

가중 네트워크를 위한 일반화된 지역중심성 지수*

A Generalized Measure for Local Centralities in Weighted Networks

이재윤 (Jae Yun Lee)**

초 록

네트워크 분석이 확산되면서 매개중심성이나 연결정도중심성과 같은 다양한 중심성 지수가 개발되어 활용되고 있으나, 가중 네트워크에서 지역중심성을 측정할 수 있는 지수로는 최근접이웃중심성 이외에는 거의 알려져 있지 않다. 이 연구에서는 가중 네트워크를 위한 일반화된 지역중심성 지수인 이웃중심성 지수를 새롭게 제안한다. 이웃중심성 지수는 파라미터 α 를 사용하여 이진 네트워크를 위한 연결정도중심성 지수와 가중 네트워크를 위한 최근접이웃중심성 지수를 일반화한 것이다. 6가지 실제 네트워크 데이터를 대상으로 하여 제안된 지수의 특징과 적정 파라미터 값을 살펴보는 실험을 수행하고 결과를 보고하였다.

ABSTRACT

While there are several measures for node centralities, such as betweenness and degree, few centrality measures for local centralities in weighted networks have been suggested. This study developed a generalized centrality measure for calculating local centralities in weighted networks. Neighbor centrality, which was suggested in this study, is the generalization of the degree centrality for binary networks and the nearest neighbor centrality for weighted networks with the parameter α . The characteristics of suggested measure and the proper value of parameter α are investigated with 6 real network datasets and the results are reported.

키워드: 네트워크 분석, 중심성 지수, 가중 네트워크, 지역중심성, 이웃중심성
network analysis, centrality measures, weighted networks, local centrality,
neighbor centrality

* 이 논문의 일부는 제20회 한국정보관리학회 학술대회에서 발표된 바 있음.

** 명지대학교 문헌정보학과 부교수(memexlee@mju.ac.kr)

■ 논문접수일자: 2015년 5월 21일 ■ 최초심사일자: 2015년 5월 28일 ■ 게재확정일자: 2015년 6월 9일

■ 정보관리학회지, 32(2), 7-23, 2015. [<http://dx.doi.org/10.3743/KOSIM.2015.32.2.007>]

1. 서론

네트워크 분석은 최근 인문사회과학의 전 영역에서 활용되고 있다. 한국학술정보인용색인 KCI에서 인문학, 사회과학, 복합학 영역을 대상으로 네트워크 분석을 검색해보면 2013년과 2014년에는 각각 110건과 127건의 논문이 검색될 정도로 연구가 확산되어 있다. 네트워크 분석의 여러 기법 중에서 가장 많이 활용되고 있는 것은 중심성 분석이라고 할 수 있으며, 여러 분야에서 다양한 지수가 개발되어 왔다. 최근에는 계량서지적 데이터를 비롯한 가중 네트워크 데이터를 분석하기 위한 가중 네트워크 중심성 지수에 대한 관심도 점차 증가하고 있다.

중심성 분석을 활용한 대부분의 연구에서는 한 가지 지수만이 아닌 여러 개의 중심성 지수를 동시에 사용한다. 이는 여러 지수를 사용하여 한 개체에 대한 다면적인 분석을 수행하기 위한 것이다. 중심성 지수는 측정 범위에 따라서 노드의 주변만 고려하는 지역중심성 지수와 네트워크 전체를 고려하는 전역중심성 지수로 나눌 수 있다. 그런데 기존에 제안된 가중 네트워크 중심성 지수들은 대부분 네트워크 전체 수준에서 중심이 되는 정도인 전역중심성을 측정하는 것이며, 지역중심성을 측정하기 위한 용도로 고안된 가중 네트워크 중심성 지수는 흔하지 않다. 지역중심성은 네트워크의 각 부분에서 지역적으로 중심적인 역할을 하는 노드를 파악해주는 역할을 하므로 지적구조 분석과 같이 국지적인 해석이 중요한 분야에서는 반드시 적용해야 하는 지수이다.

가중 네트워크에서 지역중심성을 측정하기 위한 방법으로는 한 노드가 가진 링크의 가중치를

단순 합산하는 노드 강도(Node Strength)가 여러 연구에서 사용되어 왔다(Barrat, Barthelemy, Pastor-Satorras, & Vespignani, 2004; Newman, 2004). 일부에서는 이진 네트워크의 지역중심성은 연결정도(degree)로 측정하고 가중 네트워크의 지역중심성은 연결강도로 손쉽게 측정할 수 있다는 주장도 있다(Fagiolo, Reyes, & Schiavo, 2010, p. 486). 그러나 저자동시인용 네트워크나 상관계수 네트워크처럼 가중 네트워크가 조밀할 경우에는 네트워크의 거의 모든 다른 노드와의 관계를 고려하게 되므로 연결강도는 사실상 '지역'중심성이 아닌 전역중심성이 된다는 문제가 있다. 가중 네트워크용 지역중심성 지수로 Opsahl과 Agneessens, Skvoretz (2010)이 발표한 tnet의 가중연결정도중심성도 있는데, 이 지수는 연결정도중심성에 연결강도를 곱하는 방식으로 산출되므로 조밀한 가중 네트워크에서는 지역중심성이 아닌 전역중심성을 측정하게 된다는 비판에서 역시 자유롭지 못하다.

