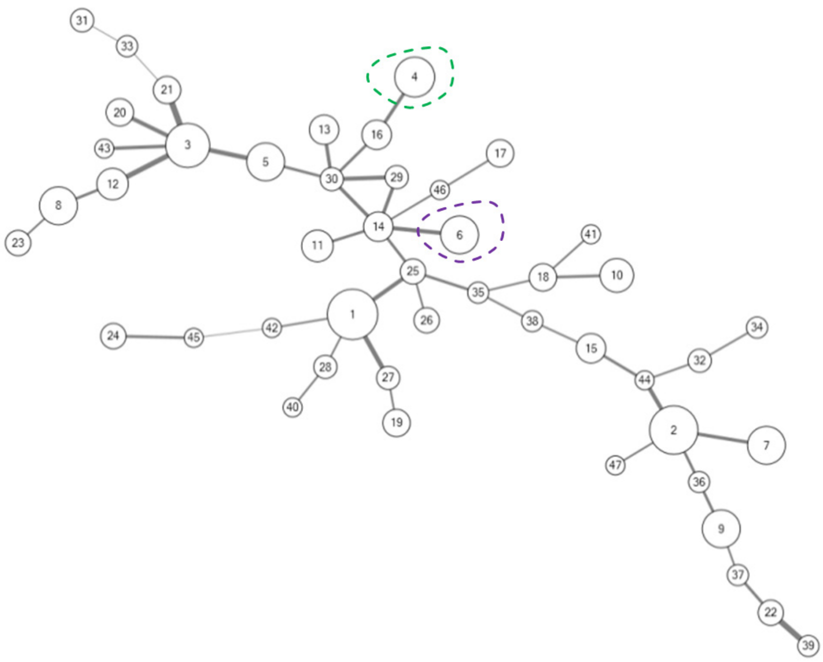
〈Figure 3〉 Pathfinder network (p=n-1, r=∞) of m3



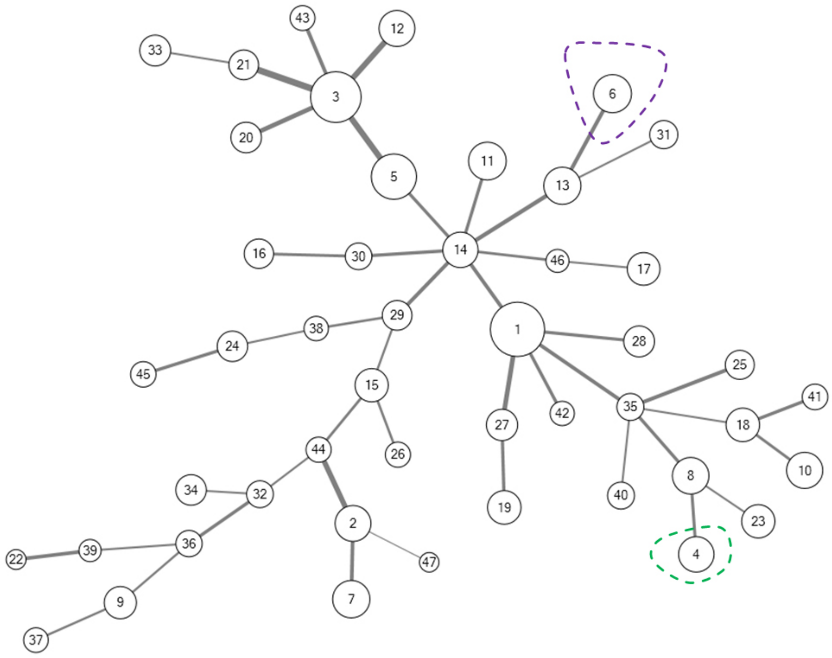
〈Figure 4〉 Pathfinder network (p=n-1, r=∞) of m4



