

계량서지적 분석에서 지적구조 매핑을 위한 링크 삭감 알고리즘의 적합도 측정*

**Measuring the Goodness of Fit of Link Reduction Algorithms for
Mapping Intellectual Structures in Bibliometric Analysis**

이재윤 (Jae Yun Lee) **

초 록

지적구조 분석을 위해 각종 네트워크를 시각화해야 하는 경우에 페스파인더 네트워크와 같은 링크 삭감 알고리즘이 널리 사용되고 있다. 이 연구에서는 네트워크 시각화를 위한 링크 삭감 알고리즘의 적합도를 측정하기 위한 지표로 NetRSQ를 제안하였다. NetRSQ는 개체간 연관성 데이터와 생성된 네트워크에서의 경로 길이 사이의 순위 상관도에 기반하여 네트워크의 적합도를 측정한다. NetRSQ의 타당성을 확인하기 위해서 몇 가지 네트워크 생성 방식에 대해 정성적으로 평가를 했었던 선행 연구의 데이터를 대상으로 시험적으로 NetRSQ를 측정해보았다. 그 결과 품질이 좋게 평가된 네트워크일수록 NetRSQ가 높게 측정됨을 확인하였다. 40가지 계량서지적 데이터에 대해서 4가지 링크 삭감 알고리즘을 적용한 결과에 대해서 NetRSQ로 품질을 측정하는 실험을 수행한 결과, 특정 알고리즘의 네트워크 표현 결과가 항상 좋은 품질을 보이는 것은 아니며, 반대로 항상 나쁜 품질을 보이는 것도 아님을 알 수 있었다. 따라서 이 연구에서 제안한 NetRSQ는 생성된 계량서지적 네트워크의 품질을 측정하여 최적의 기법을 선택하는 근거로 활용될 수 있을 것이다.

ABSTRACT

Link reduction algorithms such as pathfinder network are the widely used methods to overcome problems with the visualization of weighted networks for knowledge domain analysis. This study proposed NetRSQ, an indicator to measure the goodness of fit of a link reduction algorithm for the network visualization. NetRSQ is developed to calculate the fitness of a network based on the rank correlation between the path length and the degree of association between entities. The validity of NetRSQ was investigated with data from previous research which qualitatively evaluated several network generation algorithms. As the primary test result, the higher degree of NetRSQ appeared in the network with better intellectual structures in the quality evaluation of networks built by various methods. The performance of 4 link reduction algorithms was tested in 40 datasets from various domains and compared with NetRSQ. The test shows that there is no specific link reduction algorithm that performs better over others in all cases. Therefore, the NetRSQ can be a useful tool as a basis of reliability to select the most fitting algorithm for the network visualization of intellectual structures.

키워드: 링크 삭감 알고리즘, 네트워크 시각화, 지적구조 매핑, 페스파인더 네트워크, 클러스터링 기반

네트워크, 적합도

link reduction algorithm, network visualization, mapping intellectual structure,

pathfinder network, clustering-based network, goodness of fit

* 이 논문은 2021년도 명지대학교 교내연구비를 지원받아 수행되었음.

** 명지대학교 인문대학 문헌정보학과 교수(memexlee@mju.ac.kr)

■ 논문접수일자 : 2022년 5월 15일 ■ 최초심사일자 : 2022년 6월 7일 ■ 게재확정일자 : 2022년 6월 7일

■ 정보관리학회지, 39(2), 233-254, 2022. <http://dx.doi.org/10.3743/KOSIM.2022.39.2.233>

※ Copyright © 2022 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited. the use is non-commercial and no modifications or adaptations are made.

1. 서 론

다차원척도법(Multidimensional Scaling; 이하 MDS로도 표기)은 Torgerson(1952)이 심리학 분야에서 다차원의 유사성 또는 비유사성 자료를 저자원(주로 2차원)으로 시각화하기 위해서 처음 제안한 이후에 계속 발전해오면서 다양한 분야에서 광범위하게 활용되어왔다. 저자동 시인용분석(White & Griffith, 1981) 기법을 비롯해서 여러 지적구조 분석 연구에서도 다차원척도법이 한동안 널리 활용되었다. 다차원척도법은 대부분의 통계분석 패키지에서 기본 도구로 제공되고 있으며, 생성된 2차원 지도가 입력된 데이터를 얼마나 잘 반영하고 있는가를 스트레스와 같은 불일치도(badness of fit)나 RSQ와 같은 적합도(goodness of fit) 지표로 나타내고 있다. 통계검증 결과를 논문에 제시할 때 통계적 유의수준을 함께 보고하듯이 다차원척도법 적용 결과는 White와 Griffith(1981)의 시도 이후 대부분의 지적구조분석 연구에서 적합도와 함께 보고되어왔다. 그러나 다차원척도법은 세부 구조의 표현력이 떨어지며 대량의 개체를 표현하기 어렵다는 단점이 있다(이재윤, 2006). 이 때문에 흔히 사용되는 SPSS의 ALSCAL 프로시저에서는 100개 이하의 개체만 다룰 수 있도록 한계가 설정되어 있다.

최근에는 네트워크분석 기법이 발전하고 간편한 소프트웨어가 보급되면서 다차원척도법을 대신하여 네트워크 형태로 동시인용이나 키워드 동시출현 관계를 시각화하는 사례가 급증하였다. 그런데 이와 같은 계량서지적 분석 데이터인 동시인용 네트워크나 키워드 네트워크 등은 동시인용 빈도나 동시출현 빈도에 근거하

여 관계를 측정하므로 본질적으로 가중 네트워크이고 대부분 밀도가 높다. 밀도가 높은 네트워크의 연결구조 전체를 그대로 시각적으로 표현하면 텔몽치 그래프(hairball graph)라고 불릴 정도로 엉킨 형태로 나타나서(Zweig, 2016, 9) 네트워크에 내재된 구조 정보를 드러내기가 어렵게 된다.

복잡한 네트워크 구조를 단순화해서 표현하기 위해서는 중요한 링크와 그렇지 않은 링크를 선별하는 알고리즘이 필요하다. 1980년대 말에 Schvaneveldt(1990)가 개발한 패스파인더 네트워크(Pathfinder Network; 이하 PFNet으로 표기) 알고리즘이 대표적으로, White(2003)가 저자동시인용 분석 결과를 시각화하는 도구로 PFNet을 사용한 이후에 널리 확산되었다. 그런데 다차원척도법의 경우에는 생성된 2차원 지도의 품질을 평가하는 지표가 보편적으로 사용되고 있는 반면에, PFNet과 같은 링크 삭감 알고리즘으로 생성된 지적구조 네트워크의 품질을 확인하는 도구는 아직 개발되지 않았다. PFNet 알고리즘을 다른 연구들은 대부분 높은 계산복잡도를 낮춰서 수행속도를 향상시키는 것을 목적으로 수행되어왔을 뿐이다(Arleo, Kwon, & Ma, 2017; Hauguel, Zhai, & Han, 2009; Quirin et al., 2008).

한편 PFNet을 대체할 수 있는 알고리즘으로 클러스터링 기반 네트워크(Clustering-based Network; 이재윤, 2007: 2008; 이하 CBNet으로 표기)와 같은 경쟁적인 기법도 제안된 이상, 여러 방법으로 네트워크 지도를 생성해본 후에 가장 최적의 결과물을 선택하기 위한 수단이 필요하다.

이 연구에서는 연관성 행렬을 입력하여 도출

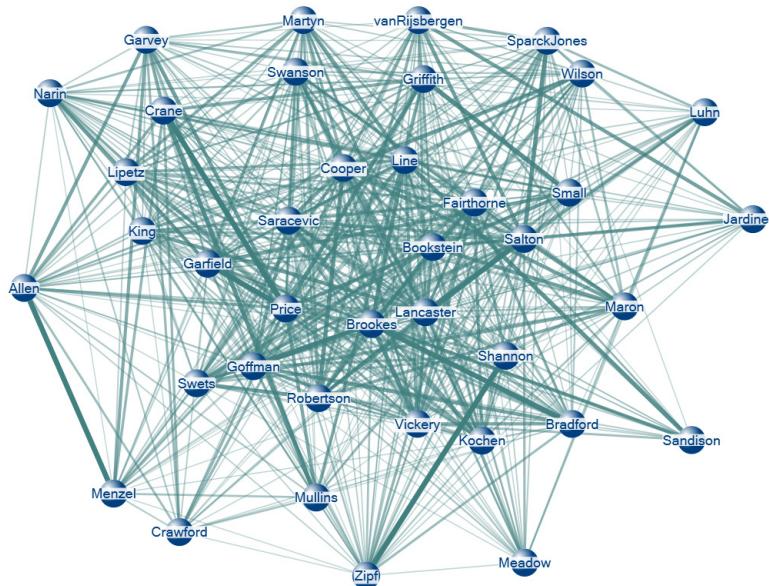
된 네트워크 표현에 대해서 입력 데이터와의 적합도(goodness of fit)를 측정하기 위한 방법을 개발하고자 한다. 이를 위해서 다양한 네트워크 표현 방식에 대해서 먼저 알아본 다음, 네트워크 표현의 적합도를 평가하는 새로운 지표를 제안하고 가상 데이터와 실제 데이터에 대해서 적용해보는 실험을 수행하였다.