계량서지적 네트워크를 고려하여 개발된 가중 네트워크용 지역중심성 지수로 최근접이웃중심성(Nearest Neighbor Centrality) 지수(이재운, 2006)가 있으며, 최근 여러 계량서지학 연구에서 최근접이웃중심성으로 지역중심성을 측정하는 사례가 보고되고 있다(강범일, 박지홍, 2013; 강범일, 이재운, 2014; 김원진, 정영미, 2010; 김관준, 2015; 서선경, 정은경, 2013; 최상희, 2014; Shim, 2013).

이처럼 최근접이웃중심성 지수는 가중 네트워크에서 지역중심성을 측정하는 유용한 지수이지만, 실제 가중 네트워크 데이터에 대해 적용하였을 때 동률 순위가 많이 산출되어 변별

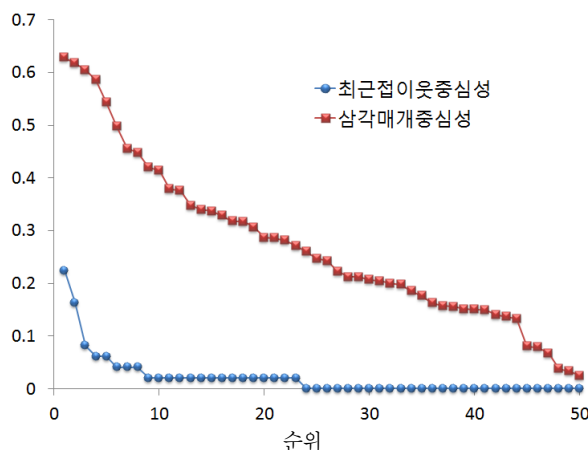
력이 떨어진다는 단점이 있다. 최근접이웃중심성 지수는 네트워크에 속한 전체 노드 각각에 대해서 다른 노드로부터 최근접이웃으로 꼽히는 횟수를 산출하여 측정한다. 그 결과 소수의 노드를 제외한 상당수의 노드가 최근접이웃으로는 0회나 1회만 꼽히게 되어 동물이 된다. 이와 같이 낮은 변별력은 최근접이웃중심성 지수의 활용 가능성에 제약이 되고 있다. 그렇다고 가중 네트워크에서 전통적인 연결정도중심성을 측정하면 링크 강도를 무시하고 링크의 수만 고려하게 되는 문제가 발생한다.

이 연구에서는 가중 네트워크를 위한 최근접이웃중심성 지수와 이진 네트워크를 위한 연결정도중심성 지수를 일반화하여 하나로 설명할 수 있는 새로운 지수를 제안한다. 새로운 지수는 파라미터의 설정에 따라 최근접이웃중심성 지수가 되거나 연결정도중심성 지수가 될 수 있으며, 가중 네트워크에서 변별력이 높은 지역중심성 지수로서 네트워크 분석의 활성화에 기여할 수 있을 것이다. 다음 2장에서는 가중

네트워크 분석에서 최근접이웃중심성 지수의 낮은 변별력 문제를 살펴보고, 이를 극복할 수 있는 방안으로 3장에서 새로운 이웃중심성 지수를 제안한 후, 4장에서 다양한 실제 계량서지적 네트워크에 적용해보고 적절한 파라미터 값을 파악해보았다.

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 가중 네트워크 중심성 지수 측정 프로그램인 WNET(이재운, 2013)을 이용하여 전역 가중 네트워크 중심성 지수인 삼각매개중심성과 지역 가중 네트워크 중심성 지수인 최근접이웃중심성 지수를 측정해본 결과는 <그림 1>과 같다.



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

〈그림 1〉을 보면 삼각매개중심성은 1위에서부터 50위까지 값이 점차 낮아지면서 동률 순위가 없는 것으로 측정되었지만, 최근접이웃중심성은 4위 이하에서 동률 순위가 다수 발생하였다. 이와 같은 최근접이웃중심성의 낮은 변별력은 지수를 활용하기 어렵게 만드는 제약사항이 된다.

〈표 1〉 〈표 2〉를 보면 50명 저자들이 가지는

최근접이웃중심성은 11, 8, 4가 각각 1명, 3이 2명, 2가 3명, 1이 15명, 0이 27명이어서 7가지 값에 불과하며, 가장 작은 값인 0으로 측정된 저자는 전체의 절반이 넘는 27명(54%)나 되어 변별력이 매우 부족하다. 반면에 연결정도중심성은 〈표 3〉 〈표 4〉와 같이 최고인 48부터 최저인 12까지 21가지 값이 측정되며 가장 많은 동률이 발생하는 경우도 4명(8%)에 불과했다.

