

공저 네트워크 분석을 위한 중심성 척도 비교 분석

Comparing Centrality Measures for Analyzing Co-authorship Networks

이재윤, 명지대학교 문헌정보학과, memexlee@mju.ac.kr

Jae Yun Lee, Library & Information Science Department, Myongji University

공동연구 네트워크의 대표적인 사례인 공저 네트워크는 오랫동안 네트워크 분석의 대상으로 다루어져 왔다. 최근에는 가중 네트워크로서 공저 네트워크에 대한 연구가 활발해지면서 연구자의 영향력을 측정하려는 몇 가지 척도가 제안되었다. 이 연구에서는 공저 네트워크에서의 중심성을 측정하기 위해서 사용된 척도인 가중페이지랭크, 공동연구 h-지수와 공동연구 hs-지수, 복합연결정도중심성, c-지수에 대해서 비교 분석해본다.

1. 서론

공저 네트워크는 대표적인 사회 네트워크로서 오랫동안 연구의 대상이 되어왔다. 공저 네트워크에 대한 연구는 네트워크 전체에 대한 구조적 특성을 밝히고자 하는 관점과, 개별 저자의 입지를 살펴보고자 하는 관점으로 나눌 수 있다. 이에 더해서 시간 흐름에 따른 변화를 살펴보는 연구가 종종 수행된다.

공저 네트워크에서 각 저자의 영향력, 또는 중요도를 분석하려는 시도는 오랫동안 이진네트워크 중심성을 측정하는 것이 대부분이었으며 최근까지도 이런 경향은 계속되고 있다(이수상, 2011; Yan & Ding, 2009).

한편 공저자의 수 이외에 공저 횟수를 고려하여 공저 네트워크를 가중 네트워크로 간주하고 중심성을 측정하려는 시도가 간헐적으로 발표되었다. 이 연구에서는 공저 네트워크에서의 중심성을 측정하기 위해서 사용된 척도인 가중페이지랭크, 공동연구 h-지수와 공동연구 hs-지수, 복합연결정도중심성, c-지수에 대해서 비교 분석해본다.

2. 공저자 중심성 계산에 사용된 척도

2.1 가중 페이지랭크

가중 페이지랭크는 링크의 강도를 고려하도록 페이지랭크 공식을 수정한 것이다. Liu 등 (2005)은 비방향성 네트워크인 공동연구 네트워크에서 페이지랭크를 산출하기 위해 비방향성인 공저 관계 한 건을, 주고받는 두 개의 관계로 변환하는 방법을 사용하였다. 공저 횟수를 반영할 수 있도록 수정된 가중 페이지랭크 공식은 다음과 같이 나타낼 수 있다.

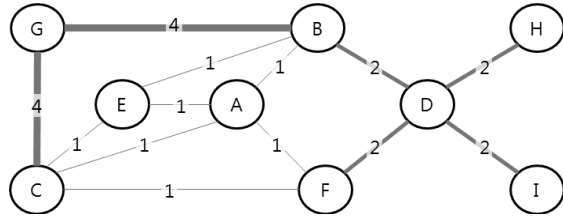
$$WPR_i = \frac{1-d}{N} + d \cdot \sum_{j=1}^N WPR_j \frac{CO_{ji}}{C_j}$$

이 공식에서 각 항의 의미는 다음과 같다.

- WPR_i - 연구자 i의 가중 페이지랭크
- CO_{ji} - 연구자 j와 연구자 i의 공저 횟수
- C_j - 연구자 j가 다른 연구자와 공저한 개별 횟수의 합계
- d - 상수, 보통은 0.85
- N - 전체 연구자 수

<그림 1>의 9명의 저자로 구성된 가상 공

저네트워크에서 페이지랭크와 가중 페이지랭크를 구하면 <표 1>과 같다.