2. 네트워크 표현 방식

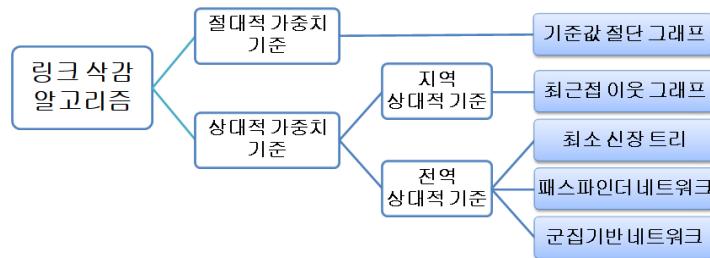
계량서지적 네트워크 데이터는 동시인용, 동시출현, 또는 상관관계 데이터를 기반으로 생성되는데, 이들은 대부분 밀도가 높은 네트워크이다. 예를 들면 White와 Griffith(1981)의 논문에 제시된 저자동시인용 데이터를 그대로 네트

워크로 표현하면 <그림 1>과 같이 저자들이 뒤엉켜서 구조를 알아보기 힘든 형태가 된다.

<그림 1>처럼 복잡한 각종 네트워크는 중요한 링크만 남기고 관계가 약한 링크는 삭제한다면 전체적인 구조가 잘 보이게 된다. 전체 링크 중에서 사소한 링크는 제외하고 중요한 링크만 남기는 방식을 링크 삭감 알고리즘(link reduction algorithm), 또는 네트워크 가지치기 알고리즘(network pruning algorithm)이라고 부른다(이재윤, 2006). 반대로 중요한 링크만 선택하여 표현하는 것을 네트워크 생성 알고리즘(network generation algorithm)이라고 부르는데 중요한 링크는 표현하고 사소한 링크는 배제한다는 점에서 둘은 같은 의미이다. 지적구조 분석에 활용되는 주요 링크 삭감 알고리즘은 <그림 2>와 같다.



<그림 1> White와 Griffith(1981)의 저자동시인용 데이터를 링크 삭감없이 네트워크로 표현한 결과



〈그림 2〉 지적구조 분석에 적용되는 주요 링크 삭감 알고리즘의 구분

기준값 절단 방식은 가장 간단한 링크삭감 알고리즘으로서 기준값을 어떻게 설정하는가에 따라서 복잡한 형태가 될 수도 있고 간단한 형태가 될 수도 있다. 코사인유사도 데이터라면 ‘0.1 이상’이나 ‘0.05 이상’과 같은 기준값을 경험적으로 설정할 수 있다. 데이터에 따라서 링크의 가중치 수준은 매우 다를 수 있기 때문에 ‘0.1 이상’이라는 기준이 매우 엄격한 기준일 수도 있고 느슨한 기준일 수도 있다. 링크의 가중치가 대체로 낮은 데이터에서는 ‘0.1 이상’이 엄격한 기준이 되어서 상당수의 링크가 삭제되 버릴 수도 있다. 지나치게 많은 링크를 삭제해 버리면 일부 노드는 링크를 하나도 가지지 못하는 고립 노드가 되버릴 수도 있으므로 주의해야 한다. 이 때문에 데이터마다 임의로 링크 삭감 기준을 설정하기보다는 전체 링크 중 가중치 상위 몇% 이상을 기준으로 하거나, 링크 가중치의 평균을 기준값으로 삼거나, 각 노드마다의 최대링크 가중치 중에서 최솟값을 기준값으로 삼는 것이 합리적이다.

기준값 절단 방식에서 전체 링크 중 가중치 상위 일정한 비율을 기준으로 할 때에는 10%나 20%의 링크만 남겨도 전체 네트워크의 구조가 잘 드러날 수 있다. 이 경우에도 몇 %를 기준으로 할 것인지는 경험적으로 결정해야 하

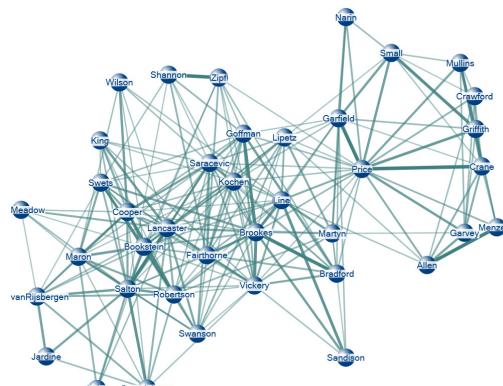
는 문제가 남는다. 이 때문에 아예 링크 가중치의 평균을 기준값으로 해서 평균 이상의 가중치를 가진 링크만 남기는 방식을 쓰는 것도 가능하다. 대부분의 경우 링크 가중치는 빈익빈부익부 분포를 보이므로 가중치가 평균 이상인 링크가 평균 미만인 링크보다 적다. 따라서 평균 이상의 링크만 남겨서 네트워크를 표현하면 보통은 상위 20~30% 정도의 링크만 남기는 효과를 얻게 된다.

전체의 20% 정도의 링크만 남기더라도 생성되는 네트워크는 꽤 복잡한 구조를 가지게 된다. 노드의 수가 많을수록 가능한 전체 링크의 수는 기하급수적으로 늘어나므로 노드의 수가 매우 많은 네트워크에서 링크 삭감 알고리즘을 적용하려면 남길 링크의 비율은 더욱 낮게 설정해야 한다. 따라서 고립 노드가 발생하지 않는 것을 보장하면서도 최소의 링크만 남길 수 있는 기준값을 결정하기 위해서는 각 노드마다의 최대링크 가중치 중에서 최솟값을 선택하는 것을 고려해볼 수 있다. 이때는 먼저 네트워크에 있는 각 노드마다 이웃과 가진 링크 가중치 중에서 최댓값을 파악한다. n개의 노드가 있는 네트워크라면 n개의 노드별 최댓값이 얻어진다. 그런 후 이를 n개의 최댓값 중에서 최솟값을 기준값으로 하여 이보다 작은 링크는 모두

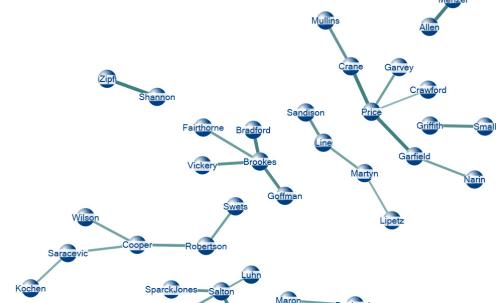
삭감하면 어떤 노드라도 최소한 1개의 링크는 가질 수 있도록 보장하게 된다. 이처럼 기준값 절단 방식에서는 기준값을 어떻게 설정하는가에 따라서 표현된 네트워크의 구조가 다양하게 나타날 수 있다.

〈그림 3〉은 〈그림 1〉에 제시했던 White와 Griffith(1981)의 저자동시인용 네트워크에 링크 삭감 알고리즘을 적용하여 기준값 절단, 최근접이웃 그래프, PFNet, CBNet(Ward 군집화 기반으로 WARD를 부기함)의 4가지 방

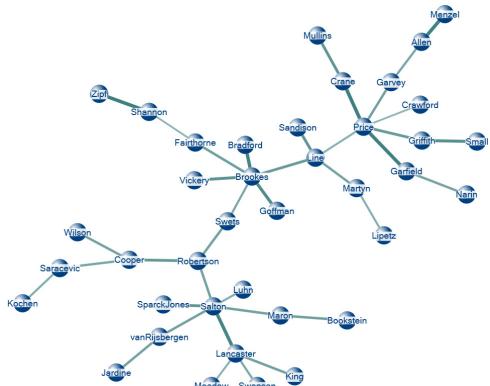
식으로 표현한 결과이다. 최소신장트리는 따로 표현하지 않았는데, PFNet의 가장 엄격한 옵션 설정($q=n-1, r=\infty$)에서 동률값인 링크가 존재하지 않으면 최소신장트리와 같은 모양이 되므로 〈그림 3〉의 (c)는 최소신장트리 알고리즘을 적용한 결과와 같다. 표현 결과에서 보듯이 (a)의 기준값 절단 네트워크는 오른쪽에 계량서지학 연구자 집단이 형성되는 등 집단 구조를 조금 더 드러내는 형태가 된다. 그러나 여전히 개별 연구자의 입지가 잘 드러날 정도로



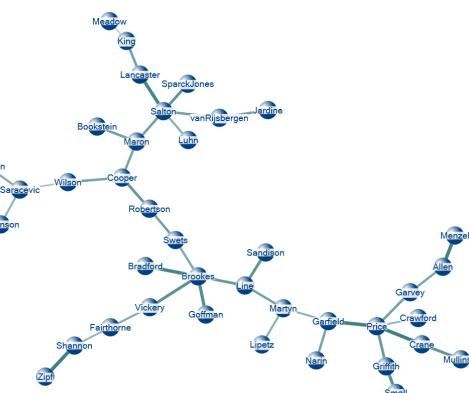
(a) 기준값(평균값) 절단 네트워크



(b) 최근접이웃 그래프



(c) PFNet



(d) CBNet-WARD

〈그림 3〉 White와 Griffith(1981)의 저자동시인용 데이터를 네트워크 형태로 표현한 결과

구조가 선명하게 보이지는 않는다. (b)의 최근 접 이웃 그래프는 연구자 집단이 소그룹으로 잘 구분되지만 전체가 하나로 연결되지 못한다는 문제가 있다. 전역 상대적 기준을 적용한 (c)와 (d)는 전체가 하나의 네트워크로 연결되면서도 구조가 선명하게 드러난다는 장점이 있다. 그러나 구조를 잘 보여주기 위해서 상당수의 링크를 제거하였으므로 노드 간의 가깝고 먼 관계가 일부 왜곡될 위험이 있다. 이는 다차원 척도법에서 고차원의 데이터를 2차원 평면에 표현하면서 개체 간 거리가 일부 왜곡되는 것을 피할 수 없는 것과 같다. 따라서 전역 상대적 연결 방식으로 표현된 네트워크에서 노드 간의 관계가 어느 정도로 왜곡되어 표현되었는지를 측정한 후, 관계 왜곡 정도가 상대적으로 덜한 네트워크 표현 결과를 선택하는 것이 바람직하다.