〈표 1〉 정보학 분야 저자 50명의 저자동시인용 네트워크에서 측정한 최근접이웃중심성

저자	최근접이웃중심성	저자	최근접이웃중심성	저자	최근접이웃중심성
Salton	11	Kuhlthau	1	Small	0
Bates	8	Brooks	1	Marchionini	0
Garfield	4	Cronin	1	Spink	0
Saracevic	3	Losee	1	Wilson	0
Egghe	3	Price	1	Dervin	0
Harter	2	Rousseau	1	Chen	0
Borgman	2	Lancaster	0	Fox	0
vanRijsbergen	2	Bookstein	0	Meadow	0
Cooper	1	Ingwersen	0	Shaw	0
Swanson	1	Schamber	0	Kantor	0
Buckland	1	Robertson	0	Maron	0
McCain	1	Blair	0	Wong	0
Belkin	1	Brookes	0	Narin	0
Markey	1	Croft	0	Bush	0
Shannon	1	Larson	0	Kling	0
White	1	SparckJones	0	Burrell	0
Fidel	1	Harman	0		

〈표 2〉 정보학 저자 50명의 최근접이웃중심성 값 분포

최근접이웃중심성	해당 저자 수 (점유율)	최근접이웃중심성	해당 저자 수 (점유율)
11	1 (2%)	2	3 (6%)
8	1 (2%)	1	15 (30%)
4	1 (2%)	0	27 (54%)
3	2 (4%)		

〈표 3〉 정보학 분야 저자 50명의 저자동시인용 네트워크에서 측정된 연결정도중심성

저자	연결정도 중심성	저자	연결정도 중심성	저자	연결정도 중심성
Bates	48	Markey	42	Fox	35
Salton	47	Brookes	42	Meadow	35
Saracevic	47	Croft	42	Shaw	35
Harter	47	Garfield	41	Brooks	34
Lancaster	47	Larson	41	Cronin	34
Borgman	45	SparckJones	41	Kantor	34
Bookstein	45	Shannon	40	Maron	34
Ingwersen	45	Harman	40	Losee	33
Schamber	45	Small	40	Price	33
Cooper	44	White	39	Egghe	31
Swanson	44	Marchionini	39	Rousseau	26
Robertson	44	Spink	39	Wong	25
vanRijsbergen	43	Wilson	39	Narin	24
Buckland	43	Fidel	37	Bush	22
McCain	43	Dervin	37	Kling	16
Blair	43	Kuhlthau	36	Burrell	12
Belkin	42	Chen	36		

〈표 4〉 정보학 저자 50명의 연결정도중심성 값 분포

연결정도중심성	해당 저자 수 (점유율)	연결정도중심성	해당 저자 수 (점유율)
48	1 (2%)	35	3 (6%)
47	4 (8%)	34	4 (8%)
45	4 (8%)	33	2 (4%)
44	3 (6%)	31	1 (2%)
43	4 (8%)	26	1 (2%)
42	4 (8%)	25	1 (2%)
41	3 (6%)	24	1 (2%)
40	3 (6%)	22	1 (2%)
39	4 (8%)	16	1 (2%)
37	2 (4%)	12	1 (2%)
36	2 (4%)		

이처럼 동시인용 네트워크와 같은 가중 네트워크에서 최근접이웃중심성을 이용하여 지역중심성을 측정하면 변별력이 부족하며, 연결정도중

심성으로 지역중심성을 측정하면 동시인용횟수와 같은 링크의 가중치를 무시하게 되는 문제가 발생한다.

3. 이웃중심성 지수

이 연구에서는 최근접이웃중심성 지수를 일반화한 이웃중심성(Neighbor Centrality) 지수를 제안하고자 한다. 여기서 의도하는 일반화란 지수 공식에 파라미터를 도입하여 이진 네트워크부터 가중 네트워크까지 다양한 상황에서 지역중심성을 측정할 수 있도록 지수를 개선함을 의미한다. 이를 통해 기존의 이진 네트워크용 지역중심성 지수인 연결정도 중심성과 가중 네트워크용 지역중심성 지수인 최근접이웃중심성을 이 연구에서 제안한 지수의 양극단에 해당하는 특수한 사례로 포함하고자 한다.

최근접이웃중심성 지수에 파라미터를 도입하여 일반화한 공식은 '이웃중심성(Neighbor Centrality: NC)'이라고 명명하며 다음과 같다.

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

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 무한대(∞)까지 설정 가능하다. 0일 때에는 순위를 전혀 고려하지 않고 값이 커질수록 순위를 크게 고려하게 되며, 무한대일 경우에는 1위만 고려하고 나머지 순위는 무시하게 된다.

파라미터 α 가 0이면 순위의 차이를 무시하고 링크 하나 당 1의 값이 동일하게 누적되므로

다음과 같이 이진 네트워크에서 링크의 수를 측정하는 연결정도중심성과 같아진다.

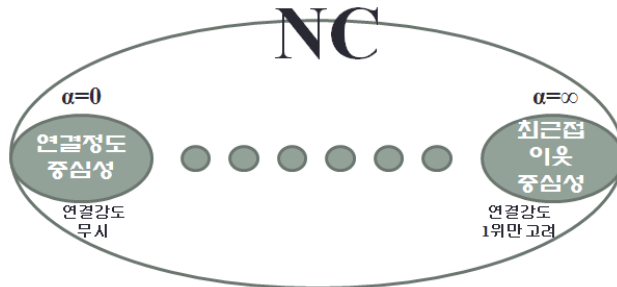
$$C_N^{\alpha=1}(n_i) = \sum_{j \neq i} \frac{1}{r_{ji}^0} = \sum_{j \neq i} \frac{1}{1} = \text{링크의 갯수}$$

α 가 ∞ 이면 순위가 1위인 링크로부터는 그 대로 $1(=1/1^\infty)$ 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.

