

가중 네트워크 분석을 위한 최근접이웃중심성 척도의 일반화

Generalizing Nearest Neighbor Centrality for Weighted Network Analysis

이재윤, 명지대학교 문헌정보학과, memexlee@mju.ac.kr
Jae Yun Lee, Library & Information Science Department, Myongji University

네트워크 분석이 확산되면서 여러 분야에서 다양한 중심성 척도가 개발되어 활용되고 있으나 가중 네트워크에서 지역중심성을 측정할 수 있는 척도로는 최근접이웃중심성 이외에는 거의 알려져 있지 않다. 최근접이웃중심성 척도는 동물값이 흔히 나타나므로 변별력이 낮다는 단점을 가지고 있다. 이 연구에서는 최근접이웃중심성 척도를 일반화한 이웃중심성 척도를 제안하고 가상 자료 및 실제 자료에 대해 적용하여 검증해보았다.

1. 서론

네트워크 분석의 여러 기법 중에서 가장 많이 활용되고 있는 것은 중심성 분석이라고 할 수 있다. Freeman(1979)이 지금까지도 널리 활용되고 있는 대표적인 중심성 척도를 정립한 이후에 여러 분야에서 다양한 척도가 개발되어 왔다. 최근에는 계량서지적 데이터를 비롯한 가중 네트워크 데이터를 분석하기 위한 가중 네트워크 중심성 척도에 대한 관심도 점차 증가하고 있다.

중심성 분석을 활용한 대부분의 연구에서는 한 가지 척도만이 아닌 여러 개의 중심성 척도를 동시에 사용한다. 이는 여러 척도를 사용하여 한 개체에 대한 다면적인 분석을 수행하기 위한 것이다. 그런데 기존에 제안된 가중 네트워크 중심성 척도들은 대부분 네트워크 전체에서의 중심성인 전역중심성을 측정하는 것이다. 지역중심성을 측정하기 위한 용도로 고안된 가중 네트워크 중심성 척도는 흔하지 않은데, 계량서지적 네트워크를 고려하여 개발된 최근접이웃중심성(Nearest Neighbor

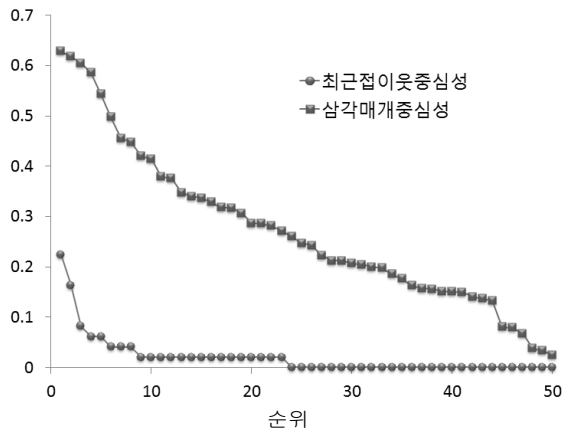
Centrality) 척도(이재윤, 2006)가 그중의 하나이다. 가중 네트워크에서 한 노드의 전역중심성과 지역중심성을 비교하면 중심성의 높고 낮음에 따라 대상을 네 가지 유형으로 구분할 수 있는 것으로 알려졌다(이재윤, 2006).

이처럼 최근접이웃중심성 척도는 가중 네트워크에서 지역 중심성을 측정하는 유용한 척도이지만, 실제 가중 네트워크 데이터에 대해 적용하였을 때 동물 순위가 많이 산출되어 변별력이 떨어진다는 단점이 있다. 최근접이웃중심성 척도는 전체 N 개의 노드 각각에 대해서 다른 노드로부터 최근접이웃으로 꼽히는 횟수를 산출하여 측정한다. 그 결과 소수의 노드를 제외한 상당수의 노드가 최근접이웃으로는 0회나 1회만 꼽히게 되어 동물이 된다. 이와 같이 낮은 변별력은 최근접이웃중심성 척도의 활용 가능성에 제약이 되고 있다.

이 연구에서는 최근접이웃중심성 척도를 일반화한 새로운 척도를 제안한다. 새로운 척도는 가중 네트워크에서 변별력이 높은 지역중심성 척도로서 가중 네트워크 분석의 활성화에 기여할 수 있을 것이다.