3. 네트워크 표현의 적합도 지표 제안

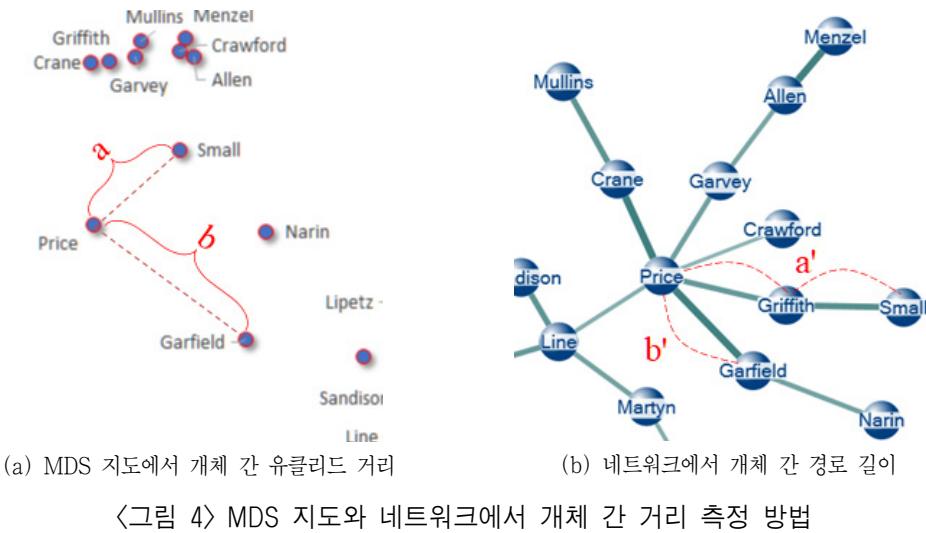
3.1 경로 길이에 기반한 네트워크 표현 적합도 지표

다차원척도법 결과의 적합도를 측정할 때에는 생성된 2차원 지도에서 개체 간 거리를 유클리드 거리 공식으로 산출한 다음 입력된 연관도와 비교한다. 링크 색감 알고리즘으로 도출된 네트워크의 적합도를 평가할 때에도 네트워크를 구성하는 개체 사이의 거리 측정 방법이 필요하다.

네트워크에서는 물리적 거리가 경로 길이(path length)로 대체된다(Barabási, 2016, 56). 경로

길이는 두 개체 사이가 몇 개의 링크로 이어져 있는가를 나타낸다. 둘 이상의 경로가 존재할 때에는 가장 짧은 경로 길이를 채택한다. 예를 들어 White와 Griffith(1981)의 저자동시인용 데이터에서 Price와 Small 사이의 코사인 계수 0.551 보다 Price와 Garfield 사이의 코사인 계수 0.865가 더 크므로 Price는 Small보다 Garfield 와 더 가까운 자리에 위치해야 한다. 그런데 이 저자동시인용 데이터로 고전적 다차원척도법을 수행하여 얻어진 MDS 지도에서는 <그림 4>의 (a)와 같이 Price와 Small 사이의 거리 0.153 보다 Price와 Garfield 사이의 거리 0.257이 더 멀어서 관계가 다르게 표현되어 있다. 이와 달리 PFNet에서는 <그림 4>의 (b)와 같이 Price 와 Small 사이의 경로 길이가 2로서, Price와 Garfield 사이의 경로 길이 1보다 더 길어서 동시인용관계가 올바르게 반영되어 있다.

입력된 연관성 행렬에서 두 개체 사이의 연관성 값이 클수록 네트워크에서는 개체 사이의 최단 경로길이가 짧게 표현되어야 한다. 따라서 입력 데이터인 연관도를 결과 네트워크에서의 개체 간 최단 경로길이와 비교하여 네트워크의 적합도를 산출하기로 한다. 이는 다차원 척도법 중 SPSS의 ALSCAL 프로시저에서 적합도 지수로 사용되는 RSQ(R SQuared)와 유사한 접근법이다. RSQ는 MDS 지도에서 산출한 개체 사이의 유클리드 거리와 입력 데이터에서 개체 사이의 상위(disparity: 비유사성)을 단조 변환한 것) 사이의 상관성을 제곱하여 구한다(Young, 1987). McCain(1986)을 비롯한 지적구조 분석 연구의 초기부터 여러 연구에서 다차원척도법의 결과에 대한 평가 수치로 RSQ가 사용되어왔다. 개체 사이의 연관성이나



〈표 1〉 Price와 Small, Price와 Garfield 사이의 유사도와 거리

개체 쌍	동시인용 cosine 계수	MDS 지도에서 유클리드 거리	PFNet에서 경로 길이
Price - Small	0.551	0.153	2
Price - Garfield	0.865	0.257	1

경로 길이는 정규분포가 아닌 경우가 흔하기 때문에 연관성 순위와 경로 길이 순위 사이의 상관계수를 제곱한 값으로 네트워크의 적합도를 평가하며 이를 NetRSQ(Network R SQuared)라고 부르기로 한다.

연관성 데이터 S 를 입력하여 도출한 네트워크에서 노드 간 경로 길이의 합을 D 라고 할 때, 이 네트워크를 평가하기 위한 NetRSQ(S, D)는 연관성 값의 순위 $R(S)$ 와 노드 간 경로길이 값의 순위 $R(D)$ 사이의 피어슨 상관계수를 구한 후 제곱하여 산출한다.

$$\text{NetRSQ}(S, D) = \text{correl}(R(S), R(D))^2$$

이때 연관성 값의 순위는 내림차순으로 산출하고 경로길이 값의 순위는 오름차순으로 산출한다. NetRSQ는 0에 가까울수록 생성된 네트워크의 적합도가 낮음을 의미하고, 반대로 1에 가까울수록 적합도가 높음을 의미한다.

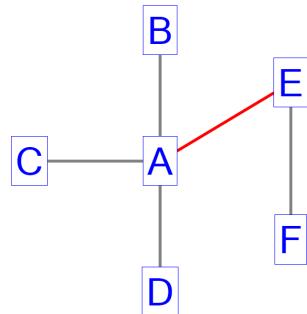
3.2 가상 네트워크에 대한 시험 측정¹⁾

〈표 2〉는 NetRSQ를 시험적으로 적용해보기 위한 가상 연관성 데이터이다. 이 개체들 사이의 관계를 네트워크로 표현한 두 가지 결과를 〈그림 5〉와 〈그림 6〉에 제시했다. 사례 2는 실제로 〈표 2〉의 연관성 데이터에 대해서 PFNet

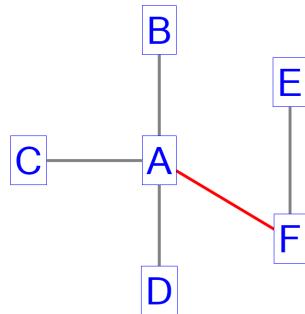
1) 이 절의 내용은 제19회 한국정보관리학회 학술대회(이재윤, 2012)에서 소개된 바 있음.

〈표 2〉 6개 개체의 가상 연관성 행렬

	[A]	[B]	[C]	[D]	[E]	[F]
[A]	1.0	0.9	0.8	0.6	0.2	0.3
[B]	0.9	1.0	0.6	0.4	0.1	0.0
[C]	0.8	0.6	1.0	0.5	0.1	0.0
[D]	0.6	0.4	0.5	1.0	0.1	0.0
[E]	0.2	0.1	0.1	0.1	1.0	0.7
[F]	0.3	0.0	0.0	0.0	0.7	1.0



〈그림 5〉 가상 네트워크 사례 1



〈그림 6〉 가상 네트워크 사례 2

알고리즘을 적용한 결과이기도 하다. 〈그림 5〉와 〈그림 6〉은 대부분의 링크가 동일하게 생성되어 있고 빨간색으로 표시된 링크 하나씩만 다르다.

개체 간 연관도와 두 사례 네트워크에서의 최단 경로길이를 비교한 〈표 3〉을 보면, 6개 개체 사이에 존재하는 15개의 개체 쌍 중에서 상위 7개까지는 두 네트워크가 같다. 그러나 8번 째와 9번째 개체 쌍인 [A]-[F]와 [A]-[E]의 최단 경로길이 순위가 두 네트워크에서 뒤바뀌었으며, 이후의 순위도 모두 다르게 되어 있다. 사례 1에서는 연관도가 더 높은 [B]-[E], [C]-[E], [D]-[E]의 경로길이가 더 짧지만, 사례 2에서는 연관도가 더 낮은 [C]-[F], [D]-[F], [B]-[F]의 경로길이

가 더 짧게 표현되어 있다. 입력된 연관도와 네트워크에서의 최단 경로길이 사이의 NetRSQ는 사례 1이 0.686, 사례 2가 0.458로 나타난다. 연관도와 최단 경로길이의 순위가 더 일치하는 사례 1의 NetRSQ가 사례 2의 NetRSQ보다 더 높으므로 NetRSQ가 네트워크의 적합도를 올바르게 측정한다는 것이 확인되었다.

〈그림 6〉의 네트워크 표현 사례 2는 가상 연관성 데이터에 PFNet 알고리즘을 적용한 결과인데, NetRSQ를 기준으로 할 때 이보다 더 나은 네트워크 표현(사례 1)이 존재한다는 것이 확인되었다. 따라서 NetRSQ로 네트워크 표현 결과를 평가한 후 여러 대안 중에서 최선의 결과를 선택하는 것이 중요함을 알 수 있다.