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

결국 이웃중심성 지수 NC는 <그림 2>와 같이 파라미터 α 가 0일 때에는 링크의 연결강도를 무시하고 링크의 수만 세는 연결정도 중심성 지수가 되고, 파라미터 α 가 무한대일 때에는 상대방으로부터 1위로 꼽힌 링크만 고려하고 2위 이하로 꼽힌 링크는 무시하는 최근접이웃중심성 지수가 된다. 파라미터 α 를 0보다 크고 무한대보다 작은 값으로 설정하면 두 극단 사이에서 각 링크의 상대적인 강도, 즉 순위를 고려하는 지수가 된다. 파라미터 α 가 0에 가깝게 작으면 지수 측정 결과가 링크의 상대적인 강도보다는 링크의 개수에 좌우되게 되고, 파라미터 α 가 커질수록 상대적으로 강한 강도의 링크에 측정 결과가 좌우되게 된다.

α 가 1이면 1위인 링크로부터는 1이 누적되고, 2위인 링크로부터는 1/2, 3위인 링크로부터는 1/3의 값이 누적된다. α 가 2이면 1위인 링크



〈그림 2〉 파라미터 α 에 따라 연결정도중심성과 최근접이웃중심성을 양극단으로 포함하는 이웃중심성의 개념

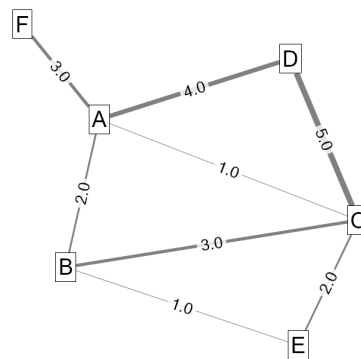
로부터는 1이 누적되고, 2위인 링크로부터는 1/4, 3위인 링크로부터는 1/9의 값이 누적된다. 두 링크의 가중치가 같은 경우는 모두 동률 순위를 부여한다. 예를 들어 1위가 3건이면 모두 1위로 설정하고 그 다음 순위를 4위로 정한다. 가상의 예로 6개의 노드가 〈표 5〉, 〈그림 3〉과 같이 8개의 링크로 연결되어 있을 때, 각 노드가 링크를 통해 받는 순위는 〈표 6〉과 같다. 예를 들어 노드 B와 연결된 링크의 순위를 산출해보면 가중치가 3.0인 노드 C와의 링크가 1위, 가중치가 2.0인 노드 A와의 링크가 2위, 가중치가 1.0인 노드 E와의 링크가 3위가 된다. 노

드 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 4〉와 같다. 결과적으로 원래의 네트워크는 비방향성 네트워크였지만 지수값을 산출하기 위해서 변환한 결과는 방향성 네트워크가 된다. 물론 원래 네트워크가 방향성 네트워크인 경우에도 문제없이 그대로 적용하여 방향성 순위 네트워크로 변환한 후 지수를 산출하게 된다.

이처럼 상대적인 순위를 이용하는 이웃중심성 지수에서는 두 노드가 주고받는 동일한 가

〈표 5〉 가상 네트워크의 링크 가중치

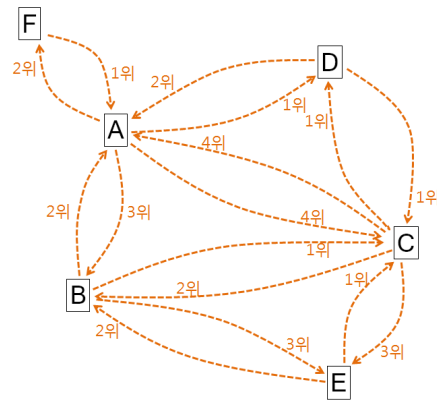
	A	B	C	D	E	F
A		2.0	1.0	4.0		3.0
B	2.0		3.0		1.0	
C	1.0	3.0		5.0	2.0	
D	4.0		5.0			
E		1.0	2.0			
F	3.0					



〈그림 3〉 6개 노드 가중 네트워크 가상 사례

〈표 6〉 가상 네트워크에서 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 4〉 6개 노드가 주고받는 순위

중치를, 상대방 노드가 가진 링크 가중치의 수준에 따라서 상대적인 값으로 변환해서 해석하게 된다. 예를 들면 서로 2회 동시인용된 두 저자라고 할지라도 총 2회 인용된 저자의 입장에서 2회 동시인용된 상대방 저자의 중요도와, 총 100회 인용된 저자의 입장에서 2회 동시인용된 상대방 저자의 중요도는 매우 다르다. 따라서 이와 같이 주변 정보를 고려하여 상대적인 수준으로 링크 가중치를 해석함으로써 각 노드의 지역적인 입지, 즉 지역중심성을 더욱 세밀하게 측정하게 된다.

