The Distinct Impact Dimensions of the Prestige Indices in Author Citation Networks*

저자 인용 네트워크에서 명망성 지표의 차별된 영향력 측정기준에 관한 연구

Hyerim Ahn (안혜림)**
Ji-Hong Park (박지홍)***

ABSTRACT

This study aims at proposing three prestige indices－closeness prestige, input domain, and proximity prestige- as useful measures for the impact of a particular node in citation networks. It compares these prestige indices with other impact indices as it is still unknown what dimensions of impact these indices actually measure. The prestige indices enable us to distinguish the most prominent actors in a directed network, similar to the centrality indices in undirected networks. Correlation analysis and principal component analysis were conducted on the author citation network to identify the differentiated implications of the three prestige indices from the existing impact indices. We selected simple citation counting, h-index, PageRank, and the three kinds of centrality indices which assume undirected networks as the existing impact measures for comparison with the three prestige indices. The results indicate that these prestige indices demonstrate distinct impact dimension from the other impact indices. The prestige indices reflect indirect impact while the others direct impact.

초 록

본 연구는 명망성 지표(closeness prestige, input domain, proximity prestige)를 인용 네트워크 내에서 특정 노드의 영향력 측정을 위한 유용한 척도로 제안하는 것을 목적으로 한다. 명망성 지표의 영향력 측정기준에 대해 알려진 바가 없으므로 본 연구는 이 세 개의 명망성 지표와 다른 영향력 지표를 비교하고자 한다. 무방향 네트워크의 중심성 지표와 유사하게 명망성 지표는 유방향 네트워크에서 특정 중심 노드들을 차별화 시켜준다. 저자 인용 네트워크에서 수행된 상관분석과 주성분분석을 통하여 본 연구는 기존 영향력 지표와 차별된 명망성 지표만의 측정기준을 발굴하였다. 세 개의 무방향 네트워크 중심성 지표와 더불어 단순인용수, h-index, PageRank를 본 연구에서 제시한 명망성 지표와 비교하였다. 이러한 명망성 지표는 기존 영향력 지표와는 차별화된 영향력을 측정하고 있다는 결과를 도출하였으며 명망성 지표는 간접적인 영향력을 기존의 다른 영향력 지표는 직접적인 영향력을 반영한다.

Keywords: prestige indices, impact measure, citation network
명망성 지표, 영향력 측정, 인용 네트워크

---

* 본 연구는 한국연구재단 BK21 Plus 지원을 받았음(2015-11-0025).
** 연세대학교 대학원 문헌정보학 박사과정, 한국연구재단 연구원(hrahn@nrf.re.kr) (제1저자)
*** 연세대학교 문헌정보학과 부교수, Univ. of Pittsburgh Visiting Scholar(jihongpark@yonsei.ac.kr) (교신저자)
■ 논문접수일자 : 2016년 5월 18일 ■ 최초심사일자 : 2016년 5월 25일 ■ 게재확정일자 : 2016년 6월 6일
■ 정보관리학회지, 33(2), 61-76, 2016. 〔http://dx.doi.org/10.3743/KOSIM.2016.33.2.061〕

1. Introduction

Citation has been a widely used tool for measuring the performance and impact of researchers. As it functions to transfer knowledge and form the intellectual network of the relevant areas (Leydesdorff, 1998), accumulation of citations reflects the impact of a specific article, author, or journal (Yan, Ding & Sugimoto, 2011). In particular, simple citation counting has arguably been used as the public tool of quantitative scientific evaluation for several decades because of its straightforwardness with regards to understanding and calculation. However, some limitations have been reported with regard to the simple citation counting. Specifically, they cannot reflect the linking structure of citing journals, citing authors, or citing articles (Yan & Ding, 2012). To overcome this problem, starting with Pinski and Narin (1976), several researchers have been making an effort to develop citation network approach to measure the impact other than the simple citation count (Bollen, Rodriguez, & Van de Sompel, 2006; Ding, 2011; González-Pereira, Guerrero-Bote, & Moya-Anegón, 2010; Yan, Ding, & Sugimoto, 2011).

Such structural characteristics of citation networks provide an useful alternative like the PageRank, however it is still unknown what dimensions of impact these alternatives actually measure and how they are different. We have found no study that compares citation-network-based impact indicators to identify the dimensional differences of impact.

There are also potential benefits of the citation network approach. Unlike interpersonal relationship networks and co-authorship networks being reciprocal, citation networks are directed networks whose relationships are one-way. In citation networks, the more citations a node gets, the more the node is regarded as prestigious since citations are positive choices (De Nooy, Mrvar, & Batagelj, 2011). The concepts of centrality and prestige can be applied in undirected networks and directed networks to find out the prominent nodes in a network (Musial, Kazienko, & Bródka, 2009; Wasserman & Faust, 1994). The prestige indices reflect the scope on which one node can have influence directly or indirectly. In citation networks, a kind of information network (Yan & Ding, 2012), the prestige indices can be used to prescribe the scope on which one node can impart knowledge and exert its impact.