〈표 3〉 가상 연관성 데이터에서 개체 간 연관도와 네트워크 표현 결과에서의 최단 경로길이

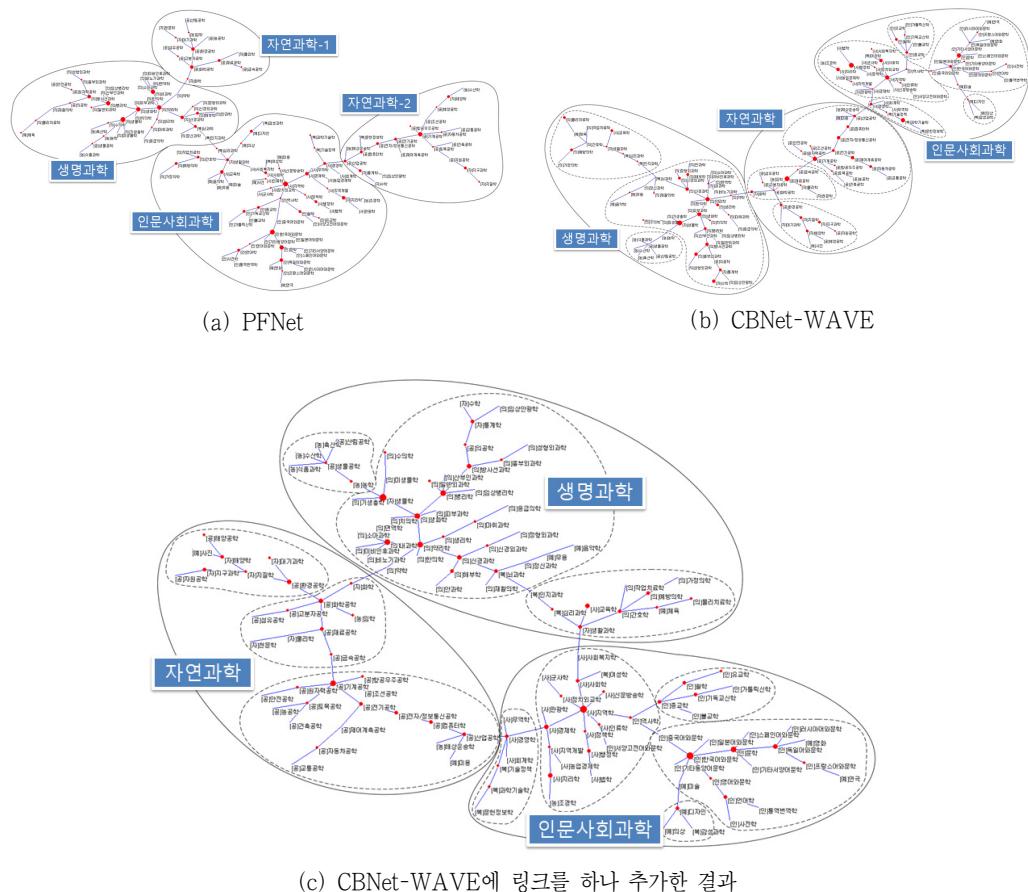
개체 쌍		연관성	최단 경로길이	
개체1	개체2		사례 1	사례 2
A	B	0.9	1	1
A	C	0.8	1	1
E	F	0.7	1	1
A	D	0.6	1	1
B	C	0.6	2	2
C	D	0.5	2	2
B	D	0.4	2	2
A	F	0.3	2	1
A	E	0.2	1	2
B	E	0.1	2	3
C	E	0.1	2	3
D	E	0.1	2	3
C	F	0	3	2
D	F	0	3	2
B	F	0	3	2

3.3 선행연구의 정성적 평가와 NetRSQ 측정 결과 비교

한국학술진흥재단(현재는 한국연구재단으로 통합)에 신청된 학문분야별 연구과제 신청서의 텍스트를 분석한 이재윤(2008)의 연구에서는, 국내 학문분야 140개 사이의 주제 연관성을 분석하였다. 이 연구에서는 140개 학문분야 사이의 연관성 데이터를 입력하여 〈그림 7〉의 (a) PFNet과 (b) CBNet-WAVE(평균연결군집화 기법 기반)를 도출하였다. 그런데 PFNet에서는 자연과학 분야가 화학공학 중심의 자연과학-1과 기계공학 중심의 자연과학-2로 분리되는 부적절한 결과가 얻어졌다. 이는 자연과학-1 영역에 속한 화학 분야가 생명과학 영역에 속한 약학 분야와 밀접하게 관련되어 있고, 자연과학-2

영역에 속한 산업공학 분야가 사회과학 영역에 속한 경영학 분야와 밀접하게 관련되어 있기 때문이다. 즉 자연과학-1 영역과 자연과학-2 영역에 속한 개별 학문분야들 사이의 관계보다 화학-약학 사이의 관계와 산업공학-경영학 사이의 관계가 더 밀접했기 때문에 자연과학-1 영역과 자연과학-2 영역이 PFNet에서 이웃으로 이어지지 못하고 둘로 분리되어 각각 생명과학 영역과 사회과학 영역으로 이어지는 결과를 낳은 것이다.

PFNet과는 달리 CBNet-WAVE에서는 인문사회과학 영역, 자연과학 영역, 생명과학 영역으로 삼분되는 자연스러운 연결 구조가 도출되었다. 또한 CBNet-WAVE에 링크 하나를 추가하였을 때, 〈그림 7〉의 (c)와 같이 인문사회과학-자연과학-생명과학-인문사회과학의 순서로 학문구조가



〈그림 7〉 학문분야 네트워크의 3가지 표현 결과
(출처: 이재윤, 2008)

순환구조로 이어져서 더욱 자연스러운 네트워크가 도출되었다. 이때 추가한 링크는 PFNet에서 표현된 링크 중에서 CBNet-WAVE에서는 가장 멀리 떨어져 있는 두 분야 사이이다. 이 세 가지 경우의 적합도를 측정해보면, PFNet는 NetRSQ가 0.2390, CBNet-WAVE는 NetRSQ가 0.3468, 링크 추가로 변형한 CBNet-WAVE는 NetRSQ가 0.4104로 나타났다. 즉, 네트워크 구조에 대한 정성적인 평가 결과와 NetRSQ의 평가 수치가 일치하게 나타난 것이다.

4. 계량서지 데이터의 적합도 측정 실험

4.1 실험 데이터와 방법

여러 선행 연구에서 분석된 38가지 실제 데이터와 직접 수집한 2가지 데이터를 합하여 총 40가지 데이터를 대상으로 NetRSQ를 측정해 보았다. 40가지 데이터에는 단어동시출현 데이터 10종, 저자동시인용 데이터 6종, 프로파일링

데이터 6종, 문헌동시인용 데이터 4종, 저자프로파일링 데이터 3종, 저자서지결합 데이터 2종, 저널동시인용 데이터 2종, 웹사이트 동시 링크 데이터 2종, 문헌서지결합 데이터 1종, 키워드서지결합 데이터 1종, 문헌유사도 데이터 1종, 문헌인용이미지구축자프로파일링 1종, 작가동시추천 데이터 1종 등의 다양한 종류가 포함되어 있다. 이중에서 직접 수집한 CW5 데이터는 빅카인즈에서 2013년부터 2017년 5월 까지 약 4년 5개월 동안의 ‘공공도서관’으로 검색된 국내 신문기사 6,824개에서 500회 이상 출현한 단어 183종의 동시출현 데이터이다. 또 하나의 직접 수집 데이터인 CW10은 저널 Scientometrics에 2010년부터 2016년 사이에 게재된 논문 1,820건에서 10회 이상 출현한 저자키워드 108종의 동시출현 데이터이다. 실험 대상인 40가지 계량서지적 데이터의 구체적인 특성과 출처는 <표 4>와 같다.

비교 대상 네트워크 표현 기법으로는 널리 사용되고 있는 전역 상대적 기준 방식인 PFNet 을 비롯해서 완전연결군집화 기법 기반인 CBNet-COMP, 평균연결군집화 기법 기반인 CBNet-WAVE, 와드군집화 기법 기반인 CBNet-WARD 의 네 가지를 실험하였다. PFNet 도출을 위해서 가중네트워크 분석 도구인 WNet(이재윤, 2013)을 사용하였고, 군집기반 네트워크 CBNet 3가지 기법은 R의 hclust 함수에 기반하여 네트워크를 도출하는 cbnet 함수를 직접 개발하여 사용하였다. NetRSQ 측정을 위해서는 연관성 행렬과 네트워크 링크 목록을 비교하는 프로그램 netrsq를 파이썬으로 직접 개발하여 사용하였다. 40가지 실험 데이터에 대한 NetRSQ 측정 결과를 비교하기 위해서는 3개 이상의 대

응표본에 대한 비모수검정 기법인 Friedman 검정 기법을 수행하였고, 사후 검정을 위해서는 Bonferroni 수정을 통한 Conover-Inman 검정을 수행하였다.

4.2 다양한 계량서지 데이터의 네트워크 표현 평가

40가지 데이터에 대해서 PFNet, 완전연결군집화 기법 기반인 CBNet-COMP, 평균연결군집화 기법 기반인 CBNet-WAVE, 와드군집화 기법 기반인 CBNet-WARD의 네 가지 네트워크 표현을 생성한 다음, $40 \times 4 = 160$ 가지 네트워크에 대해서 NetRSQ로 적합도를 측정 해보았다. NetRSQ 측정 결과는 <표 5>의 왼쪽에 제시하였다.