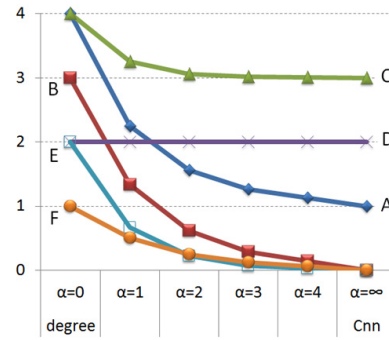
노드 A가 받는 순위는 1위 1개, 2위 2개, 4위 1개인데 α 가 0이면 노드 A의 이웃 중심성은 연결정도중심성과 같이 링크 수인 4가 된다. α 가 1이면 노드 A의 이웃 중심성은 $1+1/2+1/2+1/4=2.25$ 가 되고, α 가 2이면 노드 A의 이웃 중심성은 $1+1/4+1/4+1/16$ 인 1.56이 된다. α 가 ∞ 이면 노드 A의 이웃 중심성은 1위로 꼽힌 링크의 개수인 1이 된다. α 를 0에서 1, 2, 3, 4, 그리고 ∞ 로 설정하였을 때 가상 네트워크의 각 노드가 가지게 되는 이웃중심성은 각각 〈표 7〉

과 같다. 〈표 7〉을 보면 α 가 2 이상이 될 경우에는 노드의 순위에 변화가 없는 것으로 나타나며, α 가 0이거나 무한대이면 동률 순위가 많이 나타난다.

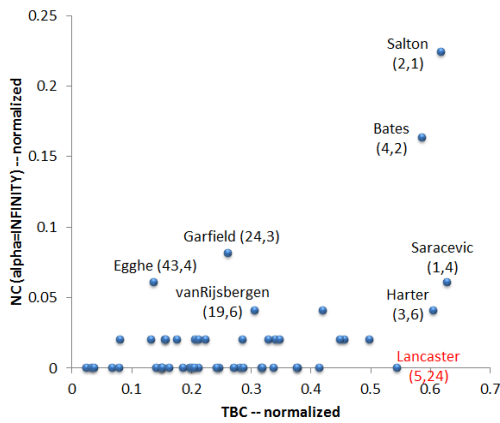
이웃중심성 지수를 실제 계량서지적 자료인 1990년대 정보학 분야 저자 50명의 동시인용 네트워크(이은숙, 정영미, 2002)를 대상으로 적용해보았다. 지숫값 산출을 위한 프로그램은 Python과 R의 두 가지 언어로 각각 구현하여 사용하였다. 저자 50명에 대해서 $\alpha=\infty$ 인 이웃중심성(최근접이웃중심성)과 $\alpha=2$ 인 이웃중심성을 측정한 다음, 전역중심성 지수인 삼각매개중심성(TBC) 측정 결과와 함께 〈그림 6〉과 〈그림 7〉의 산포도 형식으로 제시하였다. 실제 데이터에서도 최근접이웃중심성 지수($\alpha=\infty$ 인 이웃중심성 지수)로는 〈그림 6〉과 같이 세로축 좌표가 동일한 경우, 즉 동률 순위가 다수 발생한다. 반면에 $\alpha=2$ 로 설정한 이웃중심성 지수는 〈그림 7〉과 같이 동률 순위가 거의 발생하지 않으면서 저자들의 입지는 최근접이웃중심성을 적용한 결과와 유사하였다. 예를

〈표 7〉 α 값에 따른 각 노드의 이웃 중심성

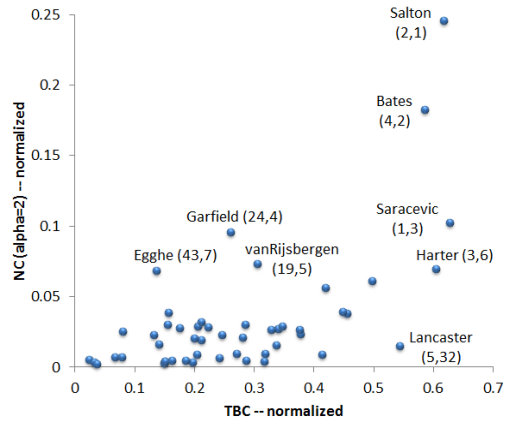
노드	C_N $\alpha=0$	C_N $\alpha=1$	C_N $\alpha=2$	C_N $\alpha=3$	C_N $\alpha=4$	C_N $\alpha=\infty$
A	4	2,250	1,563	1,266	1,129	1
B	3	1,333	0,611	0,287	0,137	0
C	4	3,250	3,063	3,016	3,004	3
D	2	2,000	2,000	2,000	2,000	2
E	2	0,667	0,222	0,074	0,025	0
F	1	0,500	0,250	0,125	0,063	0



〈그림 5〉 α 값의 증가에 따른 각 노드의 이웃 중심성 변화



〈그림 6〉 1990년대 정보학 분야 저자동시인용 네트워크에서의 삼각매개중심성과 이웃중심성($\alpha = \infty$)
(괄호 안은 삼각매개중심성과 이웃중심성 순위)



〈그림 7〉 1990년대 정보학 분야 저자동시인용 네트워크에서의 삼각매개중심성과 이웃중심성($\alpha = 2$)
(괄호 안은 삼각매개중심성과 이웃중심성 순위)

들어 최근접이웃중심성 1위와 2위인 Salton과 Bates는 이웃중심성($\alpha=2$)으로도 1위와 2위이며, 최근접이웃중심성 7위 이내에 포함된 Garfield, Egghe, Saracevic, van Rijsbergen, Harter는 이웃중심성으로도 7위 이내에 포함된다. 전역중심성에 비해서 지역중심성이 극단적으로 낮은 사례인 Lancaster의 사례를 보아도 최근접이웃중심성으로는 최하위인 공동 24위인데 이웃중

심성($\alpha=2$)으로는 오히려 더 낮은 단독 32위로 측정되어서 낮은 지역중심성이라는 결과는 같았다. 오히려 이웃중심성의 경우 동물 순위가 거의 발생하지 않기 때문에 변별력이 높아졌으며, 그로 인해 Lancaster의 지역중심성 순위가 공동 24위에서 32위로 더 낮게 측정되는 효과를 얻어서 이례적인 Lancaster의 입지가 더 두드러지는 효과도 얻게 되었다.