Considering the gap and potential contribution to impact study, this study proposes the three prestige indices－closeness prestige, input domain, and proximity prestige- as useful measures to evaluate the impact of a particular node in citation networks. In addition, we attempt to identify the implications of differentiating the three prestige indices from the existing impact indices through principal component analysis. The results of this study show the interesting possibility of utilizing the prestige indices for estimating the novel aspects of impact.

In this study, we focus on author citation networks and apply the prestige indices to evaluate the impact of a particular author. Simple citation counting, h-index, PageRank, and the centrality indices (which assume undirected networks) are chosen from the existing impact measures for comparison with the three prestige indices.

2. Literature Review

2.1 Evaluating the impact in citation networks

There are three types of networks in written scholarly communication. Yan and Ding (2012) classified scholarly networks as follows: (1) co-authorship networks are social networks that take note of contacts or interactions between actors; (2) co-citation networks, bibliographic coupling networks, and co-word networks are similarity-based networks that put emphasis upon investigation of research topics or academic fields; and (3) citation networks are information networks where nodes and links describe sources of knowledge and flows of knowledge, respectively.

Between the information networks and the social networks, there are differences in terms of directionality and heterogeneity (Romero & Kleinberg, 2010). Information networks are commonly represented in directed networks as the direction of flow becomes often meaningful interpretation clue. Whereas, social networks generally assume that relationships between people are reciprocal and are depicted as undirected networks even though some social networks such as friendship network can be directional. Information networks tend to have a few important nodes with an extremely large numbers of links, while social networks have relatively small disparities in the numbers of links each node has. In the same vein of information networks, citation is directional and citation networks are scale-free networks where distribution of links are consistent with the power law (Redner, 1998).

In evaluating impacts based on citation networks, the concept of prestige has been introduced to reflect the quality of citations. Prestige is often measured by the number of citations that an author gets from other authors who are highly cited while popularity depends on the mean number of citations that an author gets. Ding and Cronin (2011) argue that popularity and prestige may well have strong positive correlation, but are not necessarily equal. Starting with the journal-weighted PageRank developed by Bollen, Rodriguez and Van de Sompel (2006), plenty of studies that consider a combination of the prestige of citations have been conducted, including the SJR indicator (González-Pereira, Guerrero-Bote, & Moya- Anegón, 2010), author-weighted PageRank (Ding, 2011), and the P-Rank (Yan, Ding, & Sugimoto, 2011).

Similarly, centrality indices based on the structural characteristic of undirected networks are widely used to evaluate the impact in co-authorship networks (Yan & Ding, 2009) and co-citation networks (Lee, 2006). However, as far as we find out, it is rare to evaluate the impact based on the structural characteristics of directed networks in citation networks.

2.2 Prominence in Network: Centrality and Prestige

One of the basic aims of network analysis is to find out the most important actors in the network (Wasserman & Faust, 1994). If a particular node is more visible to other nodes in a network due to its links, the node is regarded as a prominent node (Knoke & Burt, 1983). When determining the prominence of a node, the nodes indirectly connected through intermediaries should also be considered as well as the directly neighbouring nodes (Wasserman & Faust, 1994).

Knoke and Burt (1983) classified prominence in a network by centrality and prestige. With regards to undirected links, it only matters how many links a node is involved in. In such a system, prominence is based on centrality. Whereas, with directed links, where to begin and where to head are important. If links are formed through positive choices, a node which gets more links than others becomes a prominent node. In this case, prominence is prestige. The difference between centrality and prestige also can be understood as distinguishing between in-degree and out-degree (Wasserman & Faust, 1994).

Centrality analysis is the most widely applied method for bibliometric data among social network analysis techniques (Lee, 2006). Centrality is measured in several ways depending on the interpretation on the concept of centrality (Sohn, 2002). Centrality is measured (1) by degree that a node connects with other nodes, defined as degree centrality, (2) by how close a node is to other nodes, defined as closeness centrality, and (3) by the extent to which a node acts as an intermediary or a bridge among other nodes in a network, defined as betweenness centrality.

Prestige analysis assumes directed networks. As aforementioned, among scholarly networks, citation networks are directed networks. To the best of our knowledge, Erman (2009) is alone in applied prestige analysis to citation networks.

Erman (2009) conducts citation analysis of the e-Government studies using 399 articles published in the proceedings of International Conference on e-Government (EGOV) from 2002 to 2008. He also employs prestige analysis to the data of 2,850 authors who published articles in the proceedings between 2005 and 2008. Input degree, input domain, and proximity prestige are chosen as prestige indices. The results are shown as top 10 authors on the three indices. He selects the most prestigious researchers (1. Grönlund, A., 2. Heeks, R.) on the e-Government field and points out that the state-of-the-art e-government research, the common research area of the top two researchers, is the most influential area on the field. Though he calculates the values of the prestige indices in a citation network and uses them for determining of the rankings of the authors, he does not identify the differentiated implication of the prestige indices applied on citation networks from the existing impact indices.

