

계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구

Centrality Measures for Bibliometric Network Analysis

이 재 윤(Jae-Yun Lee)*

목 차

1. 서론	3.3 중심성 분석
2. 중심성 척도	3.4 측정 범위에 따른 삼각매개중심성의 표출 유형
2.1 사회 네트워크 분석과 중심성 척도	4. 기타 계량서지적 네트워크의 중심성 분석
2.2 계량서지적 네트워크를 위한 중심성 척도	4.1 용어 동시출현 네트워크 분석
3. 저자동시인용 네트워크의 중심성 분석	4.2 동시링크 네트워크 분석
3.1 저자동시인용 데이터	5. 결 론
3.2 저자동시인용을 통한 지적구조의 시각화	

초 록

최근 일부 계량서지학 연구에서 사회 네트워크 분석 분야의 도구인 중심성 분석 기법을 도입하려는 시도가 나타나고 있다. 그런데 사회 네트워크를 대상으로 개발된 중심성 척도는 노드간 연결에 대한 가중치를 반영하지 않는 단점이 있다. 이 연구에서는 가중 네트워크인 계량서지적 자료에 대해서 중심성을 분석할 수 있는 새로운 척도를 네 가지 제시하였다. 제안된 척도의 유용성을 검증하기 위해서 저자동시인용 네트워크, 용어 동시출현 네트워크, 웹 사이트 동시링크 네트워크의 세 가지 실제 자료에 대해서 중심성 분석을 수행하였다. 분석 결과에서는 제안된 중심성 척도가 계량서지적 네트워크를 대상으로 개별 노드의 입지와 영향력을 파악하는데 있어서 매우 유용함이 확인되었다.

ABSTRACT

Recently, some bibliometric researchers tried to use the centrality analysis methods and the centrality measures which are standard tools in social network analysis. However the traditional centrality measures originated from social network analysis could not deal with weighted networks such as co-citation networks. In this study, new centrality measures for analyzing bibliometric networks with link weights are suggested and applied to three real network data, including an author co-citation network, a co-word network, and a website co-link network. The results of centrality analyses in these three cases can be regarded as promising the usefulness of suggested centrality measures, especially in analyzing the position and influence of each node in a bibliometric network.

키워드: 계량서지학, 네트워크 분석, 중심성 분석, 중심성 척도, 가중 네트워크, 저자동시인용 분석, 용어동시출현 분석, 동시링크 분석

Bibliometrics, Network Analysis, Centrality Analysis, Centrality Measures, Weighted Networks, Author Co-citation Analysis, Co-word Analysis, Co-link Analysis

* 경기대학교 문헌정보학과 조교수(memexlee@kgu.ac.kr)
논문접수일자 2006년 8월 14일
게재확정일자 2006년 9월 17일

1. 서론

사회 네트워크 분석은 1960년대와 1970년대에 걸쳐서 사회학과 사회심리학 분야의 연구자들이 개발한 학제적인 방법론이다. 1970년대 이후 수학, 통계학, 전산학 분야 연구자들이 가세하여 분석 기법을 체계화함에 따라서 경제학, 마케팅, 산업공학 등의 여타 분야에 유용한 분석 기법으로 자리잡게 되었다(Scott 2000).

문헌정보학 영역에서 사회 네트워크 분석과 관련된 연구는 크게 보면 두 분야에서 수행되어 왔다. 한 분야는 정보이용자 집단에 대한 분석이고 다른 분야는 계량서지학 분야이다. 물론 두 계열의 출발점은 “보이지 않는 대학(invisible college)”에 대한 연구라고 할 수 있다. Price (1963)가 현대화된 용어인 “보이지 않는 대학”에 대한 연구는 이후 동시인용분석을 비롯한 계량서지적 분석 기법의 발전에 따라서 출판물을 통해서 접근하는 방향으로 발전해왔다. 한편 계량사회학 계열의 사회 네트워크 연구자들도 일찍부터 “보이지 않는 대학”에 관심을 두고 비공식 채널을 통한 커뮤니케이션의 주요 사례로 연구해왔다(Crane 1969; 1972). “보이지 않는 대학”에 대한 연구에 있어서 계량서지적 분석은 주로 학자들의 상호 인용 위주의 상호작용에 초점을 맞추는 반면에, 계량사회학 분야에서는 학자들이 동일한 국제 학술대회에 참석하는 것과 같이 더욱 사회적인 상호작용을 밝히는 데 초점을 둔다고 할 수 있다(Zuccalla 2006).

일리노이 대학 문헌정보학 대학원 교수이며 여러 연구에서 사회 네트워크 분석을 적용한 C. Haythornthwaite는 사회 네트워크 분석이 정보에 대한 연구를 수행할 때 있어서 새로운 차

상과 분석 기법을 풍부하게 제공해준다고 강조하였다. 그녀는 사서나 정보제공자가 사회 네트워크 분석을 도입하여 정보 요구, 정보 노출, 정보 합리화, 정보 경로, 정보 기획 등과 관련된 측면에 대해서 연구할 여지가 있다고 설명하였다(Haythornthwaite 1996). 이는 계량사회학적인 측면에서 문헌정보학에 사회 네트워크 분석을 도입하려는 것이라고 할 수 있다. 이와 같은 입장의 연구에 대해서는 국내에서도 유사라(1991)가 소개한 바 있다. 사회 네트워크 분석을 정보이용자 집단 분석에 도입한 구체적인 사례 연구는 지역사회 의료기관간 정보흐름을 분석한 HISP 프로젝트(Dosa 1982; National Library of Medicine 1979), 프로젝트 수행자들의 학제간 협력 행태에 대해서 조사한 Haythornthwaite(2006), 노숙자들 사이의 정보 교환과 확산에 대해서 조사한 Hersberger(2003), 전장(戰場)에서의 집단적이면서 역동적인 정보 이용 행태에 대해 연구한 Sonnewald와 Pierce (2000) 등과 같이 매우 다양한 영역에 걸쳐서 이루어졌음을 알 수 있다.

인용망을 비롯한 계량서지적 자료에 대해서 사회 네트워크 분석을 도입하는 연구는, 인용망이 본질적으로 더 네트워크에 가까움에도 불구하고 이용자집단에 대한 적용에 비해서 뒤늦게 확산되고 있다. 물론 사회 네트워크 연구의 선구자로서 문헌정보학 분야 학술지에 학술지 인용망에 대한 분석 연구(Doreian 1985; Doreian & Fararo 1985)를 수차례 발표한 사회학자 P. Doreian이 있었으나 문헌정보학계의 후속 연구를 이끌어 내지는 못하였다. Otte와 Rousseau (2002)는 사회 네트워크 분석이 정보학계에도 유용한 도구가 될 수 있다고 지적하면서 주로

인용을 비롯한 계량서지 자료에 대한 적용 사례를 소개하였다. 계량정보학 영역 중에서는 특히 공저자를 비롯한 협동연구 분석에서 네트워크 분석 기법이 주로 사용되어왔다. 협동연구 네트워크는 공저자 네트워크에 대한 분석이 대부분인데, 공저의 단위를 개인저자로 하는 경우(Newman 2004a; Yoshikane & Kageura 2004)나 기관 단위(Nagpaul 2002) 혹은 국가 단위(Dastidar 2004; Glänzel et al. 2003; Schubert 2002)로 하는 경우가 있다. 2003년 ASIST 연례 학술대회의 패널세션으로 사회 네트워크 분석과 인용 네트워크 분석에 대한 토론이 진행되었다. 이 토론회의 준비 자료에서는 인용분석에 익숙한 정보화학자들이 아직은 사회 네트워크 분석에 대해서는 잘 알지 못한다고 언급하면서, 두 기법이 동일한 측면을 평가하는 것이 아니므로 학술 공동체에 대한 분석에 있어서 상호 보완적인 도구가 될 수 있다고 결론을 맺었다(Marion et al. 2003).

사회 네트워크 분석 기법 중에서 계량서지적 자료에 대해서 가장 활발히 적용되는 것은 중심성 분석이다. 중심성 분석은 중심성 척도라는 지표를 이용해서 조직 내 개체의 지위와 역할을 해석하는 기법으로서 1950년대부터 시도되어왔다. 사회 네트워크 분야에서 중심성 척도에 대한 이론은 Freeman(1979)이 체계적으로 정립하였으며, 그가 제시한 세 가지 중심성 척도가 대표적인 척도로 널리 사용되고 있다.

이 연구는 사회 네트워크 분석 분야에서 사용되어온 전통적인 중심성 척도를 계량서지적 네트워크에 효율적으로 적용하기 위한 방안을 모색하는 것을 목적으로 한다. Freeman(1979)이 정리한 사회 네트워크 분석 분야의 중심성

척도는 연결 여부만 구분하는 이진 네트워크에 적용하기 위한 것이고, 링크에 가중치가 부여된 가중 네트워크에 대해서는 제대로 적용되지 못하기 때문이다. 이 연구에서는 기존 중심성 척도를 참고하여 계량서지적 네트워크에서 중심성을 분석하기 위한 네 가지 새로운 척도를 제안하였다. 제안한 척도를 시험적으로 저자동시인용 자료에 적용해본 다음, 성격이 다른 용어 동시출현 자료와 웹 사이트 동시링크 자료를 대상으로 중심성 분석을 수행하여 제안된 척도의 유용성을 검증하였다.

2. 중심성 척도

2.1 사회 네트워크 분석과 중심성 척도

사회 네트워크 분석에서는 그래프 이론에 따라서 분석 대상을 노드(node, vertex, 결절점, 결점, 정점, 꼭지점, 점 등으로 표현함)와 링크(link, edge, 연결, 연결선, 변, 간선 등으로 표현함)로 파악한다. 특정 노드의 중심성은 국지적인 시각에서 보는 경우와 전체적인 시각에서 보는 경우로 나눌 수 있다. 특정 노드와 인접한 노드들과의 관계만으로 살펴보는 경우는 지역 중심성(local centrality), 분석 대상 전체 노드들과의 관계를 살펴보는 경우는 전역 중심성(global centrality)이라고 한다(Scott 2000).