4. 이웃중심성의 적정 파라미터 값 결정

4.1 파라미터 α 설정 시 고려할 조건

이웃중심성 지수에서 링크 강도를 반영하는 수준을 조절하는 파라미터 α 는 링크의 개수가 중요할 경우에는 0에 가깝게 낮은 값으로 설정해야 하고, 강한 링크를 가지고 있는지 여부가 중요할 경우에는 매우 큰 값으로 설정해야 한다. 그러나 일반적인 경우라면 파라미터 α 는 다음 조건을 만족하는 수준이 바람직하다

첫째, 이웃중심성 지수는 가중 네트워크에 적용하는 지수이므로 가급적 측정 결과가 이진 네트워크용 지수인 연결정도중심성보다는 가중 네트워크용 지수인 최근접이웃중심성 쪽에 더 가까운 값이 되어야 한다. α 가 커질수록 이웃중심성 측정 결과가 연결정도중심성과 달라지고 최근접이웃중심성에 가까워지므로, 이는 α 가 너무 작은 값이 아니어야 함을 의미한다.

둘째, 반면에 파라미터 α 가 너무 크면 이웃중심성 지수가 최근접이웃중심성 지수와 유사해지므로 변별력 면에서 유효성이 저하된다.

따라서 α 는 너무 큰 값이어도 곤란하고, 연결정도 중심성 지수보다 최근접이웃중심성 지수에 더 가까운 최소한의 수준으로 설정하는 것이 바람직하다.

이와 같은 두 가지 조건을 고려할 때 결국 이웃중심성 지수의 파라미터 α 는 너무 작지도, 너무 크지도 않은 적정 수준으로 설정해야 한다.

4.2 파라미터 α 설정 실험

앞에서 제시한 두 가지 조건을 만족할 수 있는 파라미터 α 의 적정 수준을 파악하기 위해서 <표 8>과 같이 다양한 가중 네트워크 데이터를 대상으로 중심성 측정을 시도해보았다.

중심성 측정에 사용된 6개의 실제 네트워크는 비방향성 네트워크 4종과 방향성 네트워크 2종으로서 저자, 문헌, 저널, 지역 등으로 단위가 다양하며 관계의 성격도 동시인용, 공저, 서지결합, 상호인용, 상호대차 등으로 다양하다. AC50은 앞에서 살펴본 1990년대 정보학 분야 저자 50명의 동시인용 네트워크이다(이은숙, 정영미, 2002). CA311은 마케팅 분야의 국제 학술

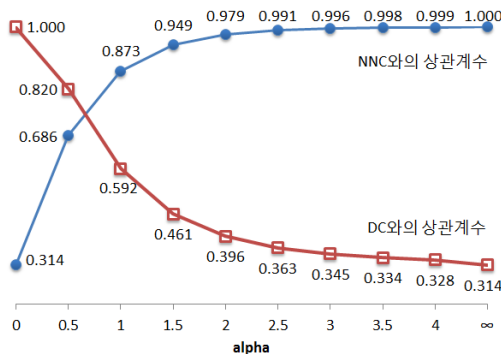
<표 8> 중심성 측정 대상 가중 네트워크 6종

네트워크 이름	AC50	CA311	BC91	JC116	J1112	IL31
종류	저자동시인용	공저 네트워크	문헌서지결합	저널동시인용	저널상호인용	지역 간 상호대차
분야	정보학	마케팅	폭소노미	약학	경영경제	도서관
방향성 여부	비방향성	비방향성	비방향성	비방향성	방향성	방향성
노드 수	50	311	91	116	112	31
링크 수	948	912	1665	6013	1212	884
최대 경로거리	2	11	2	2	5	2
평균 경로거리	1.226	4.009	1.593	1.099	2.381	1.058
네트워크 밀도	0.774	0.019	0.407	0.901	0.097	0.942

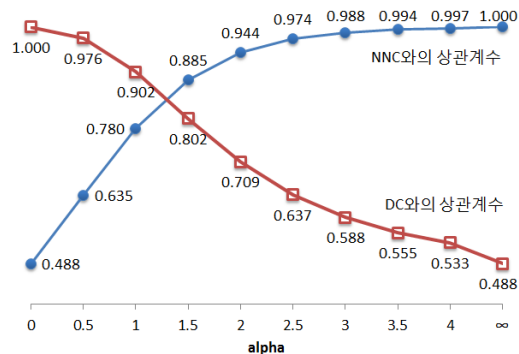
지 35중에 1973년부터 2009년까지 27년 동안 발표된 논문의 저자 중 발표논문수 상위 저자 311명 사이의 공동연구 데이터이다(이재운, 2014; Goldenberg, Libai, Muller, & Stremersch, 2010). BC91은 폭소노미 주제 논문 91편의 서지결합 네트워크이다(이재운, 2012). JC116은 약학분야 저널 116종의 동시인용 네트워크이다(조선례, 이재운, 2012). JI112는 한국학술지 인용색인 KCI의 경영학 및 경제학 분야 등재/등재후보지 112종이 2012년에 직전 3년 간의 논문을 상호 인용한 횟수로 구성된 저널상호인용 네트워크이다. 이는 KCI의 중심성지수를 측정할 때 사용되는 값이다. IL31은 경기도 지역 공공도서관 사이에 이루어진 도서관 상호대차를 31개 시 단위로 병합하여 구성한 네트워크이다(유종덕, 2013).