<그림 1> 가상 공저 네트워크 (선 위의 숫자는 공저 횟수)

<표 1> 가상 저자 9명의 가중 페이지랭크

	A	B	C	D	E	F	G	H	I
PR	0.140	0.147	0.142	0.168	0.107	0.113	0.078	0.053	0.053
WPR	0.090	0.160	0.140	0.184	0.070	0.092	0.152	0.056	0.056

2.2 복합연결정도중심성 CDC

Kretschmer와 Kretschmer(2006)는 가중 네트워크를 위해 연결정도중심성을 개선한 복합연결정도중심성 CDC(Complex Degree Centrality)를 제안하였다. 이들이 CDC를 제안하면서 고려한 원칙은 세 가지이다.

첫째, 공저한 연구자의 수가 동일하다면, 공저 횟수가 더 많은 연구자가 더 큰 CDC를 가진다.

둘째, 모든 저자 사이의 공저 횟수가 동일하다면 공저자 수가 더 많은 연구자가 더 큰 CDC를 가진다.

셋째, 공저자 수가 동일하고 공저 횟수의 합도 동일할 경우에는 각 공저자와의 공저 강도가 균일할수록 더 큰 CDC를 가진다.

연구자 A의 복합연결정도중심성 CDC_A 를 산출하는 공식은 다음과 같이 구성된다.

$$CDC_A = \sqrt{''DC_A'' \times TR_A} = \sqrt{2^{H(A)} \times TR_A}$$

이 공식에서 앞 부분인 "DC_A"는 A의 공저자 수를 반영하는 항이고 TR_A는 공저 횟수를 반영하는 항이다. 연구자 A의 공저자 “수

(number)”인 “DC_A”는 엔트로피를 2의 지수로 적용하여 측정된다. 연구자 A와 공저자 i 사이의 공저 횟수를 K_i라고 할 때, 엔트로피 H(A)는 다음과 같이 측정된다.

$$H(A) = - \sum_{i=1}^z \left\{ \frac{K_i}{TR_A} \times \log_2 \frac{K_i}{TR_A} \right\}$$

<그림 1>의 가상 공저 네트워크에서 3명과 공저한 연구자 F의 예로 CDC 계산 과정을 살펴보면 다음과 같다. 연구자 F는 공저한 연구자가 D, A, C의 3명이고 각 연구자와 2회, 1회, 1회 공저했으므로 연결정도는 3이고 공저 횟수 합계 TR은 4가 된다. 이때 연구자 F의 “DC_F”는 엔트로피인 $-(2/4 \times \log_2(2/4) + 1/4 \times \log_2(1/4) + 1/4 \times \log_2(1/4)) = 1.585$ 를 2의 지수로 취한 값인 2.828이 된다. 최종적으로 연구자 F의 복합연결정도중심성 CDC_A 는 공저횟수 합계인 4와 공저자 “수”인 2.828의 기하 평균인 3.364가 된다.

이런 방식으로 각 연구자의 복합연결정도중심성을 측정한 결과는 <표 2>와 같다. 앞에서 언급한 Kretschmer와 Kretschmer(2006)의 세 가지 원칙 중에서 세 번째 원칙에 해당하는 비교 사례로는 연구자 B와 D를 들 수 있다. 연구자 B와 D는 공저자 수가 각각 4명으로 동일하고 공저횟수의 합도 각각 8로 동일하지만, 연구자 B는 각 공저자와의 공저횟수가 불균일한 반면(4회, 2회, 1회, 1회) 연구자 D는 각 공저자와의 공저횟수가 모두 2회로 균일하므로 연구자 D가 더 큰 CDC를 가진다(5.657 > 5.187).

<표 2> 가상 저자 9명의 복합연결정도중심성

	A	B	C	D	E	F	G	H	I
CDC	4.000	5.187	4.711	5.657	3.000	3.364	4.000	1.414	1.414

2.3 공동연구 h-지수와 hs-지수

공동연구 h-지수(collaboration h-index; col-h로 약칭)는 Hirsh(2005)의 h-지수를 공

저 네트워크에 응용한 것으로서(이재윤, 2010; Lee & Choi, 2013) 다음과 같이 정의할 수 있다.