한 노드의 중심성을 측정하는 척도로는 연결정도중심성, 근접중심성, 매개중심성의 세 가지가 대표적이다(Freeman 1979). 연결정도중심성(degree centrality: C_D 로 단축 표기)은 한 노드가 연결되어 있는 링크의 수로 측정되며 지

역 중심성을 측정한다. 근접중심성(closeness centrality; C_C 로 단축 표기)은 한 노드가 다른 노드들과 얼마나 평균적으로 가까이 있는지를 측정하는 전역 중심성 척도이다. 여기에서의 거리란 링크의 수로 표현되는 경로거리를 뜻한다. 매개중심성(betweenness centrality; C_B 로 단축 표기)은 Freeman(1979)이 제안한 것으로서 한 노드가 다른 노드들 사이의 최단경로 상에 위치하는 정도를 측정하는 전역 중심성 척도이다. 상이한 집단간을 연결하는 노드일수록 매개중심성이 높게 나타난다.

앞에서 살펴본 중심성 척도 세 가지는 모두 노드간의 관계가 연결 유무로만 표현된 이진 네트워크(그래프)에 적용되는 것이다. 동시인용 네트워크와 같은 가중 네트워크에서는 연결의 강도가 가중치로 표현되므로 세 가지 척도를 그대로 적용하기가 어렵다(Liu et al. 2005). 예를 들어 연결정도중심성 척도의 경우에는 '인접' 여부의 판정 기준이 모호하다.

사회 네트워크 분석에서 사용되는 중심성 척도가 가중 네트워크를 고려하지 않고 있는 것은 그것이 계산을 간단하게 만들기 때문이기도 하지만, 일반적으로 사회 네트워크는 연결 강도를 계량화하기가 어려운 것이 더 주된 이유라고 할 수 있다. 예를 들어 친구들의 네트워크라면 어느 친구와 더 친하고 덜 친한 정도를 수치로 나타내는 것보다는 단순히 친구나 아니냐의 여부만 조사하는 것이 현실적이다. 그러나 계량서지적 분석의 대상이 되는 계량서지적 네트워크는 동시인용빈도처럼 자료의 수집 단계에서부터 빈도 자료로 수집되는 경우가 대부분이므로 연결 강도는 자연스럽게 수치로 표현된다. 따라서 계량서지적 네트워크에 대해서 전통적인 중심

성 척도를 사용하여 각 노드의 중심성을 파악하기 위해서는 특정한 방법으로 가중 네트워크를 이진 네트워크로 변환하여야 한다.

가중 네트워크를 이진 네트워크로 변경시키는 방법은 여러 가지가 제안되었지만(이재윤 2006; Newman 2004b), 가중치를 0 또는 1로 변환하는 과정에서 어느 정도의 정보가 상실되는 것은 피할 수 없다. Newman(2004b)은 가중 네트워크를 가중치만큼의 이진 연결이 중복된 다중 그래프로 변환하는 방식을 제안하였다. 이 방식에 따르면 기존의 척도는 대부분 변환된 다중 그래프에 적용이 가능하지만, 적용 결과는 본래 의미와 달라지는 경우가 많다. 예를 들어 연결정도중심성을 변환된 다중 그래프에 적용하면 특정 노드가 다른 노드와 연결된 링크의 가중치를 합하는 결과가 된다. 이로 인해 인접한 노드만이 아닌 매우 낮은 가중치로 연결된 원거리 노드까지 고려하게 되므로 결과적으로 지역 중심성이 아닌 전역 중심성을 반영하는 셈이 된다. 더군다나 가중치가 단순 빈도인 정수 값이 아니라 실수 값인 경우에는 다중 그래프로 변환하기 어렵다는 문제도 있다. 그럼에도 불구하고 제한된 상황 하에서는 적용이 가능한 방식이므로 이를 학술지 인용 네트워크에 적용하여 학술지 영향력 측정을 시도한 연구(Bollen et al. 2005)가 최근에 발표되었다.

2.2 계량서지적 네트워크를 위한 중심성 척도

지역 중심성과 전역 중심성을 다양한 가중 네트워크에 대해서 측정하기 위해서는 이에 어울리는 대안 척도를 사용하는 것이 더 바람직하다. Liu 등(2005)은 공저자 네트워크 분석을

위해서 PageRank 알고리즘(Page et al. 1999)을 수정한 AuthorRank 알고리즘을 제안하였다. 이 방식은 방향성이 있는 네트워크에서 각 저자의 중요도를 연결해나간 다른 저자로 반영시키는 것이므로 중심성이라기 보다는 위세(prestige)를 측정하는 연구(Bonacichi 1972; Knoke & Burt 1983)이다.

가중 네트워크에서의 지역·전역 중심성 측정을 위한 대안으로 기존 척도를 참고하여 이 연구에서는 네 가지 새로운 척도를 제안한다.

① 최근접중심성 : 다른 노드에 의해 최근접 이웃으로 꼽히는 정도를 최근접중심성(nearest centrality; C_N 으로 단축 표기)으로 정의하고 가중치 네트워크에서의 지역 중심성을 나타내는 지표 중 하나로 삼기로 한다. 이 최근접중심성은 가중 네트워크를 최근접이웃 그래프(이재윤 2006)로 변환한 다음 내향(indegree) 연결정도중심성을 산출하는 것과 같다. 최근접중심성은 해당 노드를 최근접 이웃으로 꼽을 만큼 최대한 인접한 노드만을 대상으로 분석하므로 가중 네트워크에서 연결정도중심성의 대안이 될 수 있다.

② 평균연관성 : 다른 노드들과의 연관성 값의 평균을 산출하면 이 노드가 다른 노드들과 전체적으로 얼마나 근접한 위치에 있는가를 측정하는 것이므로 전역 중심성을 나타내는 지표가 될 수 있다. 이를 평균연관성(mean associations; C_M 으로 단축 표기)으로 명명하고 가중 네트워크에서 근접중심성의 대안 척도로 제안한다. 다만 링크 가중치가 0인 경우, 즉 링크가 없는 노드 쌍이 지나치게 많은 경우에는 국지적인 척도가 될 위험이 있다. 이 척도는 Newman(2004b)의 방식처럼 가중 네트워크를 다중 그래프로 변환

한 후 연결정도중심성을 산출하는 것과 결과가 같다. 또한 이와 유사한 방식으로 Callon 등(1991)과 Ding 등(2001)은 동시출현단어분석에서 군집 내 노드가 군집 밖의 노드와 가지는 링크의 가중치를 평균하여 군집의 중심성이라고 한 바가 있다.

③ 평균프로파일연관성 : 전체적인 관점에서 가운데에 위치하더라도 다른 노드들과의 연관성 수준 자체가 낮은 노드는 평균연관성이 낮게 나타날 수 있다. 이런 경우에 연관성 값의 수준보다는 다른 영역 연구자들과의 연관성이 높고 낮은 패턴을 이용해서 전역 중심성을 측정할 수 있다. 동일한 이유로 저자동시인용분석을 처음으로 제안한 White와 Griffith(1981)도 동시인용빈도행렬에서 각 저자의 동시인용 프로파일 벡터간 상관계수를 2차적으로 산출한 바 있다. 이에 착안하여 다른 저자들과의 프로파일 연관성을 평균한 평균프로파일연관성(mean profile associations; C_{MP} 로 단축 표기)을 전역 중심성 척도로 제안한다. 중심이 아닌 주변 영역에 속한 노드일수록 다른 노드들과의 프로파일연관성이 낮게 나타날 것이기 때문이다. 평균프로파일연관성을 측정하기 위해서는 프로파일벡터간의 유사성을 피어슨상관계수로 측정하면 된다. 이는 White와 Griffith(1981)가 사용한 방법처럼 동시인용 또는 동시출현 행렬로부터 상관계수 행렬을 산출한 다음 이 상관계수 행렬에 대해서 평균연관성을 구하는 결과가 된다.

④ 삼각매개중심성: 한 노드가 다른 노드들 사이를 결속시켜주는 능력을 측정하는 척도로 삼각매개중심성(triangle betweenness centrality; C_{TB} 로 표기)을 제안한다. 측정은 노드 셋으로

구성된 삼각형을 기본 분석 단위로 한다. 연관성 행렬에서 분석 기준이 되는 노드를 a라고 할 때 a를 제외한 다른 노드들의 조합쌍 중에서 다음 조건을 만족하는 노드 쌍(b, c)의 비율로 측정한다.

$$sim(b, c) < MIN(sim(a, b), sim(a, c))$$

만약 유사도가 아닌 거리값을 이용할 경우에는 다음 조건을 만족하는 노드 쌍의 비율로 측정한다.

$$dist(b, c) > MAX(dist(a, b), dist(a, c))$$

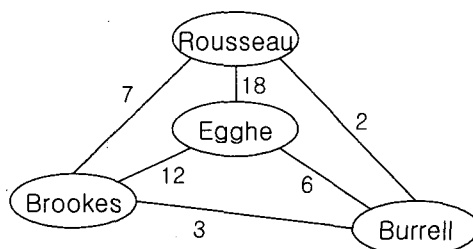
즉 서로 같지 않은 a, b, c가 삼각형의 꼭지점이라고 하면 a를 기준으로 보았을 때 a가 아닌 다른 모든 노드들의 조합으로 b와 c를 구성한 횟수 중에서 변 bc가 세 변 중 가장 긴 변이 되는 경우의 비율에 해당한다.

인간 관계에서도 서먹서먹한 두 사람 사이에 두 사람 모두와 친한 제 삼자가 개입하면 관계가 개선되기 마련이다. 삼각매개중심성은 이와 같이 행위자가 타인 간의 관계 개선에 기여하는 정도, 즉 중재력(仲裁力) 또는 결속력을 측정하는 것이라고도 할 수 있다. 중재자가 중재 대상인

두 사람 각자와 그리 친하지 않더라도 전혀 관계가 없는 두 사람 사이를 이어주는 역할은 할 수 있는 것이다. 이는 Granovetter(1973)가 주창했던 “약한 연결의 힘(strength of weak tie)”과도 관련이 있다. 결론적으로 삼각매개중심성은 소수와 집중된 관계를 맺는 노드보다는, 강도가 약하더라도 폭넓은 관계를 가진 노드에게 유리한 척도이다.

동시인용 네트워크가 <그림 1>과 같을 때 Egghe는 Burrell과 6회, Brookes와 12회 동시인용되었으므로 동시인용빈도 3인 Burrell과 Brookes 사이를 결속시켜주는 중재자 역할을 한다. 반면에 Rousseau는 Brookes와의 동시인용빈도는 7회이지만 Burrell과의 동시인용빈도가 2회에 불과하므로 3회 동시인용된 Burrell과 Brookes 사이를 결속시켜주지 못한다. 이 네 저자로 구성된 네트워크에 대해서 제안된 척도로 중심성을 측정한 결과는 <표 1>과 같다.