전체 평균은 0.2217인 PFNet이 가장 높고 평균연결기법에 기반한 CBNet-WAVE가 0.2166 으로 그 다음이며 CBNet-COMP와 CBNet-WARD의 순으로 나타났다. 그러나 개별 데이터에 대한 4개 기법 중에서 NetRSQ 성능의 순위를 집계한 <표 6>을 보면 1위를 차지한 횟수로는 CBNet-WAVE가 13회(32.5%)로 가장 높았다. 그 다음으로 PFNet과 CBNet-COMP 가 10회(25%)씩 1위를 차지했으며 평균이 가장 낮았던 CBNet-WARD도 1위를 차지한 경우가 7회(17.5%)나 되어서 기법 간 성능의 우열이 항상 일관된 것은 아님을 알 수 있다. 40 개 데이터에 대한 각 네트워크 표현 기법의 NetRSQ 성능 분포를 정리한 <그림 8>의 상자수염 그림에서도 평균이나 중앙값의 우열은 보이지만 상자 몸통이 위치한 구간은 CBNet-WAVE를 제외하면 별 차이가 없게 나타났다. 4개 기법 간

〈표 4〉 NetRSQ 측정 대상 자료집합의 특성

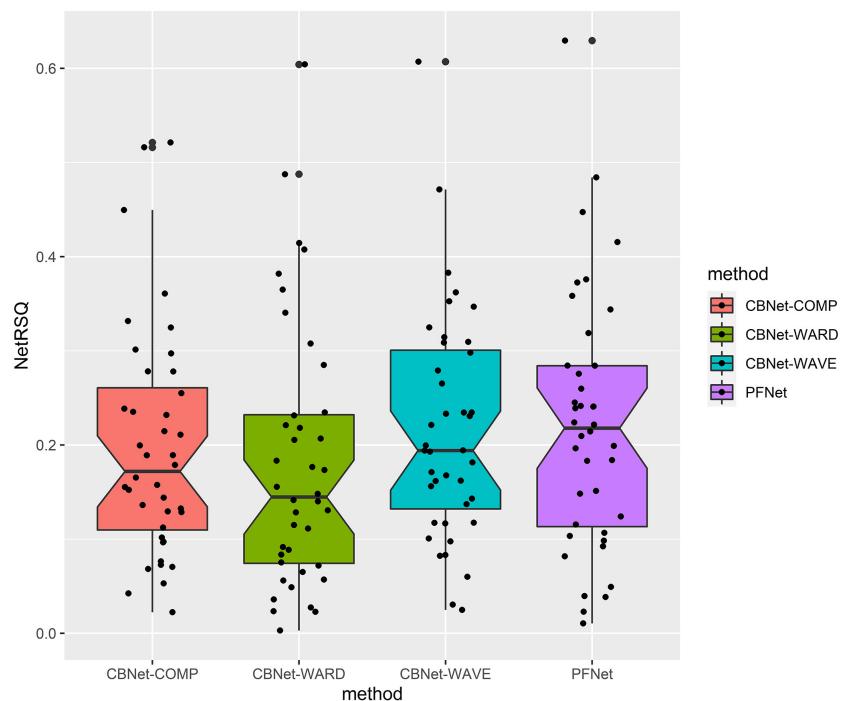
약칭	분야/주제	관계유형	대상	규모	출처
AB1	교육행정학	저자서지결합	저자	48	유희균, 엄기형(2016)
AB2	사회복지학	저자서지결합	저자	43	김희전, 조현양(2010)
AC1	정보학	저자동시인용	저자	39	White & Griffith(1981)
AC2	정보학	저자동시인용	저자	50	이은숙, 정영미(2003)
AC3	경제학	저자동시인용	저자	68	김도미(1993)
AC4	사회복지학	저자동시인용	저자	46	김희전, 조현양(2010)
AC5	언론학	저자동시인용	저자	69	김영준, 김채환(2007)
AC6	건축학	저자동시인용	저자	46	유종덕, 최은주(2011)
AP1	개혁신학	저자프로파일링	저자	41	유영준, 이재윤(2019)
AP2	국내 신학	저자프로파일링	저자	42	유영준, 이재윤(2020)
AP3	건축학	저자프로파일링	저자	46	유종덕, 최은주(2011)
AR1	문학	작가동시추천	작가	56	최상희(2017)
CL1	커뮤니케이션	동시링크	웹사이트	36	이성숙(2005)
CL2	커뮤니케이션	동시링크	웹사이트	32	정동열, 최윤미(1999)
CW1	인지과학	단어동시출현	단어	85	이재윤, 정주희(2006)
CW2	의학	단어동시출현	디스크립터	110	이재윤, 김희정(2011)
CW3	디지털도서관	단어동시출현	단어	120	Lee, Kim, Kim(2010)
CW4	오픈데이터	단어동시출현	단어	63	이재윤, 정은경(2022)
CW5	도서관 관련 신문기사	단어동시출현	단어	183	직접 수집
CW6	개혁신학	단어동시출현	단어	111	유영준, 이재윤(2019)
CW7	칼빈신학	단어동시출현	단어	56	유영준, 이재윤(2016)
CW8	국내 신학	단어동시출현	단어	92	유영준, 이재윤(2020)
CW9	재난	단어동시출현	단어	86	이재윤, 김수정(2016)
CW10	과학계량학	단어동시출현	단어	108	직접 수집
DB1	의학	문헌서지결합	문헌	268	윤정원, 정은경, 이재윤(2019)
DC1	학교도서관	문헌동시인용	문헌	91	이재윤(2020) 관련
DC2	오픈데이터	문헌동시인용	문헌	66	이재윤, 정은경(2022) 관련
DC3	딥러닝	문헌동시인용	문헌	191	이재윤(2017)
DC4	계량정보학	문헌동시인용	문헌	158	이재윤, 최상희(2013)
DP1	여성학	문헌인용이미지구축자프로파일링	문헌	148	김조아, 이재윤(2016)
DS1	산업혁명	문헌유사도	문헌	58	최상희, 이재윤(2019)
JC1	약학	저널동시인용	저널	116	조선례, 이재윤(2012)
JC2	언어학	저널동시인용	저널	38	이재윤(2019)
KB1	오픈데이터	키워드서지결합	단어	63	이재윤, 정은경(2022)
PF1	인권분야 장서	대출자 프로파일링	분류별주	63	Kim, Lee, Park(2009)
PF2	정보학	프로파일링	디스크립터	85	김판준, 이재윤(2007)
PF3	전체 학문	프로파일링	분야	140	이재윤(2008)
PF4	전체 학문	프로파일링	학문분야	38	최상희, 이재윤(2019)
PF5	재난	프로파일링	저널	36	이재윤, 김수정(2016)
PF6	디지털도서관	프로파일링	저널	54	Lee, Kim, Kim(2010)

〈표 5〉 40개 데이터에 대한 기법별 NetRSQ 측정 결과

데이터	NetRSQ (괄호 안은 순위)					NetRSQ 평균대비 성능			
	PFNet	CBNet-COMP	CBNet-WAVE	CBNet-WARD	평균	PFNet	CBNet-COMP	CBNet-WAVE	CBNet-WARD
AB1	0.3438 (1)	0.2385 (2)	0.2344 (3)	0.1284 (4)	0.2363	1.455	1.009	0.992	0.543
AB2	0.1511 (4)	0.1575 (3)	0.1815 (1)	0.1733 (2)	0.1659	0.911	0.950	1.094	1.045
AC1	0.3582 (2)	0.1523 (4)	0.2651 (3)	0.4144 (1)	0.2975	1.204	0.512	0.891	1.393
AC2	0.3187 (1)	0.1787 (3)	0.3144 (2)	0.1149 (4)	0.2317	1.376	0.771	1.357	0.496
AC3	0.2841 (2)	0.2352 (3)	0.1616 (4)	0.3076 (1)	0.2471	1.150	0.952	0.654	1.245
AC4	0.1964 (2)	0.1361 (4)	0.1620 (3)	0.2209 (1)	0.1789	1.098	0.761	0.906	1.235
AC5	0.1839 (2)	0.1552 (4)	0.1710 (3)	0.2067 (1)	0.1792	1.026	0.866	0.954	1.153
AC6	0.2408 (1)	0.1891 (3)	0.1994 (2)	0.1765 (4)	0.2015	1.195	0.939	0.990	0.876
AP1	0.2215 (1)	0.2145 (2)	0.1939 (3)	0.1305 (4)	0.1901	1.165	1.128	1.020	0.686
AP2	0.2406 (2)	0.1653 (4)	0.2789 (1)	0.2053 (3)	0.2225	1.081	0.743	1.253	0.923
AP3	0.6294 (1)	0.5212 (4)	0.6070 (2)	0.6041 (3)	0.5904	1.066	0.883	1.028	1.023
AR1	0.1240 (3)	0.1889 (1)	0.1429 (2)	0.0835 (4)	0.1348	0.920	1.401	1.060	0.619
CL1	0.2143 (2)	0.1286 (4)	0.2344 (1)	0.1555 (3)	0.1832	1.170	0.702	1.279	0.849
CL2	0.4473 (2)	0.4494 (1)	0.3828 (4)	0.4075 (3)	0.4218	1.061	1.066	0.908	0.966
CW1	0.0384 (4)	0.0726 (3)	0.0822 (2)	0.0886 (1)	0.0705	0.545	1.031	1.167	1.258
CW2	0.4841 (3)	0.5160 (1)	0.4713 (4)	0.4875 (2)	0.4897	0.989	1.054	0.962	0.995
CW3	0.2755 (2)	0.3315 (1)	0.2212 (4)	0.2344 (3)	0.2657	1.037	1.248	0.833	0.882
CW4	0.0492 (4)	0.0971 (1)	0.0831 (2)	0.0718 (3)	0.0753	0.653	1.290	1.104	0.954
CW5	0.1033 (3)	0.0965 (4)	0.1561 (1)	0.1112 (2)	0.1168	0.885	0.826	1.337	0.952
CW6	0.0983 (2)	0.1016 (1)	0.0975 (3)	0.0570 (4)	0.0886	1.109	1.147	1.100	0.643
CW7	0.1482 (3)	0.1439 (4)	0.1942 (1)	0.1832 (2)	0.1674	0.885	0.860	1.160	1.095
CW8	0.0395 (4)	0.0529 (2)	0.0599 (1)	0.0488 (3)	0.0503	0.786	1.052	1.191	0.971
CW9	0.1989 (2)	0.2318 (1)	0.1927 (3)	0.1480 (4)	0.1929	1.031	1.202	0.999	0.767
CW10	0.0228 (4)	0.0423 (1)	0.0304 (2)	0.0273 (3)	0.0307	0.743	1.378	0.990	0.889
DB1	0.0104 (3)	0.0223 (2)	0.0247 (1)	0.0029 (4)	0.0151	0.690	1.479	1.638	0.192
DC1	0.0817 (3)	0.1123 (2)	0.1172 (1)	0.0234 (4)	0.0837	0.977	1.342	1.401	0.280
DC2	0.2414 (2)	0.2549 (1)	0.1370 (4)	0.2182 (3)	0.2129	1.134	1.197	0.644	1.025
DC3	0.1154 (1)	0.0762 (3)	0.1006 (2)	0.0651 (4)	0.0893	1.292	0.853	1.126	0.729
DC4	0.2094 (1)	0.1324 (4)	0.1676 (2)	0.1414 (3)	0.1627	1.287	0.814	1.030	0.869
DP1	0.1830 (2)	0.1293 (3)	0.2306 (1)	0.0751 (4)	0.1545	1.184	0.837	1.493	0.486
DS1	0.0922 (2)	0.0682 (3)	0.1174 (1)	0.0227 (4)	0.0751	1.227	0.908	1.563	0.302
JC1	0.4155 (1)	0.2781 (4)	0.3525 (3)	0.3818 (2)	0.3570	1.164	0.779	0.987	1.070
JC2	0.2842 (4)	0.3247 (2)	0.3087 (3)	0.3403 (1)	0.3145	0.904	1.033	0.982	1.082
KB1	0.2238 (2)	0.2108 (3)	0.2330 (1)	0.0559 (4)	0.1809	1.237	1.165	1.288	0.309
PF1	0.1066 (2)	0.0704 (4)	0.1166 (1)	0.0915 (3)	0.0963	1.107	0.731	1.211	0.950
PF2	0.3758 (1)	0.3012 (3)	0.3620 (2)	0.2848 (4)	0.3310	1.136	0.910	1.094	0.861
PF3	0.2390 (3)	0.3607 (1)	0.3468 (2)	0.1400 (4)	0.2716	0.880	1.328	1.277	0.515
PF4	0.2449 (3)	0.2779 (2)	0.3248 (1)	0.2313 (4)	0.2697	0.908	1.030	1.204	0.858
PF5	0.3725 (1)	0.2972 (3)	0.2979 (2)	0.0359 (4)	0.2509	1.485	1.185	1.187	0.143
PF6	0.2597 (3)	0.1995 (4)	0.3094 (2)	0.3649 (1)	0.2834	0.916	0.704	1.092	1.288
평균	0.2217	0.1978	0.2166	0.1816	0.1447	1.052	1.002	1.111	0.835
중앙값	0.2179	0.1720	0.1941	0.1447		1.074	0.981	1.094	0.886
표준편차	0.1361	0.1214	0.1211	0.1407		0.207	0.226	0.213	0.314