2. 최근접이웃중심성 지수의 변별력

1990년대 10년 간 대표적인 정보학 분야 학술지인 JASIS에서 인용된 정보학 분야 저자 50명의 동시인용 데이터(이은숙, 정영미, 2002)로 구성된 네트워크에 대해서 전역 가중 네트워크 중심성 지수인 삼각매개중심성과 지역 가중 네트워크 중심성 지수인 최근접이웃중심성 지수를 측정해본 결과는 다음 그림과 같다.



<그림 1> 1990년대 정보학 분야 저자 50명의 저자동시인용 네트워크에서 측정한 가중 네트워크 중심성 지수의 순위별 값

이 그림을 보면 삼각매개중심성은 1위에서부터 50위까지 값이 점차 낮아지면서 동률 순위가 없는 것으로 측정되었지만, 최근접이웃중심성은 4위 이하에서 동률 순위가 다수 발생하였다. 이와 같은 최근접이웃중심성의 낮은 변별력은 척도를 활용하기 어렵게 만드는 제약 사항이 된다.

3. 이웃중심성 지수

이 연구에서는 최근접이웃중심성 지수를 일반화한 이웃중심성 (Neighbor Centrality) 지수를 제안하고자 한다. 여기서 의도하는 일반화란 척도 공식에 파라미터를 도입하여 이진

네트워크부터 가중 네트워크까지 다양한 상황에서의 지역중심성을 측정할 수 있도록 척도를 개선함을 의미한다. 이를 통해 기존의 이진 네트워크용 지역중심성 척도인 연결정도 중심성과 가중 네트워크용 지역중심성 척도인 최근접이웃중심성을 이 연구에서 제안한 척도의 양 극단에 해당하는 특수한 사례로 포함하고자 한다.

최근접이웃중심성 척도에 파라미터를 도입하여 일반화한 공식은 ‘이웃중심성(Neighbor Centrality; NC)’이라고 명명하며 다음과 같이 나타낼 수 있다.

$$C_N^\alpha(n_i) = \sum_{j \neq i} \frac{1}{r_{ji}^\alpha} = \sum_{j \neq i} r_{ji}^{-\alpha}$$

여기서 $C_N^\alpha(n_i)$ 는 노드 n_i 의 이웃중심성, r_{ji} 는 노드 n_j 와 n_i 의 링크 가중치가 노드 n_j 가 가진 링크 가중치 중에서 몇 번째로 높은지를 나타내는 링크 가중치 순위이다. 파라미터 α 는 링크 가중치 순위를 고려하는 정도를 조절하는 변수로서 0에서 ∞ 까지 설정 가능하다.

파라미터 α 가 0이면 순위의 차이를 무시하고 링크 하나 당 1의 값이 동일하게 누적되므로 이진 네트워크에서 링크의 수를 측정하는 연결정도중심성과 같아진다.

$$C_N^{\alpha=1}(n_i) = \sum_{j \neq i} \frac{1}{r_{ji}^0} = \sum_{j \neq i} 1$$

α 가 ∞ 이면 순위가 1위인 링크로부터는 그대로 $1(=1/1^\infty)$ 이 누적되고 2위 이하의 경우는 $0(=1/\infty)$ 이 누적되므로 1위보다 낮은 순위는 무시된다. 따라서 연결된 상대방으로부터 1위로 꼽힌 링크의 수를 측정하는 최근접이웃중심성 지수가 된다.

$$C_N^{\alpha=\infty}(n_i) = \sum_{j \neq i} \frac{1}{r_{ji}^\infty}$$

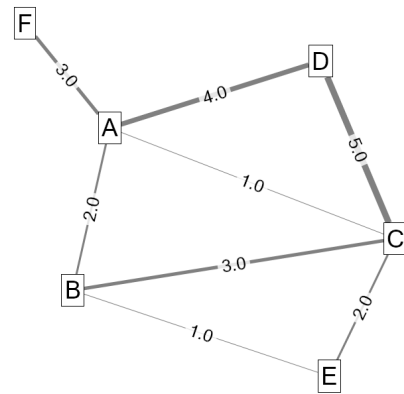
α 가 1이면 1위인 링크로부터는 1이 누적되고, 2위인 링크로부터는 $1/2$, 3위인 링크로부터는 $1/3$ 의 값이 누적된다. α 가 2이면 1위인 링크로부터는 1, 2위인 링크로부터는 $1/4$, 3위인 링크로부터는 $1/9$ 의 값이 누적된다. 두 링크의 가중치가 같은 경우는 모두 동률 순위를 부여한다. 예를 들어 1위가 3건이면 모두 1위로 설정하고 그 다음 순위를 4위로 정한다.