삼각매개중심성은 필요에 따라서 측정 범위를 전체 네트워크가 아닌 일부분으로 제한할 수 있다. 예를 들어 측정 대상과 근접한 5개 노드만을 대상으로 하거나(C_{TB-5} 로 표기), 측정 대상이 속한 군집을 범위로 하여(C_{TB-C} 로 표기) 군집을 결속시키는 능력을 파악하는 것이 가능하다. 이는 삼각매개중심성이 지역 중심



<그림 1> 동시인용 네트워크의 예

〈표 1〉 가중 네트워크에 대한 중심성 측정 예

저자	C_N		C_{TB}	C_M	C_{MP}	최근접 이웃
	절대	상대*				
Egghe	3	1.000	1.000	12.0	0.354	Rousseau
Rousseau	1	0.333	0.000	9.0	0.241	Egghe
Brookes	0	0.000	0.333	7.3	0.104	Egghe
Burrell	0	0.000	0.000	3.7	-0.397	Egghe

* 상대 척도는 이론상 가능한 최댓값으로 나눈 경우임

성이나 전역 중심성을 측정하는데 모두 사용될 수 있음을 뜻한다.

여기서 제안한 네 가지 척도는 모두 분석대상 네트워크의 크기나 밀도에 의해 값의 크기가 상당히 영향을 받는다. 따라서 비교하기 위해서는 각 척도값을 이론상 가능한 최댓값으로 나누어서 정규화한 상대적 척도를 사용하는 것이 바람직하다. 전체 노드의 수를 N 이라고 하였을 때, 최근접 이웃으로 꼽히는 최대 횟수는 $N-1$ 이므로 상대적 최근접중심성(C_N)은 최근접중심성을 $N-1$ 로 나눈 값이며 범위는 0에서 1사이이다. 평균연관성은 연관성을 자카드나 코사인 계수 등으로 정규화해서 측정하면 자체적으로 이미 정규화된 것이므로 별도의 조치가 0에서 1 사이로 범위가 제한된다. 만약 빈도 자체로 연관성을 표현할 경우에는 빈도 행렬의 최댓값으로 나누는 것을 생각해볼 수 있다. 평균프로파일연관성은 피어슨상관계수의 평균을 구하는 것이므로 역시 별도의 조치가 필요 없으며 -1에서 1 사이의 범위를 가진다. 삼각매개 중심성은 이미 최댓값 $N(N-1)/2$ 로 나누어진 비율이므로 별도의 정규화 조치가 없어도 0에서 1 사이로 제한된다.

3. 저자동시인용 네트워크의 중심성 분석

제안한 중심성 척도의 특성을 파악하고 효율성을 검증하기 위해서는 실제 계량서지적 자료에 대해 적용하여 중심성 분석을 수행할 필요가 있다. 이 장에서는 계량서지적 지적 구조 분석 연구에서 가장 많이 다루어지는 저자동시인용 자료에 대해서 중심성 분석을 수행하였다. 저자동시인용빈도 자료는 기존 연구의 것을 입수하여 이용하되 자료에 대한 분석은 이 연구에서 독자적으로 수행하였다.

3.1 저자동시인용 데이터

국내에서 수행된 저자동시인용 분석 연구 중에서 최근에 수행되었으며 자료의 규모와 범위가 중심성 분석이 필요할 만큼 방대한 이은숙과 정영미(2002)의 정보학 분야 저자동시인용 자료를 분석 대상으로 하였다. 이 자료는 1990년부터 2000년까지 11년간 *Journal of the American Society for Information Science*에 인용된 저자들 중에서 제1저자 인용빈도 순으로 〈표 2〉와 같은 상위 50인을 선정하여 동시인용 빈도를 산출한 것이다.

〈표 2〉 분석대상 정보학분야 핵심 저자

연번	저자명	연번	저자명	연번	저자명	연번	저자명	연번	저자명
1	Bates MJ	11	Chen H	21	Harter SP	31	Maron ME	41	Shannon CE
2	Belkin NJ	12	Cooper WS	22	Ingwersen P	32	McCain KW	42	Shaw WM
3	Blair DC	13	Croft WB	23	Kantor PB	33	Meadow CT	43	Small H
4	Bookstein A	14	Cronin B	24	Kling R	34	Narin F	44	Sparck Jones K
5	Borgman CL	15	Dervin B	25	Kuhlthau CC	35	Price DJD	45	Spink A
6	Brookes BC	16	Egghe L	26	Lancaster FW	36	Robertson SE	46	Swanson DR
7	Brooks TA	17	Fidel R	27	Larson RR	37	Rousseau R	47	van Rijsbergen CJ
8	Buckland MK	18	Fox EA	28	Losee RM	38	Salton G	48	White HD
9	Burrell QL	19	Garfield E	29	Marchionini G	39	Saracevic T	49	Wilson P
10	Bush V	20	Harman D	30	Markey K	40	Schamber L	50	Wong SKM

* 출처: 이은숙, 정영미(2002)

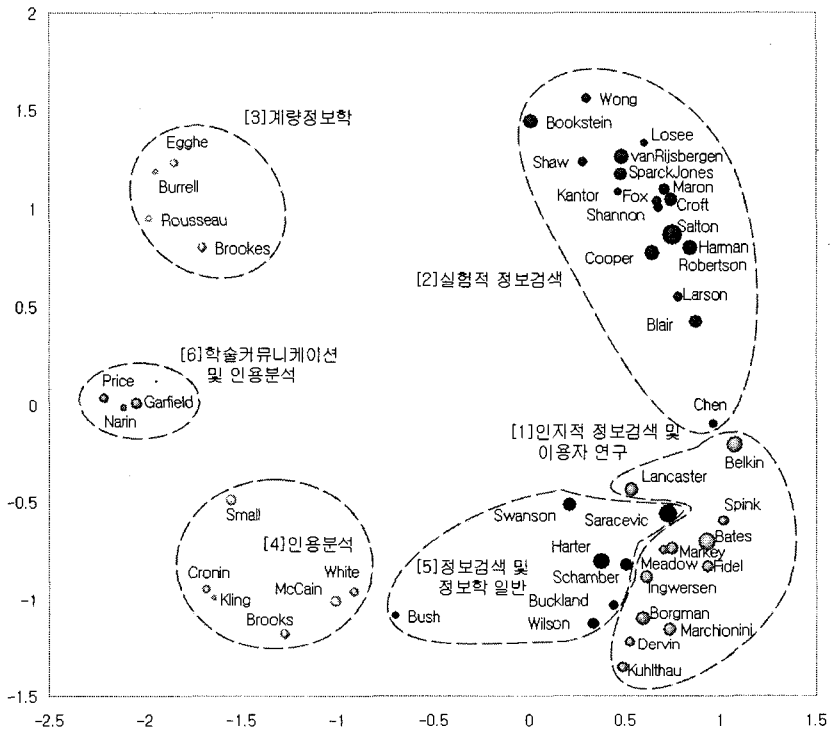
자료 처리에 있어서 이은숙과 정영미(2002)의 연구와 이 연구가 다른 점은 동시인용빈도 행렬의 대각선값 책정 방식이다. 이 연구에서는 다른 저자와의 동시인용빈도의 평균과 최고값을 구하여 최고값에 평균을 더한 값을 대각선값, 즉 자신 동시인용빈도로 책정하였다. 이는 흔히 사용되는 방식인 최댓값 세 개의 합을 2로 나눈 것과 큰 차이가 없으면서도 자신과의 동시인용빈도가 다른 저자와의 값보다 크도록 보장하기 위해서이다. 대각선 값을 이와 같이 처리하여 자기 자신과의 동시인용빈도가 가장 크도록 만들면, 필요에 따라서 동시인용빈도를 자카드 계수나 코사인 계수 등을 적용하여 정규화할 수 있는 장점이 있다.

3.2 저자동시인용을 통한 지적구조의 시각화

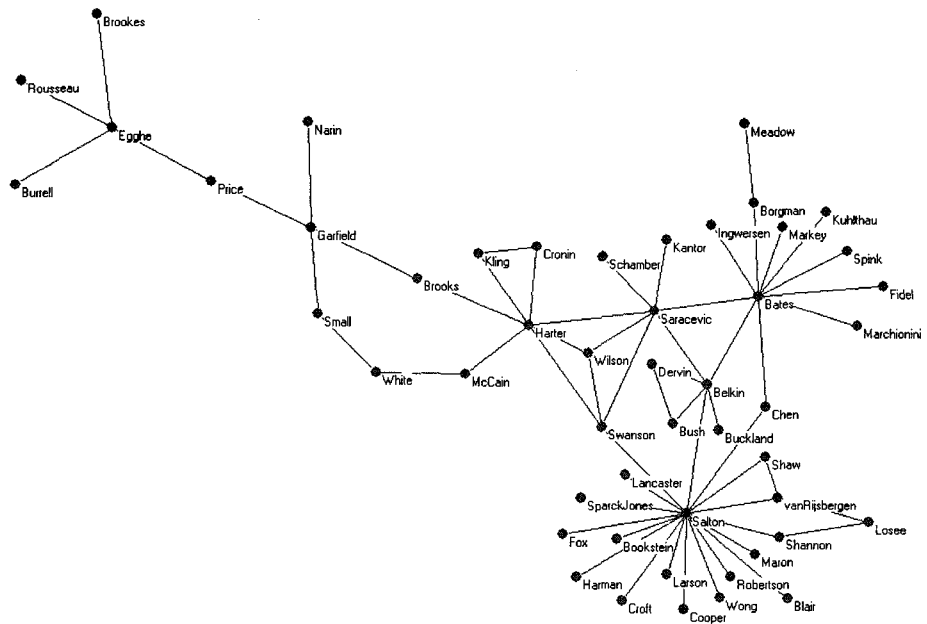
저자동시인용 자료에 대해서 중심성을 측정하기 전에 우선 두 가지 방법으로 전체적인 지적구조를 표현하였다. 중심성 분석 결과를 전통적인 시각적 표현 방식의 결과와 비교하기 위해서이다.

첫 번째 방법은 White와 Griffith(1981)의 전통적인 방식에 따른 다차원척도법과 군집분석에 의한 지적구조이다(그림 2 참조). 이 연구에서는 저자동시인용빈도 행렬에 대해서 저자간의 프로파일 상관계수를 구한 다음 이를 입력하여 MDS 지도를 표현하고 그 위에 Ward 기법으로 형성한 6개 군집을 표시하였다. 분석 처리는 SPSS v.12.0을 사용하여 수행하였다.