〈표 6〉 각 기법의 40개 데이터별 NetRSQ 순위 집계

	PFNet	CBNet-COMP	CBNet-WAVE	CBNet-WARD
1위 횟수	10 (25.0%)	10 (25.0%)	13 (32.5%)	7 (17.5%)
2위 횟수	15 (37.5%)	7 (17.5%)	13 (32.5%)	5 (12.5%)
3위 횟수	9 (22.5%)	11 (27.5%)	9 (22.5%)	11 (27.5%)
4위 횟수	6 (15.0%)	12 (30.0%)	5 (12.5%)	17 (42.5%)
합계	40 (100.0%)	40 (100.0%)	40 (100.0%)	40 (100.0%)



〈그림 8〉 40개 데이터에 대한 기법별 NetRSQ 측정 결과 분포

NetRSQ 성능 차이를 3개 집단 이상 대응표본에 대한 비모수검정 기법인 Friedman 검정으로 수행해본 결과에서는 p 값이 0.0245로서 기법 간 성능 차이가 95% 유의수준에서 통계적으로 유의하였다. 사후 검정을 위해서 Bonferroni 수정을 통한 Conover-Inman 검정을 수행한 결과, 95% 유의수준에서 평균 2위인 CBNet-WAVE의 NetRSQ와 평균 4위인 CBNet-WARD

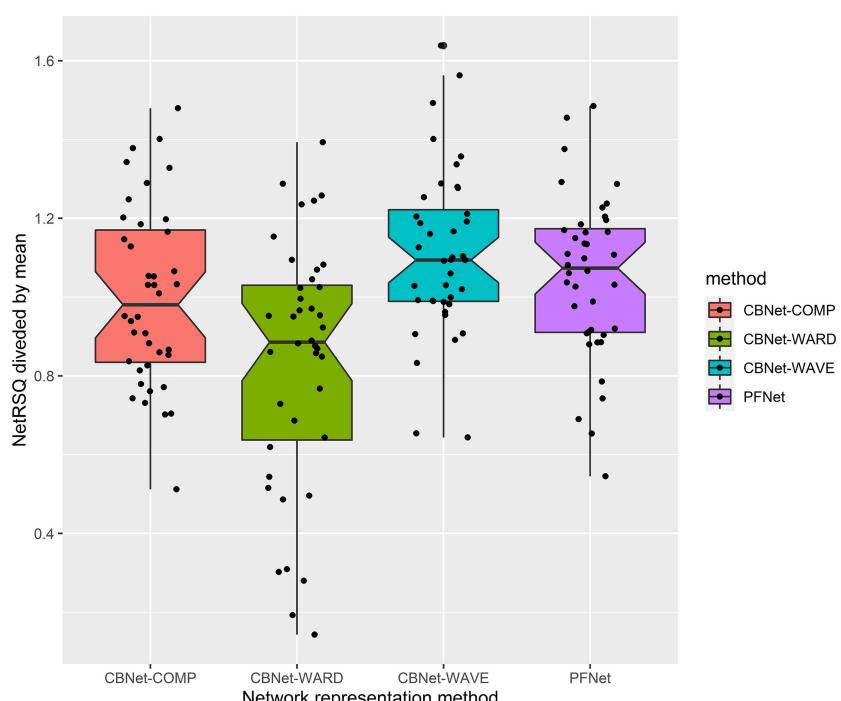
의 NetRSQ가 통계적으로 유의한 차이가 있는 것으로 나타났고($p=0.04$), 평균 1위인 PFNet의 NetRSQ와 CBNet-WARD의 NetRSQ는 통계적으로 유의하지 않은 것으로 나타났다($p=0.13$).

사후검정결과 NetRSQ 평균 차이가 가장 큰 PFNet과 CBNet-WARD 간의 통계적인 차이는 유의하지 않은데, 이보다 NetRSQ 평균 차이가 적은 CBNet-WAVE와 CBNet-WARD

간의 통계적인 차이는 유의한 이유는 40개 데이터 별로 각 네트워크 표현 기법의 NetRSQ를 측정한 것이므로 비교 집단이 독립적이지 않고 관찰값이 대응되기 때문이다. NetRSQ 측정값은 데이터의 특성에 크게 좌우되며, 특히 표현해야 하는 개체의 수가 많아지면 네트워크 표현에서 개체 간 거리가 입력 데이터에 비해서 왜곡될 가능성이 더 커진다. 실제로 각 데이터의 규모(개체 수)와 NetRSQ 수치와의 상관관계수를 구해보면 PFNet은 -0.357, CBNet-COMP는 -0.285, CBNet-WAVE는 -0.324, CBNet-WARD는 -0.310으로서 평균 -0.319의 부정적 상관관계가 얻어진다. 즉 규모와 NetRSQ는 다소 반비례하는 성향이 있는 것이다. 따라서 각 데

이터의 NetRSQ 평균으로 각 기법의 NetRSQ 수치를 나누어서 살펴보면 기법 간 성능차이를 더 뚜렷하게 살펴볼 수 있다.

〈표 5〉의 오른쪽에 데이터별 성능 평균으로 나누어 보정한 NetRSQ 값을 제시하였고, 〈그림 9〉에 보정된 값의 분포를 제시하였다. NetRSQ 값의 평균과 중앙값은 4가지 기법 중에서 PFNet이 가장 높지만, 데이터 별로 각각 4개 기법의 평균 수치로 보정한 NetRSQ 값의 평균과 중앙값은 CBNet-WAVE가 가장 높은 것으로 나타났다. 하지만 CBNet-WAVE와 PFNet, CBNet-COMP 사이의 NetRSQ 차이는 통계적으로 유의하지 않았다.



〈그림 9〉 40개 데이터에 대한 기법별 NetRSQ 측정 결과를 데이터별 평균 성능으로 나눈 값의 분포

5. 결 론

이 연구에서는 2000년대 중반 이후 지적구조 분석 연구에서 널리 사용되고 있는 PFNet을 비롯한 네트워크 표현에 대해서 적합도를 측정하는 방법을 개발하고자 하였다. 네트워크 적합도 지표로 제안된 NetRSQ는 입력된 연관성 데이터와 생성된 네트워크에서의 경로 길이 사이의 순위 상관계수를 제곱하여 산출하는 방식이다.

가상 연관성 데이터를 네트워크로 표현한 두 가지 결과에 대해서 NetRSQ를 측정해본 결과, 입력된 연관성과 최단 경로길이의 순위가 더 일치하면 NetRSQ도 더 높게 산출되는 것이 확인되었다. 실제 계량서지 데이터에 대한 정성적인 평가와 NetRSQ를 비교하기 위해서 국내 학문분야 140개 사이의 연관성을 네트워크로 표현했던 선행연구의 3가지 네트워크에 대해서 NetRSQ를 측정해보았다. 그 결과 네트워크 표현 구조에 대한 정성적인 평가가 높을수록 NetRSQ도 높게 측정됨을 확인하였다.