3. Prestige Indices for Impact Measure

We propose closeness prestige, input domain, and proximity prestige as measures of the impact in citation networks. The latter two indices were used by Erman (2009). All the three indices can be calculated by Pajek, a network analysis tool (Mrvar & Batagelj, 2014).

3.1 Closeness prestige

Closeness prestige estimates prestige based on the distances between particular nodes and all nodes linked with it, both directly and indirectly, and considering directions of the links. It is similar in concept with closeness centrality due to reflecting directionality and can be understood as in-degree closeness centrality.

The closeness prestige of node i, CPi is defined by the following equation (1), where dij is the minimum distance between node i and node j, where node j connected with node i directly or indirectly.

$$CP_i = \left[\Sigma_j d_{ij}\right]^{-1} \quad (1)$$

When we calculate closeness prestige, the difference in total distance occur depending on network size, like closeness centrality. Relative closeness prestige CP_{Ri} is obtained by multiplying n-1 to the value of CP_i.

$$CP_{Ri} = (n-1)\left[\Sigma_j d_{ij}\right]^{-1} \quad (2)$$

We use the relative closeness prestige in this study. Relative closeness prestige has a value between 0 and 1. As the value of relative closeness prestige in citation networks is nearer to 1, a node is more frequently and closely chosen by other nodes in such networks.

3.2 Input domain

The input domain of a node is the ratio of the directly or indirectly reachable nodes to all the nodes in a network (De Nooy, Mrvar, & Batagelj, 2011). If there are m reachable nodes to node i in a network in which the total number of nodes is n, input domain of node i, IDi can be calculated as equation (3).

$$ID_i = \frac{m}{n} \quad (3)$$

Input domain shows the relative size of the influential domain, which is defined as a set of the nodes affected by a particular node (Wasserman & Faust, 1994). The bigger the value of input domain, the higher a node's structural prestige. It can be understood that an author whose input domain value in citation networks is big influences a wide range of the relevant discipline.

The weakness of input domain as an index is that the values of input domain vary little in well-connected networks (Erman, 2009). Hence, restricted input domain, which reflects the distance-limited reachable nodes to a particular node, was proposed. We also use the three restricted input domains whose distance limits are 2, 3 and 4 in this study.

3.3 Proximity prestige

Proximity prestige displays how close a specific node is with all the other nodes in a network. Where Ii is the number of the nodes that are directly or indirectly connected to node i, the average distance between node i and all the nodes in the influential domain of node i is computed as $\Sigma_j d_{ij}/I_i$. Multiplying the ratio of Ii to the network size n-1 to the value of $\Sigma_j d_{ij}/I_i$ in order to remove the influence of network size, the following equation can be drawn (Wasserman & Faust, 1994).

$$PP_i = \frac{I_i/(n-1)}{\sum_j d_{ij}/I_i} \quad (4)$$

When we apply proximity prestige as an impact measure in citation networks, if the distance through indirect citations is longer, the impact decreases. Whereas, as the influence domain is larger by getting more direct and indirect citation, the impact increases.

4. Materials and Methods: Comparison of Impact Measures on Citation Network

4.1 Data Collection

This study applies the prestige indices to author citation networks for the evaluation of the impact of a specific researcher and compares the result with the existing impact measurements. For the analysis, we collected data on the field of library and information science (LIS) and established an author citation network.

Data was gathered in the top 5 LIS journals in the 'Information Science & Library Information Science' (IS & LIS) category as stated by the 2013 edition of the journal citation reports on the basis of impact factor. We classified the journals by relevant information on Ulrichsweb, and then based on impact factor (IF) and 5 year impact factor (5-Y IF), the following 5 LIS journals were selected: Journal of Informetrics (IF: 3.580, 5-Y IF: 3.609), Scientometrics (IF: 2.274, 5-Y IF: 2.294), Journal of the American Society for Information Science and Technology (Journal of the Association for Information and Technology) (IF: 2.230, 5-Y IF: 2.381), Annual Review of Information Science and Technology (IF: 1.727, 5-Y IF: 3.022), and Library & Information Science Research (IF: 1.384, 5-Y IF: 1.516).