파라미터 α 를 위한 첫 번째 조건을 검토하기 위해서 각 네트워크 마다 산출된 노드들의 이웃중심성과 연결정도중심성(DC) 및 최근접이웃중심성(NNC) 사이의 피어슨 상관계수를 측정하여 지수 사이의 유사성을 파악하였다. 파라미터 α 를 0에서 0.5씩 증가시키면서 이웃중

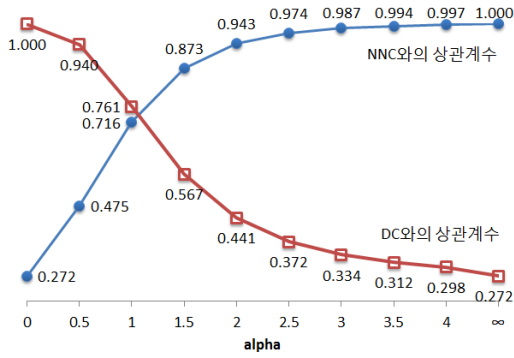
심성을 측정하면 처음에는 연결정도중심성과의 상관계수가 최근접이웃중심성과의 상관계수보다 크지만 α 가 커질수록 연결정도중심성과의 상관계수는 작아지고 최근접이웃중심성과의 상관계수는 커져서 두 값이 <그림 8>~<그림 13>과 같이 중간에서 교차된다. 최근접이웃중심성과의 상관계수가 더 커지는 지점을 확인해보면, 약학 저널 동시인용 네트워크 JC116과 지역 간 상호대차 네트워크 IL31은 $\alpha = 0.5$ 이상일 때이고, 정보학 저자 동시인용 네트워크 AC50과 경영경제 저널 상호인용 네트워크 JI112는 $\alpha = 1$ 이상일 때이며, 마케팅 분야 공저 네트워크 CA311과 폭소노미 문헌 서지결합 네트워크 BC91은 $\alpha = 1.5$ 이상일 때이다. 따라서 $\alpha = 1.5$ 이상이면 여섯 가지 네트워크 사례에서 모두 측정된 이웃중심성 지수값이 연결정도중심성(NC, $\alpha = 0$)보다 최근접이웃중심성(NC, $\alpha = \infty$)에 더 가깝게 된다. 이로서 앞에서 규정한 파라미터 α 의 설정시 고려해야 할 사항 첫 번째를 실제 데이터를 대상으로 검증해본 결과, α 는 최소 1.5 이상으로 설정해야 할 수 있다.



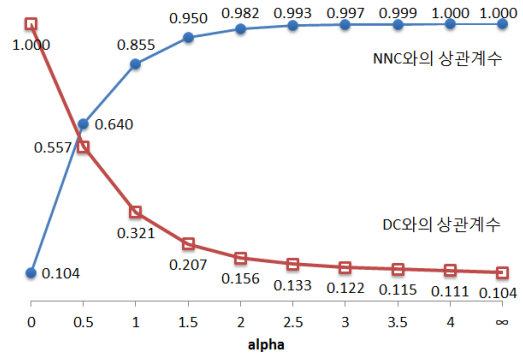
<그림 8> AC50에서 측정한 이웃중심성과 타 중심성 사이의 상관계수



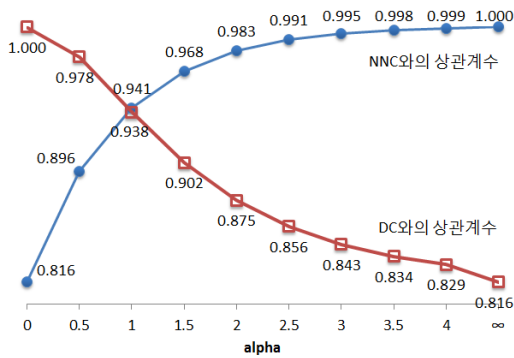
<그림 9> CA31에서 측정한 이웃중심성과 타 중심성 사이의 상관계수



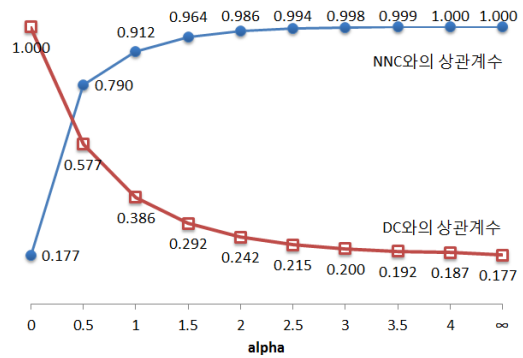
〈그림 10〉 BC91에서 측정한 이웃중심성과 타 중심성 사이의 상관계수