특정 연구자가 공저자와 발표한 개별 공저 건수를 내림차순으로 정렬한 후 공저 건수보다 작거나 같은 마지막 순위가 해당 연구자의 공동연구 h(col-h)이다.

또한 공동연구 hs-지수(collaboration hs-index; col-hs로 약칭)는 공동연구 h-지수로 산출한 col-h위 이내의 공저 건수를 각각 제곱근을 취해서 합산한 값이다(이재윤, 2010; Lee & Choi, 2013). <그림 1>의 가상 공저 네트워크에서 페이지랭크와 가중페이지랭크를 구하면 <표 3>과 같다.

<표 3> 가상 저자 9명의 타 저자와의 공저횟수 순위와 col-h, col-hs

	A	B	C	D	E	F	G	H	I
1위	1	4	4	2	1	2	4	2	2
2위	1	2	1	2	1	1	4		
3위	1	1	1	2	1	1			
4위	1	1	1	2					
col-h	1	2	1	2	1	1	2	1	1
col-hs	1.000	3.414	2.000	2.828	1.000	1.414	4.000	1.414	1.414

2.4 협동능력 지수 c

c-지수(c-index)는 연구자의 공저횟수와 공저자수 이외에 공저한 상대방의 중요도까지 고려하도록 최근 제안된 지표이다. c-지수의 c는 협동능력(collaboration competence)을 의미하므로 c-지수는 협동능력 지수라고 의역할 수 있다. c-지수는 h-지수(Hirsch, 2005)의 개념을 응용하고 있으므로 정의도 다음과 같은 문장으로 제안되었다(Yan, Zhai, & Fan, 2013).

노드(연구자) x의 협동능력 지수 c(x)는 노드 x와 연결된 이웃 노드의 중요도와, 연결된 링크의 강도를 곱한 값이 c보다 크거나 같은 가장 큰 정수 c이다.

이때 어떤 연구자의 중요도는 타 연구자와의 공저 횟수 합계, 즉 연결된 링크 가중치의 합계로 나타낸다. c-지수를 h-지수에 비유해서 설명하자면, 한 연구자에 대해서 연결된 이웃 연구자의 중요도와 해당 이웃 연구자와의 링크 강도를 곱한 값을 큰 값부터 정렬한 후 h-지수처럼 값이 순위보다 크거나 같은 마지막 순위를 산출한 것이 c-지수에 해당한다.

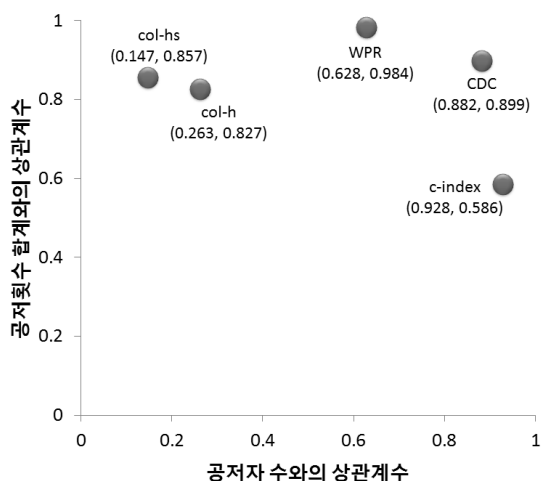
<그림 1>의 가상 공저 네트워크에서 각 저자의 c-지수를 산출해본 결과는 <표 4>와 같다. 이 네트워크에서 각 연구자의 중요도는 타 연구자와의 공저 횟수 합계가 된다. 연구자 A의 경우에 1회씩 협력한 연구자가 B, C, E, F이고 이들의 중요도가 각각 8, 7, 3, 4이므로 중요도와 협력횟수를 곱한 값을 내림차순으로 정렬하면 8, 7, 4, 3이 된다. 3번째 값이 4이고 4번째 값이 3이므로 값이 순위보다 큰 마지막 순위인 3이 연구자 A의 c-지수이다.