〈Figure 5〉 Pathfinder network (p = n - 1, r = ∞) of m5



〈Figure 6〉 Pathfinder network (p = n - 1, r = ∞) of m2



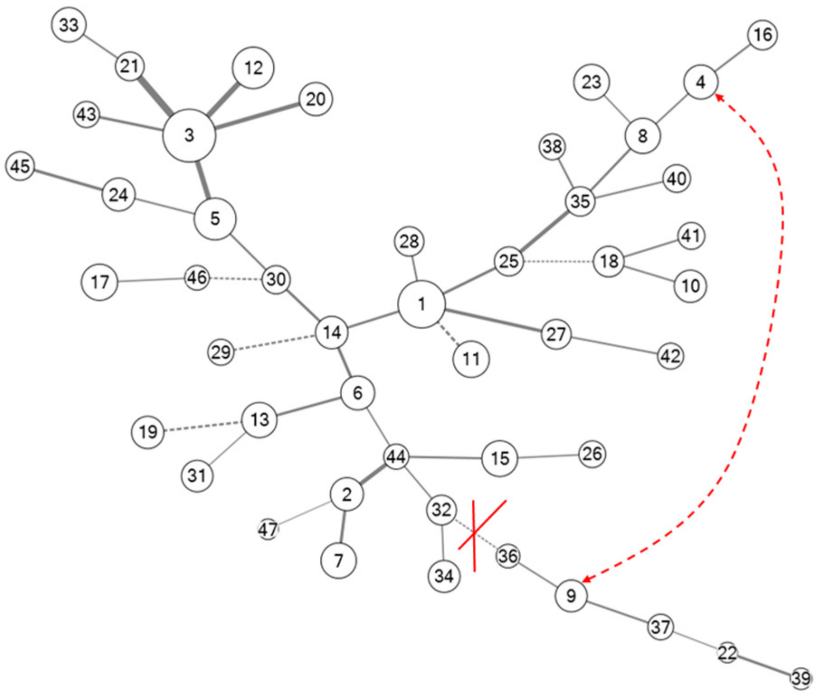
<Figure 7> Pathfinder network (p=n-1, r=∞) of m6

of citation, and the thickness of links proportionally shows the number of co-citations.

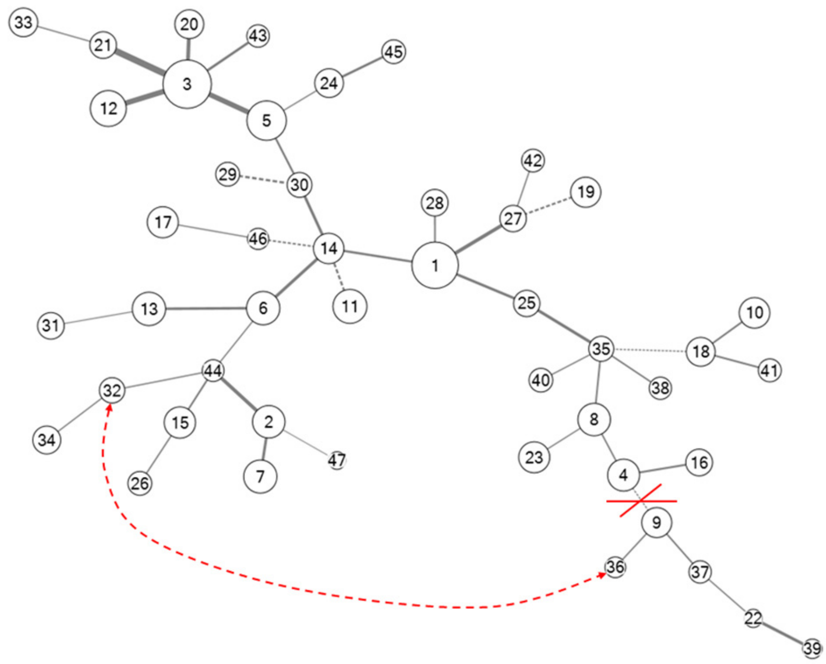
From the first and second phases of analysis, three methods, fractional counting (m3), proportional counting with a total score of 1 (m4), and proportional counting with a total score between 1 and 2 (m5), are identified to be highly correlated. In order to compare these three methods in terms of examining the pathfinder networks, three pathfinder networks (Figure 3, Figure 4, and Figure 5) are manually examined and shows distinctive differences. Three nodes with connected branches shown with dotted circles in Figure 3, Figure 4, and Figure 5 are notable for comparison. First, the node 30 shows substantial differences among m3, m4, and m5. In m3, the node 30 is connected to three branches: with the nodes 46, and 17, with the nodes 29, and with the nodes 13, 19, and 31. However, in m4, the node 30 is connected to only one branch, which contains the nodes 46 and 17. The branch containing the nodes 13, 19, and 31 is connected to the node 6, not the node 30. In m5, the node 30 is connected to only one single node, which is the node 29. Second, the node 1 with connected branches shows drastic dissimilarities among three methods. In m3, the node 1 is connected to three branches; first with the node 28, second with the nodes 27, and 42, and third with the nodes 28, 18, 10, and 41. On the other hand, in the network for m4, the node 1 is connected with four branches; first with the node 28, second with the node 11, third with the nodes 27, and 42, and fourth with the nodes 25, 18, 41, 10, 35, 38, 40, 8, 23, 4, and 16. For m5, the connection with the node 1 is much simpler compared to the network of m4. The node 1 is connected with two branches; one is the node 28 and the other is the nodes 27, 19, and 42. Third, the node 6 is connected with the nodes 44, 15, 2, 47, 7, 32, and 34 for the network of m3. For the network of m4, the connections with the node 6 include only three nodes, which are the nodes 13, 19, and 31. However, in the network of m5, the node 6 is connected with two branches, one is with the nodes 13 and 31 and the other with ten nodes, which are the nodes 13, 31, 44, 32, 34, 15, 26, 2, 7, and 47.