가상의 예로 6개의 노드가 <표 1>, <그림 2>와 같이 8개의 링크로 연결되어 있을 때, 각 노드가 링크를 통해 받는 순위는 <표 2>와 같다. 예를 들어 노드 B와 연결된 링크의 순위를 산출해보면 가중치가 3.0인 노드 C와의 링크가 1위, 가중치가 2.0인 노드 A와의 링크가 2위, 가중치가 1.0인 노드 E와의 링크가 3위가 된다. 노드 C는 이와 같은 방식으로 노드 B만 아니라 노드 D와 E로부터도 1위를 받고 노드 A로부터는 4위를 받게 된다. 6개 노드가 주고받는 순위 값으로 네트워크를 다시 표현하면 <그림 3>과 같다. 결과적으로 원래의 네트워크는 비방향성 네트워크였지만 척도값을 산출하기 위해서 변환한 결과는 방향성 네트워크가 된다.

노드 A가 받는 순위는 1위 1개, 2위 2개, 4위 1개인데 α 가 0이면 노드 A의 이웃 중심성은 $1 + 1/2 + 1/2 + 1/4 = 2.25$ 가 되고, α 가 2이면 노드 A의 이웃 중심성은 $1 + 1/4 + 1/4 + 1/16 = 1.56$ 이 된다. α 가 ∞ 이면 노드 A의 이웃 중심성은 1위로 꼽힌 링크의 개수인 1이 된다. α 를 0에서 1, 2, 3, 4, 그리고 ∞ 로 설정하였을 때 가상 네트워크의 각 노드가 가지게 되는 이웃중심성은 각각 <표 3>과 같다. <표 3>을 보면 α 가 2 이상이 될 경우에는 노드의 순위에 변화가 없는 것으로 나타나며, α 가 0이거나 무한대이면 동률 순위가 많이 나타난다.

<표 1> 가상 네트워크의 링크 가중치

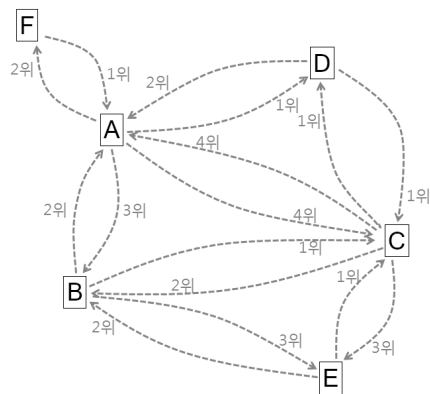
	A	B	C	D	E	F
A		2.0	1.0	4.0		3.0
B	2.0		3.0		1.0	
C	1.0	3.0		5.0	2.0	
D	4.0		5.0			
E		1.0	2.0			
F	3.0					



<그림 2> 6개 노드 가중 네트워크 가상 사례

<표 2> 각 노드가 링크를 통해 받는 순위

		순위 산출 기준 노드					
		A	B	C	D	E	F
순위값 받는 노드	A		2위	4위	2위		1위
	B	3위		2위		2위	
	C	4위	1위		1위	1위	
	D	1위		1위			
	E		3위	3위			
	F	2위					