<표 4> 가상 공저 네트워크에서 산출한 저자 9명의 c-지수

저자	A	B	C	D	E	F	G	H	I	
중요도	4	8	7	8	3	4	8	2	2	
공저자 X 공저횟수	1위	8 (8X1)	32 (8X4)	32 (8X4)	16 (8X2)	8 (8X1)	16 (8X2)	32 (8X4)	16 (8X2)	16 (8X2)
	2위	7 (7X1)	16 (8X2)	4 (4X1)	8 (4X2)	7 (7X1)	7 (7X1)	28 (7X4)		
	3위	4 (4X1)	4 (4X1)	4 (4X1)	4 (2X2)	4 (4X1)	4 (4X1)			
	4위	3 (3X1)	3 (3X1)	3 (3X1)	4 (2X2)					
c-지수	3	3	3	4	3	3	2	1	1	

3. 시험 측정결과 비교

앞에서 가상 공저 네트워크에 대해 산출한 다섯 가지 척도와 공저자 수, 공저횟수 합계 사이의 상관계수를 측정해본 결과는 <그림 2>와 같다. 공동연구 h-지수와 공동연구 hs-지수는 공저자 수와는 상관성이 약하고 공저횟수의 합계와의 상관성만 0.8 이상으로 강하게 나타났다. 가중 페이지랭크는 공저횟수의



<그림 2> 5종 척도와 공저자 수 및 공저횟수 합계와의 상관계수

합계와의 상관성이 0.984로 거의 1에 가까우므로 공저자 수보다는 공저횟수에 크게 좌우됨을 알 수 있다. 반면에 c-지수는 공저자 수와의 상관성이 0.9 이상으로 매우 높아서 공저자 수에 좌우되는 성향이 강함을 알 수 있다. CDC는 공저자 수와의 상관성과 공저횟수 합계와의 상관성이 모두 0.9에 가깝게 높게 나타났다. 이는 CDC가 두 요인을 고르게 잘 반영하는 지표임을 뜻한다.

4. 결론

공저 네트워크를 대상으로 적용된 5가지 가중 네트워크 중심성 척도를 소규모 가상 자료에 대해 적용해본 결과 복합연결정도중심성 CDC가 공저자 수와 공저횟수합계의 두 요인을 고르게 높게 반영하는 것으로 나타났다.

현실의 공저 네트워크는 이 연구에서 시험적으로 적용해본 저자 9명의 가상 공저 네트워크와 달리 저자 수가 수십, 수백명 이상이며 공저 횟수도 수십회 이상인 경우가 흔하다. 특히 분석 단위가 개인 저자가 아닌 기

관이나 국가인 경우(이재운 2010; Lee & Choi, 2013)에는 다른 결과가 나타날 가능성도 있으므로 다양한 실제 데이터에 적용하는 실험을 통해 각 척도의 성향을 검증하는 후속 연구가 필요하다.

참고문헌

이수상 (2011). 공저빈도에 따른 공저 네트워크의 속성 연구: 문헌정보학 분야 4개 학술지를 중심으로. 한국도서관·정보학회지, 42(2), 105-125.

이재운 (2010). 계량분석 기법을 활용한 연구 동향 분석: LED 분야. 산업기술진흥원.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of America*, 102(46), 16569-16572.

Kretschmer, H., & Kretschmer, T. (2006, May). A new centrality measure for social network analysis applicable to bibliometric and webometric data. Paper presented at the Second International Workshop on Webometrics, Informetrics and Scientometrics & 7th COLLNET Meeting, Nancy, France.

Lee, J. Y., & Choi, S. (2013). Collaboration networks and document networks in informetrics research from 2001 to 2011. *Journal of the Korean Society for Information Management*, 30(1), 179-191.

Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management*, 41(6), 1462-1480.

Yan, E., & 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.

Yan, X., Zhai, L., & Fan, W. (2013). C-index: A weighted network node centrality measure for collaboration competence. *Journal of Informetrics*, 7(1), 223-239.