On the other hand, two networks from m2 and m6 commonly show two nodes, 4 and 6, as a leaf, while three networks from m3, m4, m5 present the two nodes as a member node of a branch as shown in Figure 6 and Figure 7.

Manual examinations of the pathfinder networks indicate substantial differences of knowledge structure depending on different counting methods. One of possible explanations regarding different networks with different counting methods is due to link connections. One instance is shown in Figure 8 and Figure 9 for the change in the networks due to a single link connection. While the link between the nodes 32 and 36 is connected in the network of m4, it is disconnected in the network of m5. Another instance is the nodes 4 and 9. They are connected in the network of m4, but are disconnected in the



〈Figure 8〉 Pathfinder network of m4 specified by the differences compared to m5



<Figure 9> Pathfinder network of m5 specified by differences compared to m4

network of m5.

In order to systematically examine the networks, two comparisons are conducted; one is the degree of link duplicate and the other is the shortest path distance. To identify the similarity of local areas between pathfinder networks, the degree of link duplicate is measured in terms of Dice coefficient. On the other hand, the shortest path distance is calculated in order to see if there are similar global structures between two networks.

As portrayed in Table 5, three methods, m3, m4, and m5, are found to be relatively similar to each other in terms of Dice coefficients. This finding is consistent to the results of the first and second phases of analysis including correlations of authors' citation scores, and the distances of authors in the multidimensional scaling maps. In addition, regarding the similarity of the global structure, the shortest path distances are calculated and compared in terms of correlation. As shown in Table 6, the results demonstrate similar results among the five pathfinder networks. The global structures of m3, m4, and m5 networks are found to be similar to each other, while m2 and m6 are found to be relatively different to the three methods, m3, m4, m5. One of the notable features in the analysis of global structures is that some degree of similarity exists between m2 and m6.

〈Table 5〉 Dice coefficients among the links of five pathfinder networks

	m2	m3	m4	m5
m3	0.731			
m4	0.667	0.848		
m5	0.753	0.848	0.870	
m6	0.667	0.609	0.717	0.761

〈Table 6〉 Pearson's correlation coefficient (upper half) and Spearman's rank correlation coefficient (lower half) of the shortest path distance among the five methods

	m2	m3	m4	m5	m6
m2		.571	.653	.521	.762
m3	.552		.700	.874	.579
m4	.607	.719		.733	.791
m5	.498	.857	.774		.598
m6	.742	.594	.778	.631	

4.4 Comparison of the quality of author clusters

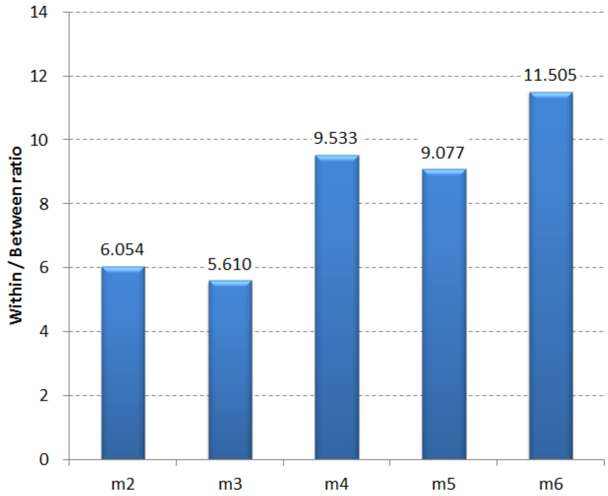
In order to assess the differences of five methods to count co-authors, the quality of clusters from the co-citation maps is assessed through internal cluster validity. For internal cluster validity, the within and between ratio is applied, which is average co-citation scores within the same clusters divided by average co-citation scores of other clusters.

internal cluster validity

$$= \frac{\textit{average within cluster author cocitation score}}{\textit{average between cluster author cocitation score}}$$

For five methods from the methods m2 to m6, seven clusters depending on author co-citation score matrix are revealed using Ward clustering method. As shown in Figure 10, the internal validity scores are demonstrated from five author counting methods, m2, m3, m4, m5, and m6.