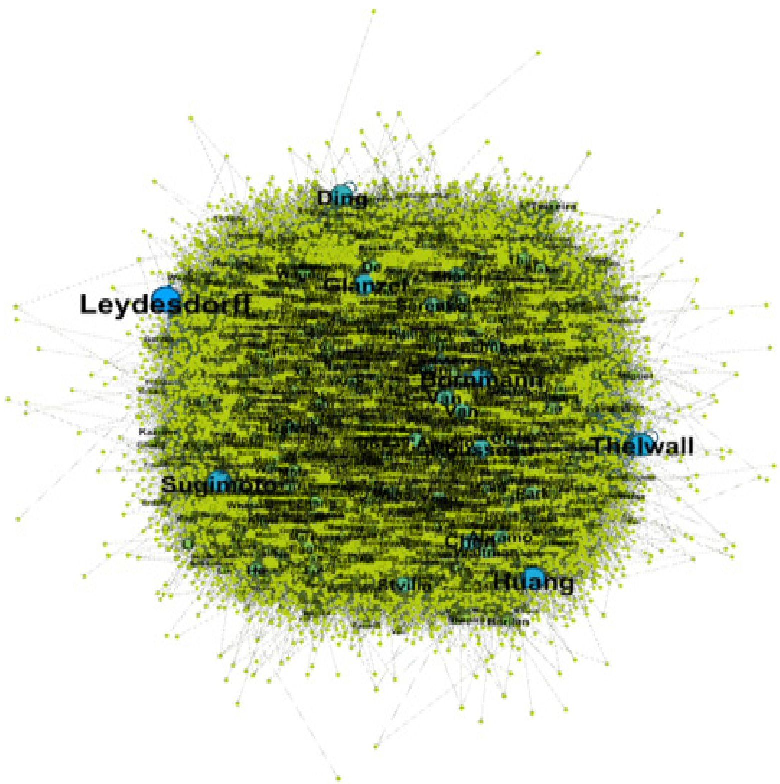
We searched all the articles published in the 5 LIS journals in 2011 and 2012 and extracted all the citation relationships from the References sections. Article search was executed using journal title on Scopus and document type was restricted to article. The details of the 981 collected articles are shown in Table 1. Because Annual Review of Information Science and Technology was only issued until 2011, there is no data on the journal in 2012.

〈Table 1〉 Number of articles collected from each journal

Journal Title	Number of Articles	Proportion (%)
Scientometrics	445	45.4
Journal of the American Society for Information Science and Technology	321	32.7
Journal of Informetrics	130	13.3
Library & Information Science Research	74	7.5
Annual Review of Information Science and Technology	11	1.1
Total	981	100

With the exclusion of 8 articles whose references were not saved properly (Annual Review of Information Science and Technology: 5 articles, Library & Information Science Research: 2, Scientometrics: 1), all the citation relations were extracted from 973 articles. We prevented potential errors in which an author was recognized as two or more different authors depending on the representation of name by removing punctuation marks and converting all letters to capitals. As a result of the pre-processing, the number of the authors decreased to 23,523 from 29,026. In author citation networks, author and citation relation are represented as node and arc respectively. The authors of the collected articles and all the cited authors in the Reference sections were included as nodes. The final author citation network has 23,523 nodes and 124,120 arcs. Though each node has more than 10 arcs on average, average degree (10.55) is much smaller than the total number of nodes. Hence, density of the network is a mere 0.00022.

Figure 1 is the author citation network visualized in Gephi. The sizes of node and name tag are configured in proportion to their degree. The names of the high-ranked researchers by degree are listed in Table 2.



〈Figure 1〉 Author citation network

〈Table 2〉 Top 10 researchers by degree

Name	Degree	Number of Articles	Number of Citations	h-index in the network
LEYDESDORFF	1030	223	424	19
HUANG	877	13	75	6
THELWALL	876	41	140	10
SUGIMOTO	822	10	55	7
BORNMANN	790	128	255	15
GLANZEL	753	167	402	18
DING	736	44	124	9
ROUSSEAU	662	29	296	13
CHEN	624	13	41	5
VAN RAAN	565	29	321	13

We assumed the author citation network as a binary network in order to calculate the impact measures. Three kinds of prestige indices were computed in Pajek. We used three restricted input domains with distance limits between nodes set at 2, 3, and 4 as input domain indices. As centrality indices suppose undirected networks, their values were calculated in Pajek after transformation of the original directed author citation network into an undirected author citation network. BibExcel and Gephi were used for the calculation of h-index and PageRank, respectively. In Appendix A, the 11 impact measures of the top 30 researchers by the mean number of citations are exemplified.

4.2 Correlation Analysis and Principal Component Analysis

Spearman rank-order correlation coefficients were calculated for each pair of impact measures for correlation analysis. For the 11 impact measures, this resulted in an 11x11 correlation matrix, shown in Appendix B. All pair-wise correlations are statistically significant ($N=11$, $p<0.01$). Also, the value of Kaiser-Meyer-Olkin Measure is larger than 0.5, and the general criterion and the results of Bartlett's Test of Sphericity are significant, as shown in Table 3. The two indices show adequacy of data for principal component analysis (PCA).

To identify the dimensions of each impact measure, we conducted PCA to deduce common factors

〈Table 3〉 KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.768
Bartlett's Test of Sphericity	Approximate Chi-Square	595758.378
	Degree of Freedom	55
	Significance Probability	.000

because it was verified that there are correlations between the impact measures used in this study. PCA is an analysis method for formulating new variables from collected multivariate data. It enables the discovery of new principal components expressed by linear combination of the exiting variables and makes summarization and interpretation of the data easy. In particular, we can find minimum principal components which do not correlate with each other through PCA when the original variables have correlations. If the prestige indices proposed in this study can depict the impact of an author on a different facet from the existing measures, it could be extracted as an individual principal component in PCA analysis. We used SPSS for PCA. The two principal components whose initial eigenvalues are bigger than 1 were selected based on total variance and initial eigenvalue (Table 4). The two components, component1 and component2, explain about 77.5% of total variance.