40가지 계량서지적 데이터에 대해서 PFNet과 군집기반 네트워크 표현 기법 3가지를 적용하여 NetRSQ를 측정해본 결과, 1위를 차지한 횟수는 CBNet-WAVE가 13회(32.5%)로 가장 많았으며 PFNet과 CBNet-COMP가 10회(25%)씩으로 그 다음이었고, CBNet-WARD가 7회(17.5%)로 가장 적었다. 통계 검증 결과에서는 CBNet-WAVE와 CBNet-WARD의 NetRSQ 차이가 유의한 것으로 나타났고 나머지 기법 간 차이는 유의하지 않았다. 특정 기법에 의한 네트워크 표현 결과가 항상 좋은 품질을 보이는 것은 아니며, 반대로 특정 기법이 항상 나쁜 품질을 보이는 것도 아님을 알 수 있었

다. 따라서 연관성 데이터를 네트워크로 표현할 경우에는 가능한 여러 표현 방법을 시도한 후 NetRSQ로 측정한 적합도가 높고 내용 해석에 무리가 없는 결과를 선택해야 할 것이다.

이 연구에서 제안한 NetRSQ는 네트워크 표현 결과의 품질을 측정하여 최적의 표현 기법을 선택하는 근거가 되므로 지적구조의 시각적인 표현 개선에 기여할 수 있다. 예를 들어 특정한 연구분야에 대해서 단어동시출현분석이나 저자 동시인용분석 등을 수행하면서 PFNet 알고리즘으로 네트워크를 도출하였을 때, NetRSQ가 낮게 측정된다면 다른 대안을 찾아봐야 할 것이다. 이 연구에서 분석한 40개 데이터집합의 PFNet에 대한 NetRSQ는 평균이 0.2217, 중앙 값이 0.2179였다. 따라서 NetRSQ가 0.22 이하라면 보통 수준에 못미치는 결과로 간주할 수 있다. 그럴 경우에는 PFNet의 대안인 CBNet을 적용해보고 조금이라도 더 나은 NetRSQ가 측정되는 네트워크를 선택하는 것이 바람직하다.

향후 지적구조를 네트워크 형태로 표현한 결과를 제시할 때 NetRSQ를 측정하여 함께 보고한다면 시각적 표현 결과에 대한 해석에도 도움이 될 수 있다. NetRSQ가 높은 네트워크라면 표현된 구조가 신뢰성이 높으므로 시각적으로 인식되는 결과를 중심으로 지적구조를 해석하면 된다. 반면에 NetRSQ가 낮게 측정된 네트워크는 시각적으로 표현된 구조에만 의존하지 말고 입력된 연관성 데이터를 보완적으로 활용하면서 네트워크 구조를 해석하는 것이 바람직하다. 또한 앞으로 PFNet이나 CBNet 알고리즘의 도출 결과를 개선하는 방법을 시도할 때에도 NetRSQ를 통한 네트워크 구조 평가가 큰 도움이 될 것으로 기대된다.

참 고 문 헌

- 김도미 (1993). 저자동시인용(著者同時引用) 분석과 인용한 문헌(文獻)의 색인어(索引語) 분석(分
析)에 의한 지적구조(知的構造)의 규명: 경제학(經濟學) 분야를 대상으로. 정보관리연구, 24(1),
32-57.
- 김영준, 김채환 (2007). 국내 언론학의 지적 구조 매핑(Mapping)에 관한 연구: 1989-2006년 저자동시
인용 분석. 커뮤니케이션학 연구, 15(3), 156-185.
- 김조아, 이재윤 (2016). 인용 이미지 구축자 프로파일링을 이용한 국내 여성학 분야 연구 전선 분석.
정보관리학회지, 33(2), 201-225. <http://doi.org/10.3743/KOSIM.2016.33.2.201>
- 김판준, 이재윤 (2007). 연구 영역 분석을 위한 디스크립터 프로파일링에 관한 연구. 정보관리학회지,
24(4), 285-303. <http://doi.org/10.3743/KOSIM.2007.24.4.285>
- 김희전, 조현양 (2010). 저자동시인용분석과 저자서지결합분석에 의한 지적 구조 분석: 사회복지학 분
야를 중심으로. 정보관리학회지, 27(3), 283-306.
<http://doi.org/10.3743/KOSIM.2010.27.3.283>
- 유영준, 이재윤 (2016). 칼빈 신학교 학술지에 대한 계량서지학적 분석에 관한 연구. 한국비블리아학회
지, 27(4), 125-145. <http://doi.org/10.14699/kbiblia.2016.27.4.125>
- 유영준, 이재윤 (2019). 외국 개혁신학 학술지에 대한 계량서지학적 연구. 한국비블리아학회지, 30(3),
149-170. <http://doi.org/10.14699/kbiblia.2019.30.3.149>
- 유영준, 이재윤 (2020). KCI 등재 신학 학술지에 대한 계량서지학적 분석. 한국비블리아학회지, 31(3),
5-27. <http://doi.org/10.14699/kbiblia.2020.31.3.005>
- 유종덕, 최은주 (2011). 저자프로파일링분석과 저자동시인용분석의 유용성 비교 검증. 정보관리학회지,
28(1), 123-144. <https://doi.org/10.3743/KOSIM.2011.28.1.123>
- 유희균, 엄기형 (2016). 한국 교원정책연구의 지적구조 탐색: 저자서지결합 분석을 중심으로. 교육행정
학연구, 34(5), 297-327.
- 윤정원, 정은경, 이재윤 (2019). An Investigation on Scholarly Publications using HINTS Data by
Bibliometric Analysis. 사회과학연구논총, 35(1), 121-152.
<http://doi.org/10.16935/ejss.2019.35.1.004>
- 이성숙 (2005). 동시링크분석을 이용한 웹정보원의 지적구조 변화에 관한 연구. 정보관리학회지, 22(2),
205-228. <http://doi.org/10.3743/KOSIM.2005.22.2.205>
- 이은숙 (2003). 복수저자를 고려한 저자동시인용분석 연구: 정보학과 컴퓨터과학을 대상으로. 석사학
위논문, 연세대학교 대학원 문헌정보학과.
- 이재윤 (2006). 지적 구조의 규명을 위한 네트워크 형성 방식에 관한 연구. 한국문헌정보학회지, 40(2),

- 333-355. <http://doi.org/10.4275/KSLIS.2006.40.2.333>
- 이재윤 (2007). 클러스터링 기반 네트워크 생성 알고리즘. 제14회 한국정보관리학회 학술대회 논문집, 147-154.
- 이재윤 (2008). 연구자의 투고 학술지 현황에 근거한 국내 학문분야 네트워크 분석. *정보관리학회지*, 25(4), 327-345. <http://doi.org/10.3743/KOSIM.2008.25.4.327>
- 이재윤 (2012). 네트워크 표현을 위한 합치도 지표. 제19회 한국정보관리학회 학술대회 논문집, 133-136.
- 이재윤 (2013). tnet과 WNET의 가중 네트워크 중심성 지수 비교 연구. *정보관리학회지*, 30(4), 241-264. <http://doi.org/10.3743/KOSIM.2013.30.4.241>
- 이재윤 (2017). 자아 중심 주제 인용분석을 활용한 딥러닝 연구동향 분석. *정보관리학회지*, 34(4), 7-32. <http://doi.org/10.3743/KOSIM.2017.34.4.007>
- 이재윤 (2019). 학술지 『언어사실과 관점』 연구동향의 자아중심 네트워크 분석. 제117회 연세대학교 언어정보연구원 학술발표회 발표자료.
- 이재윤 (2020). 인용가중 저자프로파일링을 이용한 학교도서관 연구의 지적구조 분석. *한국문헌정보학회지*, 54(2), 197-223. <http://dx.doi.org/10.4275/KSLIS.2020.54.2.197>
- 이재윤, 김수정 (2016). 국내 재난 관련 연구 동향에 대한 계량정보학적 분석. *정보관리학회지*, 33(4), 103-124. <http://doi.org/10.3743/KOSIM.2016.33.4.103>
- 이재윤, 김희정 (2011). 연구지원 서비스를 위한 계량서지적 분석: 국제백신연구소 연구동향을 대상으로. 제18회 한국정보관리학회 학술대회 논문집, 11-16.
- 이재윤, 정은경 (2022). 지적구조 규명을 위한 키워드서지결합분석 기법에 관한 연구. *정보관리학회지*, 39(1), 309-330. <https://doi.org/10.3743/KOSIM.2022.39.1.309>
- 이재윤, 정주희 (2006). 연구자 소속과 표제어 분석을 통한 국내 인지과학 분야의 학제적 구조 파악: 제13회 한국정보관리학회 학술대회 논문집, 127-134.
- 이재윤, 최상희 (2013). 계량정보학분야의 협력연구 네트워크 및 문헌네트워크 분석: 국가, 기관, 문헌 단위 분석. *정보관리학회지*, 30(1), 179-191. <http://doi.org/10.3743/KOSIM.2013.30.1.179>
- 정동열, 최윤미 (1999). 웹 정보원의 동시인용분석에 관한 실험적 연구. *정보관리학회지*, 16(2), 7-26.
- 조선례, 이재윤 (2012). 약학 분야 학술정보서비스를 위한 학술지 동시인용 분석. *정보관리연구*, 43(1), 159-185. <http://dx.doi.org/10.1633/JIM.2012.43.1.159>
- 최상희 (2017). 독자 추천도서 정보를 이용한 작가 이미지 분석 연구. *정보관리학회지*, 34(4), 153-171. <http://doi.org/10.3743/KOSIM.2017.34.4.153>
- 최상희, 이재윤 (2019). 학술용어로서 산업혁명과 4차산업혁명의 사용 양상과 관련 키워드 분석. *인문과 학연구논총*, 40(4), 157-188. <http://doi.org/10.22947/ihmju.2019.40.4.006>
- Arleo, A., Kwon, O. H., & Ma, K. L. (2017). GraphRay: distributed pathfinder network scaling. *Proceedings of the IEEE 7th Symposium on Large Data Analysis and Visualization (LDAV)*,

74-83.