두 번째 방법은 Schvaneveldt(1990)의 패스파인더 네트워크(pathfinder network) 알고리즘으로 표현한 PFNet이다(그림 3 참조). PFNet에는 주요 링크를 결정하는 제약조건이 되는 파라미터가 있는데, 여기서는 다른 연구와 마찬가지로 가장 강력한 제약조건이 되어 핵심 흐름을 파악하는데 유리한 것으로 알려진 $q=n-1, r=\infty$ 로 설정하였다(이재운 2006; Chen 2005). PFNet 형성을 위한 프로그램은 Visual FoxPro로 직접 구현하였고, 시각화를 위해서 네트워크 분석용 공개 소프트웨어인 Pajek(de Nooy et al. 2005)을 사용하였다.



<그림 2> MDS 지도와 군집분석으로 표현한 정보학 분야의 지적 구조



<그림 3> PFNet($q=n-1, r=\infty$)으로 표현한 정보학 분야의 지적 구조

3.3 중심성 분석

앞 절에서 생성한 PFNet에 대해서 전통적인 중심성 척도 세 가지를 적용하였고, 원 자료인 저자동시인용 행렬을 대상으로 이 논문에서 제안한 중심성 척도를 적용하였다. 지면 관계상 각 척도별로 상위 10위까지에 해당하는 저자만 살펴보면 <표 3>과 같다. 척도값이 같을 경우에는 인용빈도가 높은 저자에게 우선 순위를 부여하였다.

국지적 관점의 중심성인 연결정도중심성(C_D)과 최근접중심성(C_N)은 둘 다 지역 중심성을 반영하므로 순위가 4위까지 일치하면서 비슷하게 나타났다.

거시적 관점에서 다른 저자들과의 거리를 고려한 근접중심성(C_C)과 평균프로파일연관성(C_{MP})은 다소 다르게 나타났다. 근접중심성 1위인 Salton이 평균프로파일연관성은 8위였으며, 평균프로파일연관성 1위인 Saracevic의 근접중심성은 4위였다. <그림 3>의 PFNet에서 Salton이 여타 실험적 정보검색 연구자들에 대한 연결을 독점한 결과가 근접중심성 1위로 나

타났다고 판단된다. Salton의 평균프로파일연관성 순위가 상대적으로 낮게 나타난 이유는 가중 네트워크로부터 이진 네트워크인 PFNet으로 변환하면서 링크로 표현되지 못하고 생략되었던 실험적 정보검색 연구자들과 타 분야 연구자들의 동시인용빈도까지 고려하기 때문이다. <그림 2>와 <그림 3>을 보면 군집5에 속한 Saracevic이 군집 2에 속한 Salton에 비해서 여러 다른 주제 분야와의 거리가 고른 것으로 여겨진다.

삼각매개중심성은 측정 범위에 따라서 순위가 상당히 다르게 나타났다. 전체(자신 제외 49명)가 대상인 경우(C_{TB})에는 Salton이나 Bates와 같은 큰 군집의 핵심 저자, 그리고 Saracevic이나 Harter와 같은 중앙 군집의 핵심 저자가 높은 순위를 차지했다. 각 저자마다 연관도 기준 주변 5명의 저자로 범위를 제한한 경우(C_{TB-5})에는 크기가 작은 군집 6이나 군집 3의 핵심저자인 Garfield와 Egghe가 5위권으로 진입했다. 측정 범위를 소속 군집으로 제한한 경우(C_{TB-C})에는 Egghe와 Garfield가 1, 2위를 차지해서 작은 군집에 유리한 방식이라는 것을 알 수 있다.

<표 3> 척도별 상위 10위 이내 저자 비교(괄호[] 안은 해당 저자의 소속 군집 번호(1~6))

순위	PFNet 중심성			가중 네트워크(동시인용행렬) 중심성					
	C_D	C_C	C_B	C_N	C_{TB-C}^*	C_{TB-5}^{**}	C_{TB}	C_M	C_{MP}
1	[2]Salton	[2]Salton	[2]Salton	[2]Salton	[3]Egghe	[2]Salton	[2]Salton	[2]Salton	[5]Saracevic
2	[1]Bates	[1]Belkin	[5]Harter	[1]Bates	[6]Garfield	[6]Garfield	[5]Saracevic	[1]Bates	[1]Belkin
3	[5]Harter	[5]Swanson	[1]Bates	[5]Harter	[2]Salton	[3]Egghe	[5]Harter	[5]Saracevic	[1]Lancaster
4	[5]Saracevic	[5]Saracevic	[5]Saracevic	[5]Saracevic	[1]Bates	[1]Bates	[1]Belkin	[5]Harter	[5]Harter
5	[1]Belkin	[5]Harter	[4]Brooks	[6]Garfield	[2]vanRijsbergen	[1]Belkin	[1]Belkin	[5]Harter	[5]Swanson
6	[5]Swanson	[1]Bates	[5]Swanson	[1]Belkin	[1]Belkin	[5]Saracevic	[1]Lancaster	[2]Robertson	[1]Bates
7	[6]Garfield	[5]Wilson	[6]Garfield	[3]Egghe	[5]Saracevic	[4]Small	[2]Cooper	[2]vanRijsbergen	[2]Cooper
8	[3]Egghe	[2]Chen	[1]Belkin	[2]vanRijsbergen	[2]Robertson	[6]Price	[1]Borgman	[2]Cooper	[2]Salton
9	[5]Wilson	[4]Brooks	[6]Price	[1]Borgman	[1]Borgman	[4]McCain	[2]Robertson	[1]Borgman	[2]Robertson
10	[2]vanRijsbergen	[2]vanRijsbergen	[3]Egghe	[4]Small	[4]Cronin	[2]vanRijsbergen	[2]Bookstein	[1]Lancaster	[2]Blair

* C_{TB-C} 는 6개 군집별로 범위를 제한하여 각 군집 내에서 삼각매개중심성을 측정된 경우임.
 ** C_{TB-5} 는 각 저자마다 동시인용빈도가 높은 5개 인접 노드에 대해서 삼각매개중심성을 측정된 경우임.

전체적인 관점의 삼각매개중심성, 평균연관성, 평균프로파일연관성 척도는 큰 군집인 군집 1, 군집 2, 군집 5에 속한 저자들만 10위 안에 포함된 반면에, 국지적인 관점의 최근접중심성, 군집내 삼각매개중심성, 근접 5노드 삼각매개중심성은 각 군집별로 최소 한 저자 이상씩을 10위 안에 포함시킨 점이 눈에 띈다. 이는 동시인용 네트워크에서 핵심 저자를 파악할 때 주요 주제 위주로 할 것인지, 각 주제별로 고르게 할 것인지에 따라서 채택하는 척도가 달라져야 함을 뜻한다. 특히 근접 5노드 삼각매개중심성은 7위 이내에 여섯 군집의 저자가 모두 하나 이상 나타나서 각 주제별로 가장 고르게 핵심 저자를 파악할 수 있었다.

3.4 측정 범위에 따른 삼각매개중심성의 표출 유형

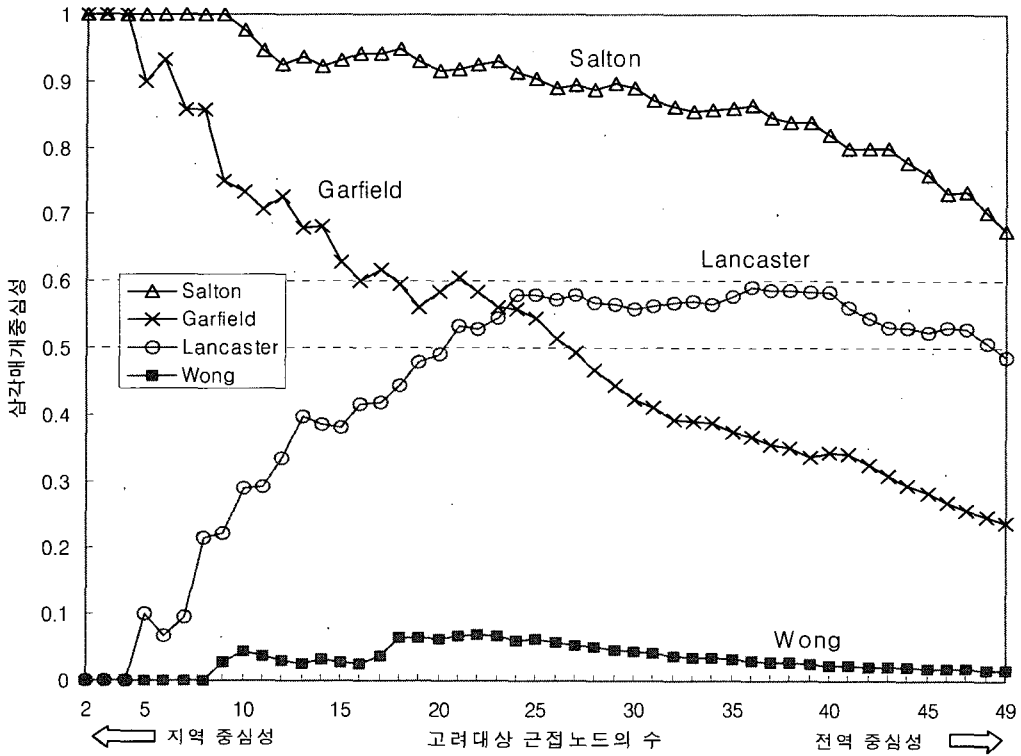
삼각매개중심성 척도는 앞에서 살펴본 바와 같이 측정 범위를 달리함에 따라서 국지적인 시각과 전체적인 시각을 모두 반영할 수 있는 척도이다. 이런 특징을 활용하여 여기서는 다양한 범위에서 각 저자의 중심성을 살펴보았다. 가장 국지적인 범위로는 연관도를 기준으로 인접 두 저자에서부터 시작하여 가장 전역적인 범위인 전체 저자에 이르기까지 삼각매개중심성의 측정 범위를 넓혀가며 각 단계마다의 척도값을 측정하였다. 시각적인 이해를 돕기 위해서 50명의 측정 결과를 모두 나타내지 않고 삼각매개중심성의 변화 양상 유형을 대표하는 저자 네 명의 측정 결과만을 <그림 4>에 제시하였다. 중심성의 표출 구역에 따라 파악된 네 유형은 다음과 같다.