As indicated in Figure 10, the method weighing more score on first author, which is m6, shows the highest score of internal validity. In contrast, the methods m2 and m3 assigning the same score to all co-authors show the lowest scores of internal validity. This result indicates that first author weight method with proportional co-author score may reveal the clusters of authors more appropriately.



<Figure 10> Internal validity scores for five author counting methods

5. Discussion and Conclusion

Co-authoring has been the fastest growing phenomenon in research communities. To understand the knowledge structures of subject fields, author co-citation analysis needs to appropriately reflect the highly collaborative practices in various research communities. With a growing concern on multiple authorship and different ways of counting in context of citation analysis (Persson, 2001; Rousseau & Zuccala, 2004; Zhao, 2006), this study aims to identify and demonstrate the influences of different credit counting methods in context of author co-citation analysis. As there have been considerable endeavors to propose proper counting methods, this study selects six different counting methods: first author counting (m1), straight full counting (m2), fractional counting (m3), proportional counting with a total score of 1 (m4), proportional counting with a total score between 1 and 2 (m5), and first-author-weighted fractional counting (m6). For the purpose of this study, the comparative analyses were conducted by correlation analysis of author citation counts, the distances of authors in the multidimensional scaling maps, and examinations on the pathfinder networks depending on six different counting methods.

First, as previous studies showed that there are differences between first-author-only and all-author (Eom, 2007; Zhao, 2006; Zhao & Strotmann, 2011), first author counting method (m1) is found to be substantially different from the rest of counting methods in this current study. Among five different counting methods, some credit counting methods are clustered from the results of three analysis phases. Three counting methods, fractional counting (m3), proportional counting with a total score of 1 (m4), proportional counting with a total score between 1 and 2 (m5), are found to be relatively similar. On the other hand, straight full counting (m2) and first-author-weighted fractional counting (m6) are found to be comparatively dissimilar with the three methods, m3, m4, and m5. This finding may suggest that although all co-authors are included for author co-citation analysis, considerable potentials that influence the knowledge structures of fields still exist depending on how to count co-authors in terms of author co-citation analysis. Second, the pathfinder networks of different counting methods are examined. Manual examinations on the pathfinder networks indicate that there are considerable differences of the networks depending on different counting methods. The differences identified in the networks even exist among three methods, m3, m4, and m5, which are found to be consistently similar. Depending on the counting methods, the dissimilarities among the networks are found as to whether specific individual nodes are connected to others or not. Such differences can be explained due to the individual link connections between the nodes. As a result, connections among the nodes affect the overall network structures according to the different ways of counting. Since presenting and interpreting the results of knowledge structures are primarily based on manual examinations in the practices of network-based analysis, the differences in the structure of networks might have a substantial impact on the knowledge structure results. Moreover, the internal validity shows that the method weighing more on first author with proportional scoring might be considered a better method to author clustering. The findings of this study have implications on the process of knowledge structure identification with author co-citation analysis. When the intellectual structures of fields are revealed by using author co-citation analysis, the differences and characteristics of different counting methods should be considered in the context of analysis and interpretation.

Acknowledgments

We thank Jong-Duk Ryoo and Eun-Ju Choi for the data collection, and an anonymous reviewer for helpful comments.

References

Chung, Y. M., & Lee, E. S. (2002). A co-citation analysis of multiple authorship in the subject field of Information Science and Computer Science. Journal of Knowledge Processing and Management, 3(2), 1-26. Retrieved from http://jkpm.yonsei.ac.kr/fulltext/v3n2a1.pdf

Cole, J. R., & Cole, S. (1973). Social stratification in science. Chicago: The University of Chicago Press.

Egghe, L., Rousseau, R., & Van Hooydonk, G. (2000). Methods for accrediting publications to authors or countries: Consequences to evaluation studies. Journal of the American Society for Information Science, 51(2), 145-157.

Eom, S. (2008). All author cocitation analysis and first author co-citation analysis: A comparative empirical investigation. Journal of Informetrics, 2(1), 53-61. http://dx.doi.org/10.1016/j.joi.2007.09.001

Garvey, W. D. (1979). Communication, the Essence of Science: Facilitating Information Exchange among Librarians, Scientists, Engineers, and Students. Oxford: Pergamum Press.

Gauffriau, M., & Larsen, P. O. (2005). Counting methods are decisive for rankings based on publication and citation studies. Scientometrics, 64(1), 85-93. http://dx.doi.org/10.1007/s11192-005-0239-6

Glanzel, W. (2002). Coauthorship patterns and trends in the Sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies. Library Trends, 50(3), 461-473.