Communality, the ratio of variation of each variable explained by component1 and component2, is shown in Table 5. We continued PCA without the exclusion of any variables because the communalities of all the variables after extracting the two components are larger than 0.5 and relatively high. The abbreviations for the impact measures are as follows - DC: degree centrality, CC: closeness centrality, BC: betweenness centrality, CP: closeness proximity, ID2, ID3, ID4: input domains whose distance limits between nodes are 2, 3 and 4, respectively, PP: proximity prestige, and PR: PageRank.

The results of PCA are shown in Table 6. Varimax rotation and Kaiser normalization method were applied in the PCA. It shows that CP, PP, ID3, and ID4 are closely related with component 1, while Citation, DC, PR, BC, and h-index are with component 2. Whereas, CC and ID2 have loadings with both components in the sense that the 0.4 is an appropriate cut-off threshold in social science. Figure 2 shows this result visually that all the impact measures are classified into three major groups.

〈Table 4〉 Total variance explained

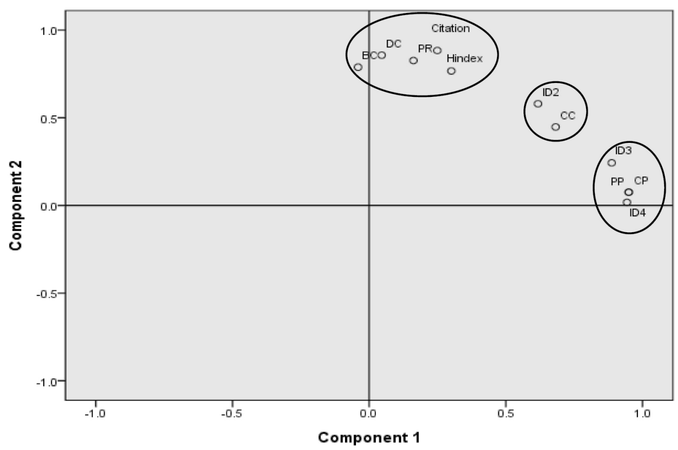
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.912	53.743	53.743	5.912	53.743	53.743
2	2.613	23.752	77.495	2.613	23.752	77.495
3	.875	7.952	85.447			
4	.610	5.548	90.995			
5	.342	3.110	94.106			
6	.258	2.348	96.454			
7	.162	1.469	97.923			
8	.102	.923	98.846			
9	.078	.708	99.554			
10	.049	.446	100.000			
11	2.180E-007	1.982E-006	100.000			

〈Table 5〉 Communalities

Impact Measure	Initial Value	Extracted Value
h-index	1.000	.678
Citation	1.000	.843
DC	1.000	.737
CC	1.000	.666
BC	1.000	.624
CP	1.000	.907
ID2	1.000	.718
ID3	1.000	.846
ID4	1.000	.890
PP	1.000	.907
PR	1.000	.709

〈Table 6〉 Rotated component matrix

Indicator	Component 1	Component 2
CP	**.950**	.076
PP	**.949**	.075
ID4	**.943**	.018
ID3	**.887**	.243
CC	**.682**	**.448**
ID2	**.618**	**.580**
Citation	.249	**.884**
DC	.046	**.857**
PR	.162	**.826**
BC	-.041	**.789**
h-index	.300	**.767**



〈Figure 2〉 Component Plot

5. Discussions

As a result of correlation analysis on the three kinds of five prestige indices, the three kinds of centrality indices, h-index, citation, and PageRank a statistically significant correlation exists between all the impact measures. The two principal components are extracted through PCA and the 11 impact measures are separated into three groups depending on correlations with the two components. The impact measures are degree centrality (DC), closeness centrality (CC), betweenness centrality (BC), closeness proximity (CP), input domains whose distance limits between nodes are 2, 3 and 4 (ID2, ID3, and ID4 respectively), proximity prestige (PP), and PageRank (PR).

Among these, CP, ID3, ID4, and PP are closely related with component1. They are based on directional citations, and are all the prestige indices proposed to evaluate the impact of a particular node in citation networks. Whereas, the existing impact measures based on citations (h-index, citation, PR, DC, and BC) are closely associated with component2. ID2 and CC have a meaningful associations with both component1 and component2 and form a separate group.

Spearman rank-order correlation coefficients between degree, in-degree, and out-degree on the author citation network were calculated to figure out the reason why some directional indices (h-index, citation, and PR) and some undirectional indices (DC and BC) occur in the same group. The h-index and PageRank are computed based on the number of citations that is same as in-degree of a directional network. Therefore, it can be considered that h-index and PageRank reflect directionality. By contrast, for DC and BC, the degree which does not take directionality into account is used.

As shown in Table 7, correlation between degree and in-degree (.897) is higher than that between degree and out-degree (.439). We can assume that in-degree becomes bigger than out-degree as degree increases. Therefore, including directionality in the calculation does not make a large difference among the impact measures.