① Salton 유형(지역 및 전역 중심): 특정 분야에 대한 영향력이 매우 강력하여 국지적으로 뿐만 아니라 전체적으로도 중심 역할을 차지하는 유형으로서 지역 및 전역 중심 노드(local and global central node)라고 부를 수 있다. 측정 범위가 넓어지더라도 삼각매개중심성이 크게 낮아지지 않고 완만하게 하락하는 경향을 보인다. 대개 큰 주제 영역의 중심 저자가 동시에 전역 중심성도 높게 되는데, Bates도 이 유형에 속한다.

② Garfield 유형(지역 중심, 전역 주변): 특정 분야에 대한 영향력만 강력하여 국지적으로 중심을 차지하는데 그치는 유형으로서 지역 중심 노드(local central node)라고 부를 수 있다. 측정 범위가 넓어질수록 삼각매개중심성이 급격히 하락하는 경향을 보인다. 작은 주제 영역의 중심 저자인 Egghe, Price, Small 등이 이에 속한다.

③ Lancaster 유형(지역 주변, 전역 중심): 특정 분야에 대한 영향력은 미미하지만 관계의 범위가 넓기 때문에 전체적인 차원에서는 결속력을 발휘하는 드문 유형으로서 전역 중심 노드(global central node)라고 부를 수 있다. 여타 중심 노드 유형의 경우에는 삼각매개중심성의 측정 범위가 주변 다섯 노드를 넘어서면 중심성이 하락하는데 반해서 이 유형은 오히려 상승하는 것이 특징이다. Cooper도 약간 이 유형에 가깝다.

④ Wong 유형(지역 및 전역 주변): 특정 분야로 관계의 폭이 제한되어 있으며 지역 중심 저자의 영향력 하에 있는 유형으로서 주변 노드(peripheral node)라고 부를 수 있다. 모든 범위에서 삼각매개중심성이 매우 낮게 나타나며 상당수 저자가 이 유형에 해당한다.



〈그림 4〉 측정 범위에 따른 삼각매개중심성의 변화 양상

4. 기타 계량서지적 네트워크의 중심성 분석

여기서는 앞에서 살펴본 계량서지적 중심성 척도를 사용하여 동시인용 자료와 성격이 다른 용어 동시출현 네트워크와 웹 사이트 동시링크 네트워크를 분석하는 사례를 제시하여 제안된 중심성 척도의 유용성을 검증하였다. 3장의 저자동시인용 분석에 적용한 사례와 마찬가지로 분석을 위한 원 자료는 기존 연구에서 취하고 네트워크 표현과 중심성 분석은 자체적으로 실시하였다.

4.1 용어 동시출현 네트워크 분석

용어 동시출현 네트워크의 사례로 유영준(2003)의 연구에서 수집된 문헌정보학 분야 3개 학회지 논문의 색인어 집합을 대상으로 중심성을 분석해보았다. 유영준(2003)은 2001년까지 정보관리학회지, 한국도서관정보학회지, 한국문헌정보학회지에 실린 1,162편의 국내 논문에 대해서 국회도서관의 국내 학술지 기사색인 데이터베이스에 색인된 색인어를 수집하여 분석한 바 있다. 여기서는 유영준(2003)의 연구에 포함된 색인어 중에서 사용 빈도가 10회 이상

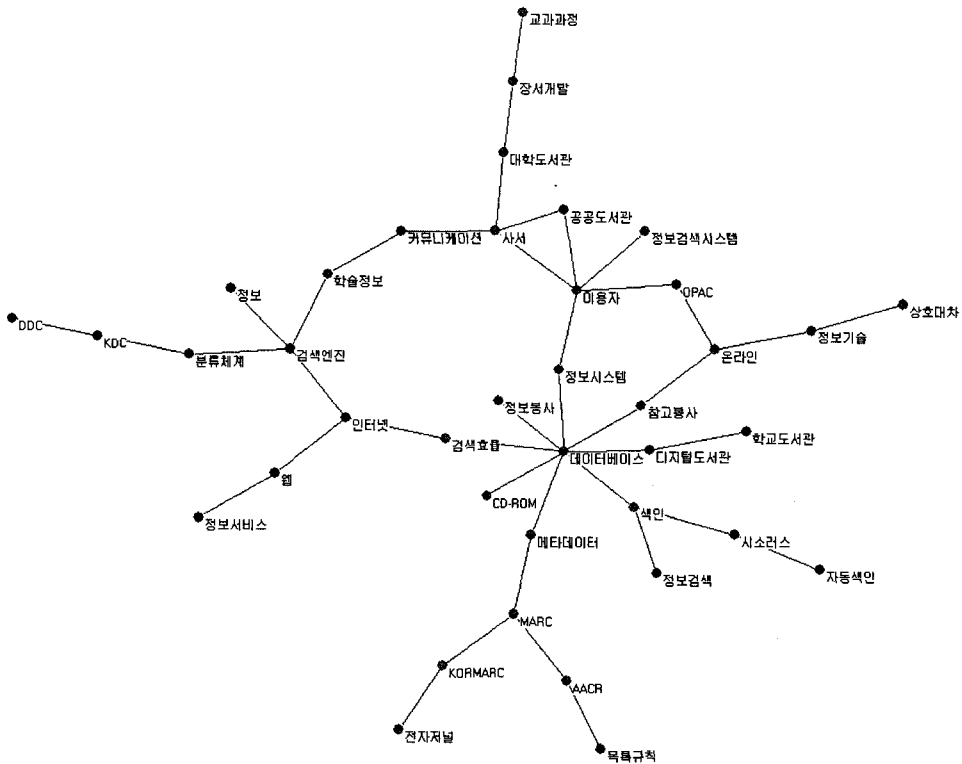
인 색인어 39개를 다음 <표 4>와 같이 선정하여 이들이 동일한 논문에 대해서 색인된 빈도로 용어 동시출현 네트워크를 구축하였다. 이 39개

용어에 대해서 동시출현 빈도를 자카드 계수로 정규화한 후 PFNet($q=n-1, r=\infty$)을 생성해보면 <그림 5>와 같다.

<표 4> 문헌정보학 분야 3개 학회지 논문의 사용 빈도 상위 색인어

빈도	색인어	빈도	색인어	빈도	색인어	빈도	색인어
105	대학도서관	25	학교도서관	14	자동색인	10	검색엔진
64	공공도서관	21	DDC	14	정보기술	10	전자저널
38	인터넷	19	웹	14	정보봉사	10	정보
32	정보검색	17	KORMARC	13	메타데이터	10	정보시스템
30	디지털도서관	16	분류체계	13	색인	10	커뮤니케이션
29	OPAC	16	시소러스	13	MARC	10	학술정보
27	정보서비스	16	정보검색시스템	12	교과과정	10	AACR
26	데이터베이스	15	검색효율	12	상호대차	10	CD-ROM
26	사서	15	장서개발	11	목록규칙	10	KDC
26	이용자	15	참고봉사	11	온라인		

* 출처: 유영준(2003)



<그림 5> PFNet($q=n-1, r=\infty$)으로 표현한 국내 문헌정보학 분야 용어 동시출현 네트워크(2001년 이전)

3개 학회지 수록 논문의 주요 색인어 39개에 대한 동시출현 행렬을 자카드 계수로 정규화한 후 산출한 각 용어의 중심성은 <표 5>와 같다. 자카드 계수로 정규화한 이유는 이 자료와 같이 동시출현 빈도 수준이 높지 않은 경우에는 빈도를 그대로 사용하면 동률이 너무 많이 나타나서 해석하기가 어렵고, 정규화를 통해서 값의 범위를 조절하는 효과를 얻을 수 있기 때문이다.

<표 5>에서 지역 중심성에 해당하는 두 척도 중에서 왼쪽의 최근접중심성(C_N)이 가장 국지적인 관점의 중심성을 측정하는 척도라고 할 수 있다. 오른쪽의 전역 중심성 척도 중에서 삼각매개중심성(C_{TB})은 해당 용어의 영향력이 미치는 범위를 반영한다고 볼 수 있고, 평균프로파일연관성(C_{MP})은 국내 문헌정보학 영역에서 해당 용어의 위치가 중립적인지 여부와 관련된다.

주요 용어의 중심성을 보면 국지적 관점의 중심성인 최근접중심성(C_N)과 근접 5노드 삼각매개중심성(C_{TB-5})은 '데이터베이스'가 1위로 나타나며 전역적 관점의 중심성인 삼각매개

중심성(C_{TB}), 평균프로파일연관성(C_{MP})으로는 모두 '인터넷'이 1위로 나타났다. 이를 통해 알 수 있는 것은, 세부 주제로는 문헌정보학에서 '데이터베이스'가 가장 영향력이 큰 용어이지만, 전반적으로 더 폭넓은 영향력을 가진 용어는 '인터넷'이라는 점이다. 1990년대 중반 이후 '인터넷'이 문헌정보학 전반에 끼친 막대한 영향이 반영된 결과로 판단된다.

특이한 용어는 전체 사용빈도 1위인 '대학도서관'이다. '대학도서관'은 최근접중심성(C_N)과 근접 5노드 삼각매개중심성(C_{TB-5})으로는 10위 밖이지만 삼각매개중심성(C_{TB})으로는 4위, 평균프로파일연관성(C_{MP})으로는 2위로 나타났다. 즉 지역 중심성은 그리 높지 않지만 전역 중심성은 매우 높은 용어임을 알 수 있다. 조심스럽게 추측해보면, '대학도서관'이 국내 문헌정보학계에서는 개별 연구 주제라기보다는 데이터베이스, 이용자, OPAC, 장서관발 등의 연구 주제를 탐구하는 배경 상황으로서 다루어지기 때문이라고 생각된다. 유영준(2003)은 색인어

<표 5> 문헌정보학 분야 용어의 중심성(척도별 상위 10위 이내)

지역 중심성				전역 중심성			
C_N		C_{TB-5}		C_{TB}		C_{MP}	
순위	용어	순위	용어	순위	용어	순위	용어
1	데이터베이스	1	데이터베이스	1	인터넷	1	인터넷
2	인터넷		디지털도서관	2	데이터베이스	2	대학도서관
	이용자	색인	3	이용자	3	이용자	
5	검색엔진	4*	인터넷	4	대학도서관	4	데이터베이스
	OPAC		이용자	5	공공도서관	5	검색엔진
	시소러스		공공도서관	6	정보검색	6	사서
	색인		정보검색시스템	7	디지털도서관	7	OPAC
	KORMARC		검색엔진		OPAC	8	웹
	장서관발		정보기술, 온라인	9	정보검색시스템	9	참고봉사
10	사서(외 15개)	정보봉사, AACR	10	정보서비스	10	정보검색	