- Barabási, A. L. (2016). Network Science. UK: Cambridge University Press.
- Hauguel, S., Zhai, C., & Han, J. (2009). Parallel pathfinder algorithms for mining structures from graphs. 2009 Ninth IEEE International Conference on Data Mining, 812-817.
<https://doi.org/10.1109/ICDM.2009.142>
- Kim, Pan Jun, Lee, Jae Yun, & Park, Ji-Hong (2009). Developing a new collection-evaluation method: mapping and the user-side h-index. Journal of the American Society for Information Science and Technology, 60(11), 2366-2377. <https://doi.org/10.1002/asi.21159>
- Lee, Jae Yun, Kim, Heejung, & Kim, Pan Jun (2010). Domain analysis with text mining: analysis of digital library research trends using profiling methods. Journal of Information Science, 36(2), 144-161. <https://doi.org/10.1177/0165551509353251>
- McCain, K. W. (1986). Cocited author mapping as a valid representation of intellectual structure. Journal of the American Society for Information Science, 37(3), 111-122.
[https://doi.org/10.1002/\(SICI\)1097-4571\(198605\)37:3<111::AID-ASI2>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1097-4571(198605)37:3<111::AID-ASI2>3.0.CO;2-D)
- Quirin, A., Cordón, O., Guerrero-Bote, V. P., Vargas-Quesada, B., & Moya-Anegón, F. (2008). A quick MST-based algorithm to obtain Pathfinder networks ($\infty, n - 1$). Journal of the American Society for Information Science and Technology, 59(12), 1912-1924.
<https://doi.org/10.1002/asi.20904>
- Schvaneveldt, R. W. (1990). Pathfinder Associative Networks: Studies in Knowledge Organization. Norwood, NJ: Ablex Publishing Corporation.
- Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. Psychometrika, 17, 401-419. <https://doi.org/10.1007/BF02288916>
- 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.
<https://doi.org/10.1002/asi.4630320302>
- White, H. D. (2003). Pathfinder networks and author cocitation analysis: a remapping of paradigmatic information scientists. Journal of the American Society for Information Science and Technology, 54(5), 423-434. <https://doi.org/10.1002/asi.10228>
- Young, F. W. (1987). Multidimensional Scaling: History, Theory, and Applications. Hillsdale, N.J.: L. Erlbaum Associates.
- Zweig, K. A. (2016). Network Analysis Literacy: A Practical Approach to the Analysis of Networks. Austria: Springer-Verlag. <https://doi.org/10.1007/978-3-7091-0741-6>

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Choi, Sanghee & Lee, Jae Yun (2019). Analyzing use of the words 'Industrial Revolution' and 'the 4th Industrial Revolution' and related keywords in academic disciplines. *The Journal of Humanities*, 40(4), 157-188. <http://doi.org/10.22947/ihmju.2019.40.4.006>
- Choi, Sanghee (2017). Analysis of author image based on book recommendation from readers. *Journal of the Korean Society for Information Management*, 34(4), 153-171. <http://doi.org/10.3743/KOSIM.2017.34.4.153>
- Jeong, Dong-Youl & Choi, Yoon Mee (1999). An experimental study of cocitation analysis on web information. *Journal of the Korean Society for Information Management*, 16(2), 7-26.
- Joe, Seon-Rye & Lee, Jae Yun (2012). Journal co-citation analysis for library services in pharmaceutics. *Journal of Information Management*, 43(1), 159-185. <http://dx.doi.org/10.1633/JIM.2012.43.1.159>
- Kim, Do-Mi (1993). A study on intellectual structure using author co-citation analysis and indexing term analysis of citing documents: application to economics. *Journal of Information Management*, 24(1), 32-57.
- Kim, Heejeon & Cho, Hyun Yang (2010). A study on intellectual structure using author co-citation analysis and author bibliographic coupling analysis in the field of social welfare science. *Journal of the Korean Society for Information Management*, 27(3), 283-306. <http://doi.org/10.3743/KOSIM.2010.27.3.283>
- Kim, Jo-Ah & Lee, Jae Yun (2016). Analyzing the research fronts of women's studies in Korea using citation image makers profiling. *Journal of the Korean Society for Information Management*, 33(2), 201-225. <http://doi.org/10.3743/KOSIM.2016.33.2.201>
- Kim, Pan Jun & Lee, Jae Yun (2007). Descriptor profiling for research domain analysis. *Journal of the Korean Society for Information Management*, 24(4), 285-303. <http://doi.org/10.3743/KOSIM.2007.24.4.285>
- Kim, Youngjoon & Kim, ChaeHwan (2007). Mapping the intellectual structure of communication research field in Korea: an author co-citation analysis, 1989~2006. *Korean Journal of Communication Studies*, 15(3), 156-185.
- Lee, Eun Suk (2003). An Author Co-citation Analysis of Multiple Authorship in the Fields of Information Science and Computer Science. Master's thesis, Graduate School of Yonsei University.

- Lee, Jae Yun & Choi, Sanghee (2013). Collaboration networks and document Networks in informetrics research from 2001 to 2011: Finding influential nations, institutions, documents. Journal of the Korean Society for Information Management, 30(1), 179-191.
<http://doi.org/10.3743/KOSIM.2013.30.1.179>
- Lee, Jae Yun & Chung, EunKyung (2022). Introducing Keyword Bibliographic Coupling Analysis (KBCA) for identifying the intellectual structure. Journal of the Korean Society for Information Management, 39(1), 309-330, <https://doi.org/10.3743/KOSIM.2022.39.1.309>
- Lee, Jae Yun & Jung, Ju Hee (2006). Examining the interdisciplinary structure of Korean cognitive science through analyzing author affiliations and title words. Proceedings of the 13th Conference of the Korean Society for Information Management, 127-134.
- Lee, Jae Yun & Kim, Heejung (2011). Bibliometric analysis for the research support service at International Vaccine Institute. Proceedings of the 18th Conference of the Korean Society for Information Management, 11-16.
- Lee, Jae Yun & Kim, Soojung (2016). A bibliometric analysis of research trends on disaster in Korea. Journal of the Korean Society for Information Management, 33(4), 103-124.
<http://doi.org/10.3743/KOSIM.2016.33.4.103>
- Lee, Jae Yun (2006). A study on the network generation methods for examining the intellectual structure of knowledge domains. Journal of the Korean Society for Library and Information Science, 40(2), 333-355. <http://doi.org/10.4275/KSLIS.2006.40.2.333>
- Lee, Jae Yun (2007). Clustering-based network generation algorithm. Proceedings of the 14th Conference of Korean Society for Information Management, 147-154.
- Lee, Jae Yun (2008). Analyzing the network of academic disciplines with journal contributions of Korean researchers. Journal of the Korean Society for Information Management, 25(4), 327-345. <http://doi.org/10.3743/KOSIM.2008.25.4.327>
- Lee, Jae Yun (2012). A measure of goodness of fit for network representations. Proceedings of the 19th Conference of Korean Society for Information Management, 133-136.
- Lee, Jae Yun (2013). A comparison study on the weighted network centrality measures of tnet and WNET. Journal of the Korean Society for Information Management, 30(4), 241-264.
<http://doi.org/10.3743/KOSIM.2013.30.4.241>
- Lee, Jae Yun (2017). Deep learning research trends analysis with ego centered topic citation analysis. Journal of the Korean Society for Information Management, 34(4), 7-32.
<http://doi.org/10.3743/KOSIM.2017.34.4.007>
- Lee, Jae Yun (2019). Analyzing research trends of the journal Language Facts and Perspectives.

- with ego-centered network analysis. Presentations at the 117th Workshop of the Institute of Language & Information Studies, Yonsei University.
- Lee, Jae Yun (2020). Analyzing the intellectual structure of school library researches with citation-weighted author profiling. *Journal of the Korean Society for Library and Information Science*, 54(2), 197-223. <http://doi.org/10.4275/KSLIS.2020.54.2.197>
- Lee, Sung-Sook (2005). A study on changes of the intellectual structure in web information using the co-links analysis. *Journal of the Korean Society for Library and Information Science*, 22(2), 205-228. <http://doi.org/10.3743/KOSIM.2005.22.2.205>
- Riew, Hee Kyun & Eum, Key Hyoung (2016). An inquiry on the 'intellectual structure' in Korea teacher policy researches: Focusing on author bibliographic coupling analysis. *Korean Journal of Educational Administration*, 34(5), 297-327.
- 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.
<https://doi.org/10.3743/KOSIM.2011.28.1.123>
- Yoo, Yeong Jun & Lee, Jae Yun (2016). A bibliographic study on the Calvin Theological Journal. *Journal of the Korean Biblia Society for Library and Information Science*, 27(4), 125-145. <http://doi.org/10.14699/kbiblia.2016.27.4.125>
- Yoo, Yeong Jun & Lee, Jae Yun (2019). A bibliometric study on foreign reformed theological journals. *Journal of the Korean Biblia Society for Library and Information Science*, 30(3), 149-170. <http://doi.org/10.14699/kbiblia.2019.30.3.149>
- Yoo, Yeong Jun & Lee, Jae Yun (2020). A bibliometric study on the KCI listed theological journals. *Journal of the Korean Biblia Society for Library and Information Science*, 31(3), 5-27. <http://doi.org/10.14699/kbiblia.2020.31.3.005>
- Yoon, JungWon, Chung, Eunkyoung, & Lee, Jae Yun (2019). An investigation on scholarly publications using HINTS data by bibliometric analysis. *Ewha Journal of Social Sciences*, 35(1), 121-152. <http://doi.org/10.16935/ejss.2019.35.1.004>