Hagen, N. T. (2008). Harmonic allocation of authorship credit: Source-level correction of bibliometric bias assures accurate publication and citation analysis. PLoS ONE, 3(12), e4021. http://dx.doi.org/10.1371/journal.pone.0004021

Hagen, N. T. (2010). Harmonic publication and citation counting: Sharing authorship credit equitably-not equally, geometrically or arithmetically. Scientometrics, 84(3), 785-793. http://dx.doi.org/10.1007/s11192-009-0129-4

Howard, G. S., Cole, D. A., & Maxwell, S. H. (1987). Research productivity in psychology based on publication in the journals of American Psychology Association. American Psychologist, 42(11), 975-986.

Kwak, Sun-Young, & Chung, EunKyung (2012). Domain analysis on economics by utilizing cocitation analysis of multiple authorship. Journal of the Korean Society for Information Management, 29(1), 115-134. http://dx.doi.org/10.3743/KOSIM.2012.29.1.115

Lee, Jongwook, & Yang, Kiduk (2011). A bibliometric analysis of faculty research performance assessment methods. Journal of the Korean Society for Information Management, 28(4), 119-140. http://dx.doi.org/10.3743/KOSIM.2011.28.4.119

Lindsey, D. (1980). Production and citation measures in the sociology of science: The problem of multiple authorship. Social Studies of Science, 10, 145-162.

Perssson, O. (2001). All author citations versus first author citations. Scientometrics, 50(2), 339-344.

Prathap, G. (2011). The fractional and harmonic p-indices for multiple authorship. Scientometrics. 86, 239-244. http://dx.doi.org/10.1023/A:1010534009428

Price, D. D. S. (1981). Multiple authorship. Science, 212, 986.

Park, Ji Yeon, & Jeong, Dong Youl (2013). A study on the intellectual structure of library and information science in Korea by author bibliographic coupling analysis. Journal of the Korean Society for Information

Management, 30(4), 31-59. http://dx.doi.org/10.3743/KOSIM.2013.30.4.031

Ryoo, Jong-Duk, & Choi, Eun-Ju (2011). A comparison test on the potential utility between author profiling analysis(APA) and author co-citation analysis (ACA). Journal of the Korean Society for Information Management, 28(1), 123-144. http://dx.doi.org/10.3743/KOSIM.2011.28.1.123

Rousseau, S., & Rousseau, R. (1998). The scientific wealth of European nations: Taking effectiveness into account. Scientometrics, 42, 75-87.

Rousseau, R., & Zuccala, A. (2004). A classification of author co-citations: Definitions and search strategies. Journal of the American Society for Information Science and Technology, 55(6), 513-529. http://dx.doi.org/10.1002/asi.10401

Schneider, J., Larsen, B., & Ingwersen, P. (2009). A comparative study of first and all-author co-citation counting, and two different matrix generation approaches applied for author co-citation analyses. Scientometrics, 80(1), 103-130. http://dx.doi.org/10.1007/s11192-007-2019-y

Sonnenwald, D.H. (2008). Scientific collaboration. Annual Review of Information Science and Technology, 41(1), 643-681. http://dx.doi.org/10.1002/aris.2007.1440410121

Trueba, F., & Guerrero, H. (2004). A robust formula to credit authors for their publications. Scientometrics, 60(2), 181-204. http://dx.doi.org/10.1023/B:SCIE.0000027792.09362.3f

Tscharntke, T., Hochberg, M. E., Rand, T. A., Resh, V. H., & Krauss, J. (2007). Author sequence and credit for contributions in multiauthored publications. PLoS Biology, 5(1), 13-14. http://dx.doi.org/10.1371/journal.pbio.0050018

Van Hooydonk, G. (1998). Standardizing relative impacts: Estimating the quality of research from citation counts. Journal of the American Society for Information Sciences, 49, 932-941.

White, H. D., & Griffith, B. C. (1981). Author co-citation: A literature measure of intellectual structure. Journal of the American Society for Information Science, 49(4), 327-355.

Zhao, D. (2006). Towards all-author co-citation analysis. Information Processing and Management, 42, 1578-1591. http://dx.doi.org/10.1016/j.ipm.2006.03.022

Zhao, D., & Strotmann, A. (2011). Counting first, last, or all authors in citation analysis: A comprehensive comparison in the highly collaborative stem cell research field. Journal of the American Society for Information Science and Technology, 62(4), 654-676. http://dx.doi.org/10.1002/asi.21495