* C_{TB-5} 기준으로는 공동 4위가 9개여서 상위 12개 용어를 제시하였음

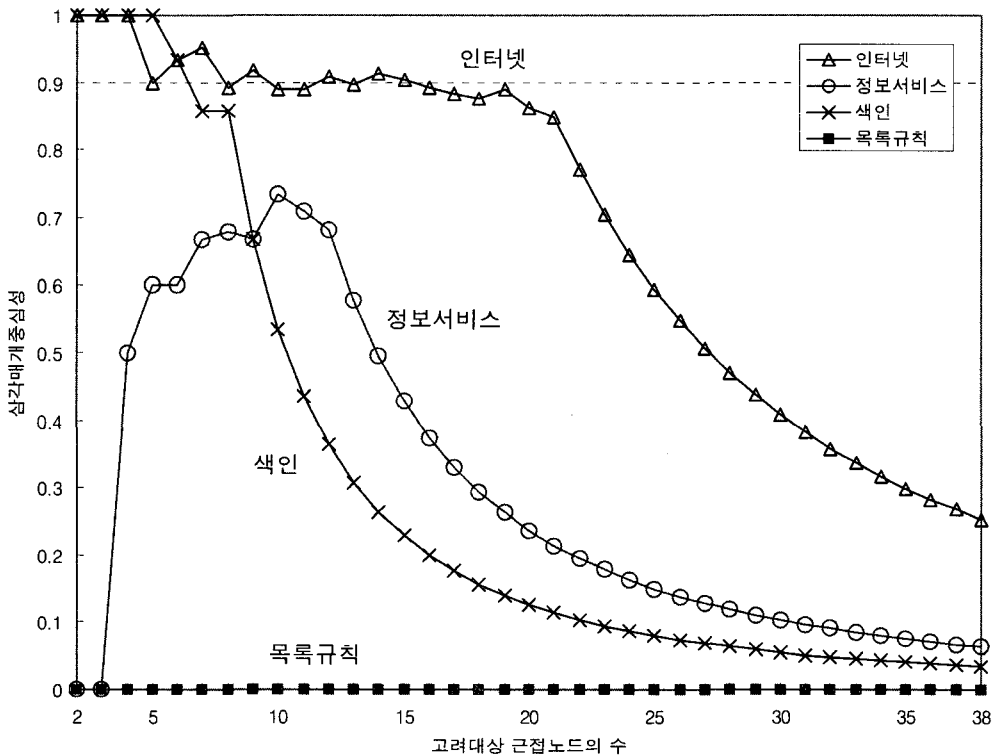
를 분석할 때 '도서관'은 거의 모든 연구에 해당되므로 제외하였는데, '대학도서관'도 어느 정도 이에 가깝다고 할 수 있다.

앞 장의 저자동시인용 분석에서의 마찬가지로 삼각매개중심성을 주변 두 용어에서부터 전체 용어에 이르기까지 측정 범위를 달리하면서 산출한 결과를 유형별로 대표적인 네 용어만 제시하면 <그림 6>과 같다. 이 그림에서 '인터넷'이 지역 및 전역 중심 유형, '색인'이 지역 중심 유형, '목록 규칙'이 주변 유형에 해당하고, '정보서비스'는 분명하지는 않으나 전역 중심 유형에 가깝다고 여겨진다. 삼각매개중심성이 가장 높은 '인터넷'의 경우에도 측정범위가 근

접 21개 노드를 넘어서면서 급격히 감소하는 것은 '인터넷'의 영향력이 그 이상 미치지 못하기 때문이다. 이는 앞 장에서 분석한 저자동시인용 자료에 나타난 저자간의 관계에 비해서 동시출현 산출 대상인 용어간의 관계가 밀접하지 못함을 뜻한다.

4.2 동시링크 네트워크 분석

웹 사이트 동시링크 네트워크의 사례로 정동열과 최윤미(1999)가 조사한 커뮤니케이션 분야 웹 사이트 32개에 대해서 중심성 분석을 수행해보았다. 정동열과 최윤미(1999)는 커뮤니



<그림 6> 측정 범위에 따른 문헌정보학분야 용어의 삼각매개중심성 변화 양상

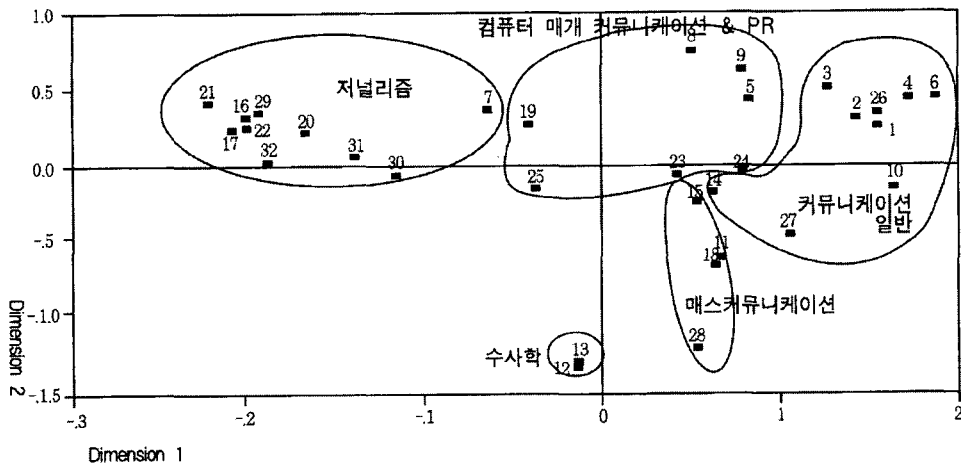
케이션 분야의 주요 웹 사이트 32개를 선정하여 각각에 <표 6>과 같은 주제를 할당하였다.

정동열과 최윤미(1999)는 선정된 사이트간의 동시링크 빈도를 검색엔진 Altavista를 이용해서 산출하고 이로부터 <그림 7>과 같은 다차원척도 분석과 군집 분석을 수행하였다. 동일한 동시링크빈도 행렬을 이용하여 본 연구에서 생성한 PFNet($q=n-1, r=\infty$)은 <그림 8>과 같다. 두 그림에서 '수사학'은 'speech communication'의 번역으로서 내용상 '화법'이나 '구두 커뮤니케이션'이 더 적절할 것으로 생각되나 비교 분석을

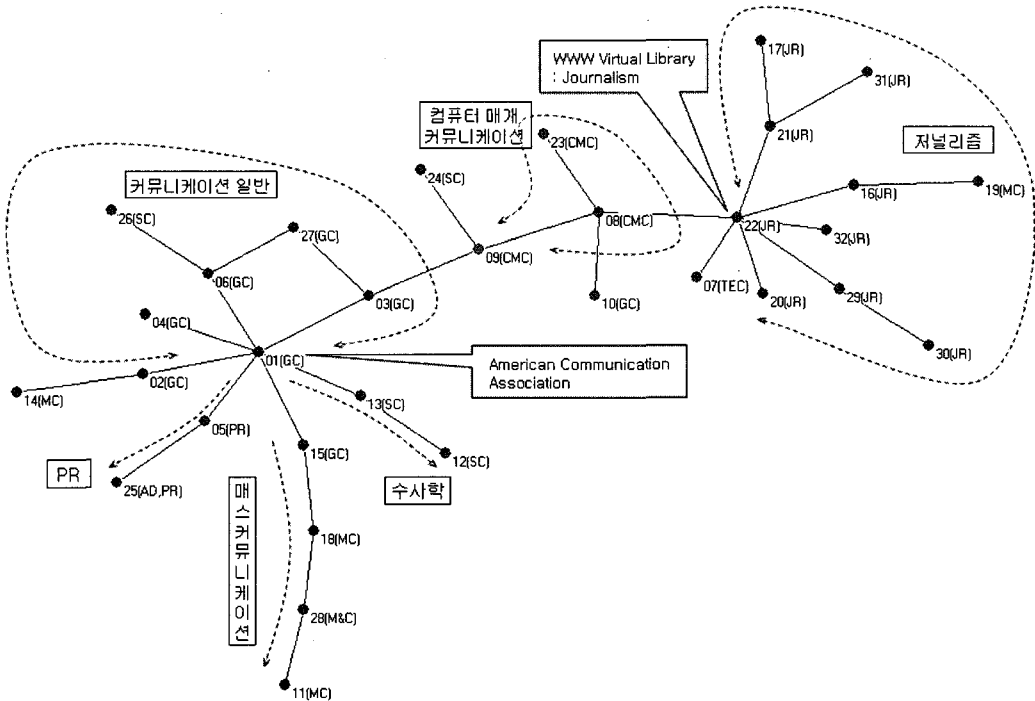
위해서 그대로 사용하였다. <그림 7>과 <그림 8>에서 공통적으로 확인할 수 있는 사항은 커뮤니케이션 일반 분야(GC)와 저널리즘 분야(JR)가 양대 핵심 분야를 이루고 있으며, 컴퓨터 매개 커뮤니케이션(CMC) 분야가 중립적인 위치에 해당한다는 점이다. <그림 8>의 PFNet에서 더 알 수 있는 것은 커뮤니케이션 일반 분야에서는 1번 사이트(American Communication Association)가 중심이고, 저널리즘 분야에서는 22번 사이트(WWW Virtual Library: Journalism)가 중심이라는 사실이다.