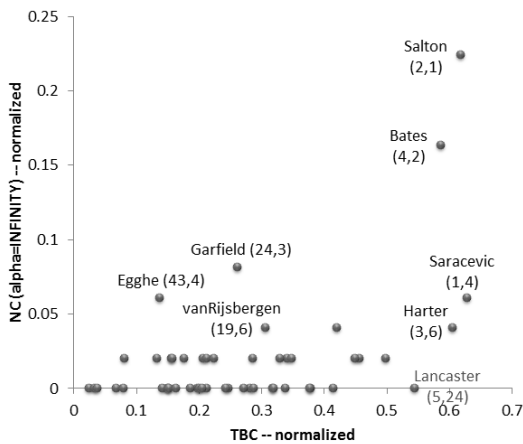
<그림 3> 6개 노드가 주고받는 순위

<표 3> α 값에 따른 각 노드의 이웃 중심성

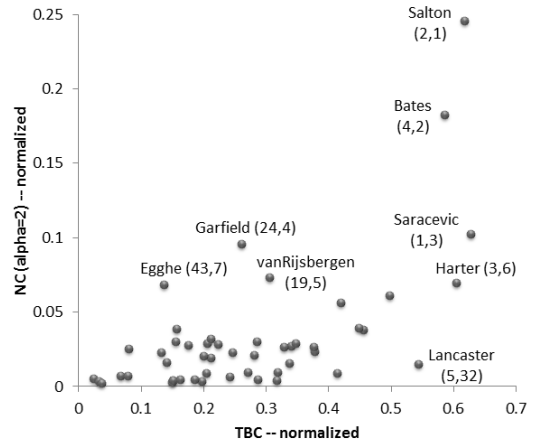
노드	$C_N^{\alpha=0}$	$C_N^{\alpha=1}$	$C_N^{\alpha=2}$	$C_N^{\alpha=3}$	$C_N^{\alpha=4}$	$C_N^{\alpha=\infty}$
A	4	2.250	1.563	1.266	1.129	1
B	3	1.333	0.611	0.287	0.137	0
C	4	3.250	3.063	3.016	3.004	3
D	2	2.000	2.000	2.000	2.000	2
E	2	0.667	0.222	0.074	0.025	0
F	1	0.500	0.250	0.125	0.063	0

이웃중심성 척도를 실제 계량서지적 자료인 1990년대 정보학 분야 저자 50명의 동시인용 네트워크(<그림 1> 참조)를 대상으로 적용해 보았다. 저자 50명에 대해서 $\alpha=\infty$ 인 이웃중심성(최근접이웃중심성)과 $\alpha=2$ 인 이웃중심성을 측정해 다음, 전역 중심성 척도인 삼각매개중심성(TBC) 측정 결과와 함께 <그림 4>와 <그림 5>의 산포도 형식으로 제시하였다.

실제 데이터에서도 최근접이웃중심성 척도($\alpha=\infty$ 인 이웃중심성 척도)로는 <그림 4>와 같이 세로축 좌표가 동일한 경우, 즉 동률 순위가 다수 발생한다. 반면에 $\alpha=2$ 로 설정한 이웃중심성 척도는 <그림 5>와 같이 동률 순위가 거의 발생하지 않으면서 저자들의 입지는 최근접이웃중심성을 적용한 결과와 유사하였다.



<그림 4> 1990년대 정보학 분야 저자동시인용 네트워크에서의 삼각매개중심성과 이웃중심성($\alpha=\infty$) (괄호 안은 삼각매개중심성과 이웃중심성 순위)



<그림 5> 1990년대 정보학 분야 저자동시인용 네트워크에서의 삼각매개중심성과 이웃중심성($\alpha=2$) (괄호 안은 삼각매개중심성과 이웃중심성 순위)

4. 결론

최근접이웃중심성 척도를 일반화한 이웃중심성 척도를 제안하고 가상 자료 및 실제 자료에 대해서 적용해본 결과, 가중 네트워크에서 지역중심성을 측정하면서도 높은 변별력을 보이는 것으로 나타났다.

제안된 이웃중심성 척도는 방향성 네트워크와 비방향성 네트워크에 모두 적용할 수 있으므로 다양한 분야의 가중 네트워크 분석에 기여할 수 있을 것으로 기대된다.

참고문헌

이은숙, 정영미 (2002). 복수저자를 고려한 동시인용분석 연구: 정보학과 컴퓨터과학을 대상으로. 지식처리연구, 3(2): 1-26.
 이재운 (2006). 계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구. 한국문헌정보학회지, 40(3): 191-214.
 Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, 1: 215-239.