BC (betweenness centrality), one of the centrality indices, considers both directed links and indirected links like the prestige indices. Nevertheless, BC is not classified in the same group with the prestige indices in PCA. We computed Spearman rank-order correlation coefficients between BC, degree, and input domain without distant limits to comprehend correlations between BC and directed/indirected links. Degree and input domain represent direct links and indirected links respectively.

<Table 7> Spearman Rank-order Correlation Coefficient of Degree, In-degree, and Out-degree

	Degree	In-degree	Out-degree
Degree	1.000	.857**	.439**
In-degree		1.000	.046**
Out-degree			1.000

** $p<0.01$ (2-tailed)

<Table 8> Spearman Rank-order Correlation Coefficient of BC, Degree, and Input Domain

	BC	Degree	Input Domain
BC	1.000	.811**	.010
Degree		1.000	.015*
Input Domain			1.000

* $p<0.05$ ** $p<0.01$ (2-tailed)

BC and degree show high positive correlation at the .01 significance level while BC and input domain have no statistically significant correlation. In other words, it can be assumed that BC becomes larger as degree representing direct links increases. However, the trends of BC cannot be inferred based on input domain representing indirect links.

The correlations can be explained on data grounds. Because 23,518 nodes among the total 23,523 nodes in the author citation networks have the same input domain value (0.99996), the values of input domain do not have the practical ability to discriminate between nodes. This weakness in the input domain is mentioned in the previous section 3.2. It can be understood that degree has a higher correlation with BC than input domain since the author citation network is a well -connected network.

PageRank, by definition, also considers indirect links in its calculation. However, it has been demonstrated that the effect of direct links on PageRank is very small. Instead, PageRank has high correlation with degree (Fortunato, Boguna, Flammini, & Menczer, 2008; Litvak, Scheinhardt, & Volkovich, 2006).

Taken together, component1 and component2 can be designated as the indirect impact through intermediaries and the direct impact on direct citation relationship respectively. As explained in the previous section, the prestige indices reflect both indirect links and direct links, while taking into account directionality. Hence, component1 is closely related with only the prestige indices and can be interpreted as indirect impact spreading with directionality through both direct and indirect links. While BC and PageRank are among the indices grouped in component2 that reflect indirect links, the effect of indirect links is very small and consideration of directionality does not make large difference due to the high correlation between degree and in-degree. Thus, component2 is unaffected by the existing impact measures which, based on citation counting, can be understood as the impact of direct links disregarding both indirect links and directionality.

The two indices that cannot be classified into either component1 or component2, ID2 and BC, also meet this interpretation. While ID2 reflects indirect links, it is limited to indirect links whose distance limit is only 1 or 2. Therefore, the effect of indirect links on ID2 is smaller than ID4 or PP and the effect of direct links is larger. Because CC is calculated based on distance, unlike other centrality indices, the effect of indirect links and direct links becomes larger and smaller, respectively. As a result, it can be assumed that the two indices are influenced by indirect impact component1 and direct impact component2 to a relatively similar extent.

6. Conclusions

In this study, we propose the three prestige indices －closeness prestige, input domain, and proximity prestige- as an useful alternative measures for the impact of a particular node in citation networks. This study is based on the observation that it is still unknown what dimensions impact indices actually measure. It compares the three prestige indices with other impact indices to identify a new measurement dimension of impact. The prestige indices enable us to distinguish the most prominent actors in a directed network, similar to the centrality indices in undirected networks. Correlation analysis and principal component analysis were conducted to identify the differentiated implications of the three prestige indices from the existing impact indices. We selected simple citation counting, h-index, PageRank, and the three kinds of centrality indices which assume undirected networks as the existing impact measures for comparison with the three prestige indices.

The author citation network in the field of library and information science was established for analysis of the 11 impact measures including the three restricted input domains whose distance limits are 2, 3 and 4. In correlation analysis using the Spearman rank-order correlation coefficient, all pair-wise correlations were statistically significant. We extracted two principal components through principal component analysis and the 11 impact measures are separated into three groups depending on their correlations with the two components.

As a result of integrating principal component analysis and supplementary correlation analysis, the first component and the second component can be designated as the indirect impact through intermediaries, and the direct impact on direct citation relationship, respectively. The first component is closely related to closeness prestige, input domains whose distance limits are 3 and 4, and proximity prestige. This first component can also be interpreted as indirect impact spreading with directionality through both direct and indirect links. The second component is applicable to the existing impact factors like citation, h-index, PageRank, degree centrality, and betweenness centrality. It can be understood as direct impact of direct links disregarding both indirect links and directionality. Closeness centrality and input domain whose distance limit is 2 are influenced by indirect impact of the first component and direct impact of the second component to a relatively similar extent.

Our results indicate that the three kinds of prestige indices proposed in this study can demonstrate indirect impact distinct from the existing impact measures which reflect direct impact based on simple citation counting. Citation networks are information networks where nodes and links describe sources of knowledge and the flow of knowledge, respectively (Yan & Ding, 2012). Information delivered through indirect links also must be considered for impact measures as well as information delivered through direct links. Like other network-based measures, this study also has some potential limitation in the sense that it may show the portion of a picture rather than the whole one because of the limited data.