<표 6> 정동열과 최윤미(1999)가 다룬 웹 사이트 주제분야

약칭	주제명
[GC]	General Communication
[CMC]	Computer-mediated Communication
[SC]	Speech Communication
[TEC]	Telecommunication
[MC]	Mass Communication
[M&C]	Media and Culture
[JR]	Journalism
[AD]	Advertising
[PR]	Public Relations



<그림 7> 커뮤니케이션 분야 웹 사이트의 MDS 지도 및 군집(정동열, 최윤미 1999)



〈그림 8〉 커뮤니케이션 분야 웹 사이트의 PFNet($q=n-1, r=\infty$)과 주제 표시

커뮤니케이션 분야 웹 사이트 동시링크빈도 행렬에 대해서 2장에서 제시한 중심성 척도를 산출하여 각 척도별 상위 10개 이내 사이트를 파악한 결과는 <표 7>과 같다. 산출 결과 중에서 지역 중심성을 살펴보면 최근접중심성(C_N)과 근접 5노드 삼각매개중심성(C_{TB-5}) 모두 주요 주제인 저널리즘, 커뮤니케이션 일반, 컴퓨터매개커뮤니케이션 관련 사이트가 고르게 상위를 차지하고 있는 것으로 나타난다. 특히 근접 5노드 삼각매개중심성(C_{TB-5})으로는 6위 이내에 다섯 주제의 사이트가 하나 이상씩 자리하므로 여러 척도 중에서 가장 고르게 주요 노드를 파악해준다는 점이 다시 확인되었다. 두 가지 지역 중심성 척도에서 공통적으로 드러나는 것은 이미 PFNet에서 파악했듯이 커뮤니케

이션 일반 분야는 1번 사이트가 중심이고, 저널리즘 분야에서는 22번 사이트가 중심이라는 사실이다. 또한 PFNet에서는 파악하기 어려웠지만 중심성 척도로 알 수 있는 것은 컴퓨터매개 커뮤니케이션 분야에서는 9번과 8번 사이트가 거의 대등하게 중심을 차지하고 있다는 사실과, 매스커뮤니케이션 분야에서는 18번 사이트, PR 분야에서는 5번 사이트가 중심에 해당한다는 점이다.

<표 7>의 산출 결과 중에서 전역 중심성 척도를 살펴보면 이 동시링크 자료에 대해서는 평균프로파일연관성(C_{MP})과 삼각매개중심성(C_{TB})의 차이가 뚜렷하게 나타난다. 삼각매개중심성(C_{TB})으로는 여러 주제의 사이트가 고르게 상위권을 차지하지만, 평균프로파일연관

<표 7> 커뮤니케이션 분야 웹 사이트 중심성 상위 사이트

지역 중심성			전역 중심성		
C _N			C _{MP}		
순위	번호 (주제)	사이트 명	순위	번호 (주제)	사이트 명
1	22(JR)	WWW Virtual Library: Journalism	1	03(GC)	Communication Institute for Online Scholarship
2	01(GC)	American Communication Association	2	01(GC)	American Communication Association
3	09(CMC)	Journal of Computer-Mediated Communication	3	04(GC)	International Communication Association
	08(CMC)	Computer Mediated Communication Magazines	4	06(GC)	Southern States Communication Association
5	21(JR)	Reporter.org	5	02(GC)	Association for Educational Communication and Technology
	18(MC)	Media and Communication Studies Site	6	05(PR)	The Public Relations Society of America
7	04(GC)	International Communication Association 외 11개	7	08(CMC)	Computer Mediated Communication Magazines
			8	26(SC)	European Speech Communication Association
			9	09(CMC)	Journal of Computer-Mediated Communication
			10	21(JR)	Reporter.org
C _{TB-5}			C _{TB}		
순위	번호 (주제)	사이트 명	순위	번호 (주제)	사이트 명
1	01(GC)	American Communication Association	1	01(GC)	American Communication Association
2	22(JR)	WWW Virtual Library: Journalism	2	09(CMC)	Journal of Computer-Mediated Communication
	09(CMC)	Journal of Computer-Mediated Communication		3	22(JR)
4	08(CMC)	Computer Mediated Communication Magazines	4	08(CMC)	Computer Mediated Communication Magazines
	05(PR)	The Public Relations Society of America	5	15(GC)	Communication & Education on the Web
6	18(MC)	Media and Communication Studies Site	6	21(JR)	Reporter.org
7	21(JR)	Reporter.org	7	05(PR)	The Public Relations Society of America
	06(GC)	Southern States Communication Association	8	03(GC)	Communication Institute for Online Scholarship
	04(GC)	International Communication Association	9	06(GC)	Southern States Communication Association
	11(MC)	Vanderbilt Television News Archive	10	17(JR)	FACSNET

성(C_{MP}) 면에서는 커뮤니케이션 일반에 속하는 사이트가 상위 5위까지를 독점하고 있다. 이는 삼각매개중심성 척도가 영향력의 넓이를 반영하는데 반해서, 평균프로파일연관성 척도는 구조적으로 중심에 위치하는가를 반영하기 때문이라고 할 수 있다. 구조적으로는 커뮤니케이션 일반에 해당하는 사이트가 커뮤니케이션 분야의 중앙에 해당하는 것이 당연하다. 즉, 삼각매개중심성이 영향력 면에서 대표적인 노드

를 찾는데 도움이 된다고 한다면, 평균프로파일연관성은 분석 범위 안에서 중립적 혹은 중간적인 위치의 노드를 찾는데 도움이 되는 척도라고 할 수 있다.

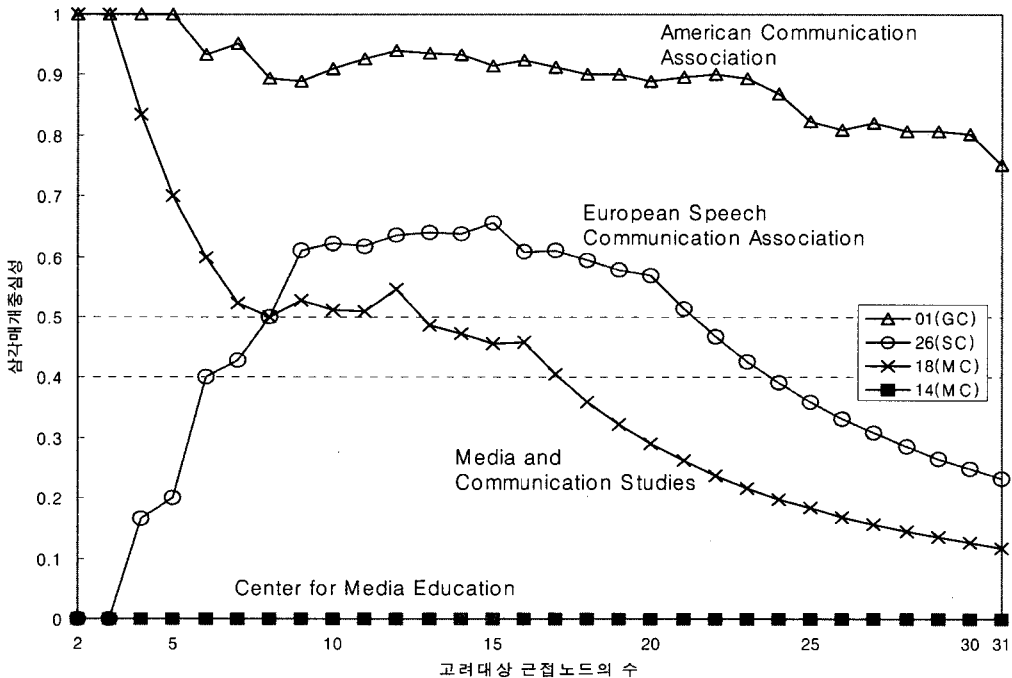
1번 사이트인 American Communication Association은 두 가지 지역 중심성 척도(C_N, C_{TB-5})와 삼각매개중심성(C_{MP})으로는 모두 1위를 차지하고 있으며 평균프로파일연관성(C_{MP})으로도 2위에 해당한다. 이는 앞 장의 저자동시인

용 분석에서 지역 및 전역 중심성 유형에 해당하는 Salton보다도 더 중심성이 높은 경우로서 (Salton은 다른 척도로는 1위이지만 평균프로파일연관성 척도로는 8위였음) 이 사이트가 커뮤니케이션 분야를 명실상부하게 대표하는 곳임을 알려준다.

한편 저널리즘 분야 사이트는 평균프로파일 연관성(C_{MP}) 면에서 전혀 상위권을 차지하지 못하는 것으로 나타났다. 이는 저널리즘 분야의 비중이 비록 상당히 크긴 하지만 커뮤니케이션 분야 전체로 볼 때에는 주변 영역 혹은 독자적인 영역에 해당하는 것으로 간주해야 함을 시사한다.

저자동시이용 및 용어동시출현에서와 마찬가지로 삼각매개중심성을 주변 두 사이트에서

부터 전체 사이트에 이르기까지 측정 범위를 달리하면서 측정할 결과를 유형별로 대표적인 네 사이트만 제시하면 <그림 9>와 같다. 이 그림에서도 1번 사이트인 American Communication Association이 지역 및 전역 중심 유형임이 분명히 드러난다. 18번 사이트인 Media and Communication Studies는 지역 중심성은 높으나 전역 중심성은 낮은 지역 중심 유형으로서, 매스커뮤니케이션 분야의 중심 노드이지만 전체적으로는 중심을 차지하지 못함을 나타낸다. 26번 사이트인 European Speech Communication Association은 측정 범위가 주변 다섯 사이트를 넘어가면서 오히려 중심성이 높아지는 것으로 나타나므로 지역 주변, 전역 중심 유형에 가깝다고 판단된다. 이 사이트는 <그림 7>의 군집



<그림 9> 측정 범위에 따른 커뮤니케이션 분야 웹 사이트의 삼각매개중심성 변화 양상

분석 결과에서도 수사학 군집에 포함되지 못하여서 전역 중심성은 높은 편이지만 수사학이라는 소주제의 중심 사이트로는 적합하지 않음을 알 수 있다. 14번 사이트인 Center for Media Education은 모든 범위에서 삼각매개중심성이 0으로 나타나서 전혀 중심성을 가지지 못하는 주변노드임을 알 수 있다. 측정 범위가 넓어짐에 따른 삼각매개중심성의 하락 경향을 살펴보면, 앞 절의 용어 동시출현 분석에서와 달리 웹 사이트 동시링크 분석에서는 가장 높은 1번 사이트의 중심성이 마지막까지 급격히 하락하지 않는 것으로 나타났다. 이는 분석대상 사이트들 간의 관계가 앞 절의 용어간 관계에 비해서 밀접하다는 것을 시사한다.

5. 결 론

동시인용이나 동시링크와 같이 계량서지적 분석의 대상이 되는 자료는 대부분 빈도행렬로 표현되며 이를 네트워크로 나타내면 연결 가중치가 있는 가중 네트워크가 된다. 이런 계량서지적 네트워크에 대한 중심성 분석을 효과적으로 수행하기 위해서 이 연구에서는 새로운 중심성 척도를 제안하였다. 제안된 척도 중에서 최근접중심성(C_N)은 지역 중심성을 측정하는 척도이고, 평균연관성(C_M), 평균프로파일연관성(C_{MP}), 삼각매개중심성(C_{TB})은 전역 중심성을 측정하는 척도이다. 지역 중심성을 측정하기 위한 용도로는 최근접중심성 이외에도 삼각매개중심성의 측정 범위를 주변 5개 노드로 제한한 근접 5노드 삼각매개중심성(C_{TB-5})을 사용할 수도 있다. 그리고 전역 중심성을 측정하는

척도중에서는 평균연관성보다 상관계수를 활용한 평균프로파일연관성이 구조적인 중심을 더 잘 찾아주는 것으로 나타났다. 따라서 지역 중심성을 측정할 때에는 최근접중심성과 근접 5노드 삼각매개중심성을 적용하고, 전역 중심성을 측정할 때에는 삼각매개중심성과 평균프로파일연관성을 적용하면 효율적인 분석이 가능할 것으로 판단된다. 필요에 따라서는 특정 군집내 노드로만 측정 범위를 제한하는 군집내 삼각매개중심성(C_{TB-C})도 유용할 것이다.