In this study, we applied the prestige indices and the existing impact measures on the author citation networks to evaluate the impact of a particular author. Importantly, the prestige indices can be useful tools to evaluate the prominence of a particular paper on citation networks whose unit is a paper focusing on the diffusion of knowledge and information.

References

Bollen, J., Rodriguez, M. A., & Van de Sompel, H. (2006). Journal status. Scientometrics, 69(3), 669-687.

De Nooy, W., Mrvar, A., & Batagelj, V. (2011). Exploratory social network analysis with Pajek (27 vols.). New York: Cambridge University Press. http://dx.doi.org/10.1017/cbo9780511996368

Ding, Y. (2011). Applying Weighted PageRank to Author Citation Networks. Journal of the American Society for Information Science and Technology, 62(2), 236-245. http://dx.doi.org/10.1002/asi.21452

Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. Information Processing & Management, 47(1), 80-96. http://dx.doi.org/10.1016/j.ipm.2010.01.002

Erman, N. (2009). Citation analysis for e-government research. Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government, 244-253.

Fortunato, S., Boguna, M., Flammini, A., & Menczer, F. (2008). Approximating PageRank from in-degree. Lecture Notes in Computer Science, 4936, 59-71. http://dx.doi.org/10.1007/978-3-540-78808-9_6

González-Pereira, B., Guerrero-Bote, V.P., & Moya-Anegón, F. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. Journal of Informetrics, 4(3), 379-391. http://dx.doi.org/10.1016/j.joi.2010.03.002

Knoke, D., & Burt, R. S. (1983). Prominence. In R. S. Burt & M. J. Minor (Eds.), Applied network analysis: A methodological introduction (pp. 195-222). Newbury Park, CA: Sage.

Lee, Jae Yun (2006). Centrality measures for bibliometric network analysis. Journal of the Korean Society for Library and Information Science, 40(3), 191-214. http://dx.doi.org/10.4275/kslis.2006.40.3.191

Leydesdorff, L. (1998). Theories of citation? Scientometrics, 43(1), 5-25. http://dx.doi.org/10.2139/ssrn.2279062

Litvak, N., Scheinhardt, W. R. W., & Volkovich, Y. V. (2006). Probabilistic relation between In-Degree and PageRank. Proceedings of 4th International Workshop, Banff, Canada.

Mrvar, A., & Batagelj, V. (2014). Pajek and Pajek-XXL: Programs for analysis and visualization of very large networks reference manual list of commands with short explanation version 4.0. Retrieved from http://pajek-and-pajekxxl.software.informer.com

Musial, K., Kazienko, P., & Bródka, P. (2009). User position measures in social networks. Proceedings of the 3rd Workshop on Social Network Mining and Analysis, 6. http://dx.doi.org/10.1145/1731011.1731017

Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. Information Processing and Management, 12(5), 297-312. http://dx.doi.org/10.1016/0306-4573(76)90048-0

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. The European Physical Journal B (EPJ B), 4(2), 131-134. http://dx.doi.org/10.1007/s100510050359

Romero, D. M., & Kleinberg, J. M. (2010). The directed closure process in hybrid social-information networks, with an analysis of link formation on Twitter. Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, 138-145.

Sohn, Dong-Won (2002). Social Network Analysis. Seoul: Kyoungmoon.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. New York: Cambridge University Press.

Yan, E. J., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. Journal of the American Society for Information Science and Technology, 60(10), 2107-2118. http://dx.doi.org/10.1002/asi.21128

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

Yan, E. J., Ding, Y., & Sugimoto, C. R. (2011). P-Rank: An indicator measuring prestige in heterogeneous scholarly networks. Journal of the American Society for Information Science and Technology, 62(3), 467-477. http://dx.doi.org/10.1002/asi.21461

[Appendices]