제안된 척도를 검증하기 위해서 저자동시인용 네트워크 용어 동시출현 네트워크 웹 사이트 동시링크 네트워크와 같은 다양한 계량서지적 자료에 적용해서 중심성을 분석해보았다. 분석 결과는 다차원척도법이나 군집분석, 패스파인더 네트워크와 같은 시각적인 분석 기법에 비해서 각 노드의 전체적인 입지와 국지적인 입지를 모두 더 분명하게 파악할 수 있는 것으로 나타났다. 그리고 전역 중심성 척도 중에서도 삼각매개중심성 척도는 영향력의 넓이를 반영하는데 반해서, 평균프로파일연관성 척도는 구조적으로 중심에 위치하는 정도를 반영하는 것으로 파악되었다. 따라서 영향력 면에서 대표적인 노드를 찾기 위해서는 삼각매개중심성 척도를 활용하고, 중립적 혹은 중간적인 위치의 노드를 찾기 위해서는 평균프로파일연관성 척도를 활용하는 것이 바람직하다. 또한 근접 5노드 삼각매개중심성 척도는 여러 소주제를 망라하여 고르게 대표 노드를 파악할 때 유리한 것으로 나타났다.

근접 5노드 삼각매개중심성이나 군집내 삼각매개중심성과 같이 삼각매개중심성 척도는 측정 범위를 임의대로 정하여 파악할 수 있는

용통성을 가지고 있다. 삼각매개중심성의 측정 범위를 최소에서 최대까지 달리하면서 중심성을 분석하면 각 노드를 ① 지역 및 전역 중심 유형, ② 지역 중심 유형, ③ 전역 중심 유형, ④ 주변 유형의 네 가지로 구분할 수가 있었다. 이는 기존의 지적 구조 분석 방식으로는 파악하기 어려운 정보이며, 단순히 각 노드가 중심성이 높거나 낮다는 평면적인 분석으로부터 한 걸음 더 나아가서 어느 정도 입체적인 분석을 가능하게 한다고 판단된다.

이 연구에서 제안한 중심성 척도의 유용성이 세 종류의 실제 자료에 대한 분석을 통해 입증된 만큼, 향후 계량서지적 분석 연구의 활성화

에 도움이 될 것으로 기대된다. 후속 연구로는 여러 분야의 실제 자료에 대해서 시각적 표현과 중심성 척도를 함께 적용한 분석을 진행할 계획이며, 특히 학술지 인용망이나 특허 인용망과 같이 분석결과와 실용적인 활용 가능성이 높은 영역에 대해서 적극적인 활용을 모색하고자 한다.

〈알림〉

이 연구에서 분석한 계량서지적 자료의 이용을 흔쾌히 허락해주신 정영미 교수님, 정동열 교수님, 유영준 박사님께 감사드립니다.

참 고 문 헌

- 유사라. 1991. 네트워크 연구의 사회적 접근과 분석이론. 『도서관학』, 20: 297-330.
- 유영준. 2003. 문헌정보학의 지식 구조에 관한 연구. 박사학위논문, 연세대학교 대학원.
- 이은숙, 정영미. 2002. 복수저자를 고려한 동시 인용분석 연구: 정보학과 컴퓨터과학을 대상으로. 『지식처리연구』, 3(2): 1-26.
- 이재윤. 2006. 지적 구조의 규명을 위한 네트워크 형성 방식에 관한 연구. 『한국문헌정보학회지』, 40(2): 333-355.
- 정동열, 최윤미. 1999. 웹 정보원의 동시인용분석에 관한 실험적 연구. 『정보관리학회지』, 16(2): 7-25.
- Bollen, Johan, Herbert Van de Sompel, Joan A. Smith, and Rick Luce. 2005. "Toward alternative metrics of journal impact: A comparison of download and citation data." *Information Processing & Management*, 41(6): 1419-1440.
- Bonacich, P. 1972. "Factoring and weighting approaches to status scores and clique identification." *Journal of Mathematical Sociology*, 2: 113-120.
- Callon, M., J. P. Courtial, & F. Laville. 1991. "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry." *Scientometrics*, 22(1): 153-205.
- Chen, C. 2005. "The centrality of pivotal points

- in the evolution of scientific networks." *International Conference on Intelligent User Interfaces(IUI 2005)*, pp. 98-105.
- Crane, D. 1969. "Social structure in a group of scientists: A test of the invisible college hypothesis." *American Sociological Review*, 34, 335-352.
- Crane, D. 1972. *Invisible Colleges: Discussion of Knowledge in Scientific Communities*. Chicago, University of Chicago Press.
- Dastidar, Prabir G. 2004. "Ocean science & technology research across the countries: A global scenario." *Scientometrics*, 59 (1): 15-27.
- de Nooy, W., A. Mrvar, and V. Batagelj. 2005. *Exploratory Social Network Analysis with Pajek*. Cambridge University Press.
- Ding, Y, G. G. Chowdhury, and S. Foo. 2001. "Bibliometric cartography of information retrieval research by using co-word analysis." *Information Processing & Management*, 37(6): 817-842.
- Doreian, P. 1985. "A measure of standing of journals in stratified networks." *Scientometrics*, 8: 341-363.
- Doreian, P., and T. J. Fararo. 1985. "Structural equivalence in a journal network." *Journal of the American Society for Information Science*, 36(1): 28-37.
- Dosa, M. 1994. *The Health Information Sharing Project(HISP): Summary of Facts*. Syracuse, NY: Syracuse University School of Information Studies. Unpublished report. Quoted in Gray et al (1998).
- Dosa, M. L. 1982. "Information transfer in gerontology and health: a centralized and a decentralized model." *Special Collection*, 1(3-4): 53-72. Quoted in Gray et al(1998).
- Freeman, L. C. 1979. "Centrality in social networks conceptual clarification." *Social Networks*, 1, 215-239.
- Glänzel, Wolfgang, et al. 2003. *Nanotechnology - An Analysis Based on Publications and Patents*.
<<http://www.steunpuntoos.be/nanotech.html>>.
- Granovetter, Mark. 1973. "The strength of weak tie." *American Journal of Sociology*, 78: 1360-1380.
- Gray, S. A., and R. O'Shea, M. E. Petty, and J. Loonsk. 1998. "The western New York health resources project: developing access to local health information." *Bulletin of the Medical Library Association*, 86(3): 335 - 339.
- Haythornthwaite, Caroline, 1996. Social network analysis: an approach and technique for the study of information exchange. *Library and Information Research*, 18: 323-342.
- Haythornthwaite, Caroline. 2006. "Learning and knowledge networks in interdisciplinary collaborations." *Journal of the American Society for Information Science and*

- Technology*, 57(8): 1079-1092.
- Hersberger, Julie. 2003. "A qualitative approach to examining information transfer via social networks among homeless populations." *New Review of Information Behaviour Research*, 4(1): 95-108.
- Knoke, D., and R. S. Burt. 1983. "Prominence." In R. S. Burt and M. J. Minor(eds.), *Applied Network Analysis: A Methodological Introduction*, SAGE Publications, pp. 195-222..
- Liu, X., J. Bollen, M. L. Nelson, and H. Van de Sompel. 2005. "Co-authorship networks in the digital library research community." *Information Processing & Management*, 41(6): 1462-1480.
- Marion, L. S., E. Garfield, L. L. Hargens, and L. A. Lievrouw. 2003. "Social network analysis and citation network analysis: complementary approaches to the study of scientific communication." *Proceedings of the American Society for Information Science and Technology*, 40(1): 486-487.
- Nagpaul, P. S. 2002. "Visualizing cooperation networks of elite institutions in India." *Scientometrics*, 54(2): 213-228.
- National Library of Medicine. 1978. *Development and Evaluation of Health Information Sharing*. NY: Syracuse University School of Information Studies. Unpublished report. 유사라(1991)에서 재인용.
- Newman, M. E. J. 2004a. "Coauthorship networks and patterns of scientific collaboration." *Proceedings of the National Academy of Sciences*, 101: 5200-5205.
- Newman, M. E. J. 2004b. "Analysis of weighted networks." *Physical Review E*, 70, 056131.
- Otte, E., and R. Rousseau. 2002. "Social network analysis: a powerful strategy, also for the information sciences." *Journal of Information Science*, 28(6): 441-454.
- Page, L., S. Brin, R. Motwani, and T. Winograd. 1998. *The Pagerank Citation Ranking: Bringing Order to the Web*. Technical report, Stanford University Database Group. Available at <<http://dbpubs.stanford.edu:8090/pub/1999-66>>.
- Price, D. J. de Solla. 1963. *Little Science, Big Science*. New York: Columbia University Press.
- Schubert, András. 2002. "The web of scientometrics." *Scientometrics*, 53(1): 3-20.
- Schvaneveldt, R. W.(ed). 1990. *Pathfinder Associative Networks: Studies in Knowledge Organization*. Norwood, NJ: Ablex.
- Scott, J. P. 2000. *Social Network Analysis: A Handbook*. London: SAGE Publications.
- Sonnenwald, Diane H., and Linda G. Pierce. 2000. "Information behavior in dynamic group work contexts: interwoven situa-

- tional awareness, dense social networks and contested collaboration in command and control." *Information Processing & Management*, 36(3): 461-479.
- White, H. D., and B. C. Griffith. 1981. "Author cocitation: a literature measure of intellectual structure." *Journal of the American Society for Information Science*. 32: 163-171.
- Yoshikane, F., and K. Kageura. 2004. "Comparative analysis of coauthorship networks of different domains: the growth and change of networks." *Scientometrics*, 60(3): 435-446.
- Zuccala, Alesia. 2006. "Modeling the invisible college." *Journal of the American Society for Information Science and Technology*, 57(2): 152-168.