A. Data of the top 30 researchers by the mean number of citations

	name	h-index	Citation	DC	CC	BC	CP	ID2	ID3	ID4	PP	PR
1	Leydesdorff L	19	424	943	0.3938	0.0490	0.0356	0.0582	0.0657	0.0669	0.0356	3.74E-04
2	Glanzel W	18	402	699	0.3818	0.0313	0.0346	0.0556	0.0656	0.0667	0.0346	3.88E-04
3	Van Raan A F J	13	321	521	0.3667	0.0156	0.0334	0.0546	0.0655	0.0669	0.0334	3.10E-04
4	Schubert A	15	316	393	0.3659	0.0121	0.0331	0.0535	0.0654	0.0667	0.0330	2.83E-04
5	Rousseau R	13	296	613	0.3785	0.0195	0.0329	0.0541	0.0649	0.0667	0.0328	3.14E-04
6	Moed H F	14	291	325	0.3646	0.0080	0.0326	0.0524	0.0656	0.0667	0.0326	2.81E-04
7	Garfield E	11	290	302	0.3697	0.0187	0.0334	0.0550	0.0663	0.0669	0.0334	3.65E-04
8	Bornmann L	15	255	718	0.3738	0.0217	0.0322	0.0519	0.0657	0.0669	0.0322	2.42E-04
9	Hirsch J	4	243	243	0.3492	0.0047	0.0317	0.0506	0.0656	0.0671	0.0317	3.51E-04
10	Van Leeuwen T N	13	237	469	0.3642	0.0107	0.0315	0.0501	0.0654	0.0669	0.0315	2.55E-04
11	Egghe L	12	214	261	0.3499	0.0041	0.0308	0.0488	0.0647	0.0667	0.0308	2.66E-04
12	Braun T	9	210	210	0.3518	0.0029	0.0309	0.0494	0.0648	0.0667	0.0309	1.96E-04
13	Daniel H	11	203	203	0.3471	0.0023	0.0312	0.0497	0.0657	0.0670	0.0312	1.86E-04
14	Small H	10	194	219	0.3582	0.0052	0.0318	0.0530	0.0655	0.0669	0.0318	1.82E-04
15	Newman M	13	190	190	0.3506	0.0057	0.0307	0.0485	0.0652	0.0668	0.0307	1.63E-04
16	Narin F	9	178	182	0.3503	0.0042	0.0303	0.0479	0.0644	0.0667	0.0302	2.05E-04
17	Persson O	9	173	191	0.3570	0.0029	0.0310	0.0508	0.0651	0.0669	0.0310	1.50E-04
18	Lariviere V	10	167	403	0.3631	0.0103	0.0302	0.0481	0.0645	0.0667	0.0302	1.56E-04
19	De Solla Price D	7	164	164	0.3506	0.0047	0.0308	0.0493	0.0659	0.0670	0.0308	1.75E-04
20	Zitt M	8	153	271	0.3491	0.0045	0.0302	0.0489	0.0643	0.0667	0.0302	1.67E-04
21	Bordons M	10	151	311	0.3567	0.0112	0.0293	0.0441	0.0644	0.0667	0.0293	1.38E-04
22	Gingras Y	9	149	230	0.3484	0.0032	0.0297	0.0469	0.0640	0.0666	0.0297	1.31E-04
23	BarIlan J	8	148	262	0.3641	0.0139	0.0307	0.0502	0.0656	0.0667	0.0307	1.78E-04
24	Meyer M	9	148	148	0.3449	0.0022	0.0298	0.0471	0.0642	0.0667	0.0298	1.45E-04
25	Martin B R	5	146	146	0.3461	0.0027	0.0295	0.0452	0.0641	0.0670	0.0294	1.53E-04
26	Chen C	8	141	351	0.3590	0.0147	0.0299	0.0471	0.0651	0.0667	0.0299	1.23E-04
27	Katz J S	6	141	141	0.3415	0.0017	0.0291	0.0440	0.0635	0.0670	0.0290	1.50E-04
28	Thelwall M	10	140	848	0.3825	0.0529	0.0300	0.0483	0.0643	0.0668	0.0300	1.20E-04
29	Thijs B	8	138	304	0.3492	0.0043	0.0290	0.0450	0.0628	0.0666	0.0290	1.42E-04
30	Visser M	10	136	203	0.3448	0.0016	0.0300	0.0479	0.0650	0.0668	0.0300	1.49E-04

* DC: Degree Centrality, CC: Closeness Centrality, BC: Betweenness Centrality, CP: Closeness Prestige, ID: Input Domain, PP: Proximity Prestige, PR: PageRank

B. Spearman Rank-order Correlation Matrix

	h-index	Citation	DC	CC	BC	CP	IP2	IP3	IP4	PP	PR
h-index	1.000	.588**	.439**	.349**	.444**	.422**	.441**	.418**	.425**	.422**	.534**
Citation	.588**	1.000	.857**	.512**	.658**	.582**	.631**	.577**	.581**	.582**	.801**
DC	.439**	.857**	1.000	.635**	.811**	.448**	.495**	.442**	.446**	.447**	.676**
CC	.349**	.512**	.635**	1.000	.579**	.762**	.767**	.761**	.761**	.762**	.300**
BC	.444**	.658**	.811**	.579**	1.000	.357**	.394**	.351**	.359**	.356**	.574**
CP	.422**	.582**	.448**	.762**	.357**	1.000	.970**	.998**	.996**	1.000**	.475**
IP2	.441**	.631**	.495**	.767**	.394**	.970**	1.000	.970**	.965**	.970**	.506**
IP3	.418**	.577**	.442**	.761**	.351**	.998**	.970**	1.000	.994**	.998**	.467**
IP4	.425**	.581**	.446**	.761**	.359**	.996**	.965**	.994**	1.000	.997**	.471**
PP	.422**	.582**	.447**	.762**	.356**	1.000**	.970**	.998**	.997**	1.000	.473**
PR	.534**	.801**	.676**	.300**	.574**	.475**	.506**	.467**	.471**	.473**	1.000

* DC: Degree Centrality, CC: Closeness Centrality, BC: Betweenness Centrality, CP: Closeness Prestige, ID: Input Domain, PP: Proximity Prestige, PR: PageRank

** Correlation is significant at the 0.01 level (2-tailed)