〈그림 11〉 JC116에서 측정한 이웃중심성과 타 중심성 사이의 상관계수



〈그림 12〉 J1112에서 측정한 이웃중심성과 타 중심성 사이의 상관계수



〈그림 13〉 IL31에서 측정한 이웃중심성과 타 중심성 사이의 상관계수

파라미터 α 의 설정시 고려해야 할 사항 두 번째는 α 가 너무 커서 최근접이웃중심성처럼 변별력이 저하되지 않도록 적절한 수준 이내로 설정해야 하는 점이다. 최근접이웃중심성 지수를 적용했을 때, 각 노드가 가지는 값의 가짓수가 너무 적었던 것이 문제라는 것에 착안하여, 이웃중심성 지수를 측정한 결과값을 정수로 반올림했을 때, 네트워크에 속한 노드가 가지는 중심성 값이 몇 가지인가를 세어서 해당 지수의 변별력으로 간주하였다. 예를 들어 어떤 네트워크에서 최근접이웃중심성 측정 결과가 4인 노

드가 1개, 3인 노드가 2개, 2인 노드가 5개, 1인 노드가 10개, 0인 노드가 20개였다면 다섯 가지 값이 측정되었으므로 변별력은 5이다. 그런데 이웃중심성으로 측정한 결과를 정수로 반올림하였을 때 중심성 값이 5, 4, 3, 2, 1, 0의 여섯 가지로 나타날 경우에는 변별력이 6이고, 최근접이웃중심성에 비하면 이웃중심성이 $6/5 = 120\%$ 의 변별력을 가지게 되는 셈이다.

여섯 가지 네트워크에 대해서 파라미터 α 를 0에서 4.0까지 0.5씩 늘려가면서 설정하고, 무한대인 최근접이웃중심성의 경우까지 설정하

여 10가지 수준에서 각 노드가 가지는 이웃중심성 값을 반올림해서 값의 가짓수를 세어보면 <표 9>와 같다. <표 9>에서 최근접이웃중심성 값의 가짓수보다 많은 경우는 음영 표시를 하였다. 또한 <그림 14>에서는 이웃중심성을 반올림한 값의 가짓수를 최근접이웃중심성 값의

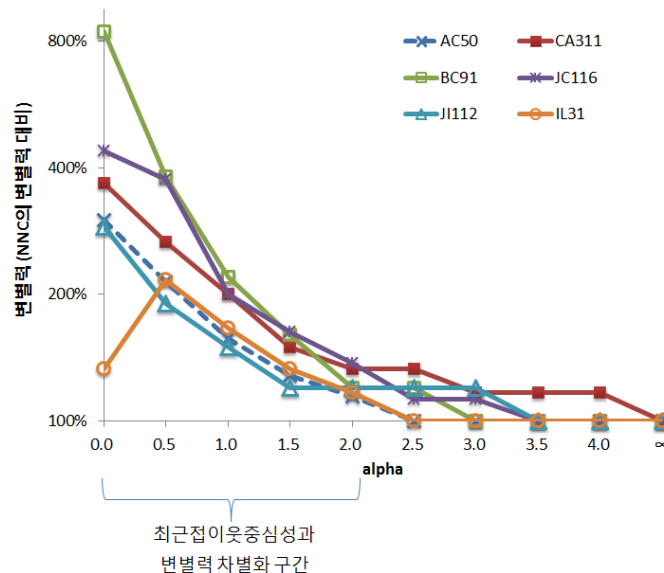
가짓수로 나누어 100을 곱한 것으로 표현한 변별력을 표시하였다. 최근접이웃중심성 값의 가짓수와 같을 경우는 이웃중심성의 변별력이 100%가 되고, 그보다 중심성 값의 가짓수가 많아지면 더 높은 변별력을 가진다.

<표 9>와 <그림 14>를 보면 파라미터 α 가 3.5

<표 9> 이웃중심성 측정 결과 값을 정수로 반올림했을 때 각 노드가 가지는 값의 가짓수

	AC50	CA311	BC91	JC116	J1112	IL31
$a=0.0$ (DC)	21	22	42	35	29	8
$a=0.5$	15	16	19	30	19	13
$a=1.0$	11	12	11	16	15	10
$a=1.5$	9	9	8	13	12	8
$a=2.0$	8	8	6	11	12	7
$a=2.5$	7	8	6	9	12	6
$a=3.0$	7	7	5	9	12	6
$a=3.5$	7	7	5	8	10	6
$a=4.0$	7	7	5	8	10	6
$a=\infty$ (NNC)	7	6	5	8	10	6

* 음영 부분은 $a=\infty$ (NNC)일 때보다 값의 가짓수가 더 많은 경우



<그림 14> 파라미터 α 의 변화에 따른 이웃중심성의 변별력 비교

이상이면 6개 사례 네트워크 중에서 마케팅분야 공저자 네트워크 CA311을 제외한 나머지 5개 네트워크에서 변별력이 최근접이웃중심성과 같은 수준으로 떨어진다. 파라미터 α 가 2.5 이상일 때에도 저자동시인용 네트워크 AC50과 상호대차 네트워크 IL31에서는 변별력이 최근접이웃중심성과 같은 수준으로 낮게 나타난다. 따라서 파라미터 α 가 2.0 이하일 때에만 6개 네트워크 모두에서 이웃중심성의 변별력이 최근접이웃중심성보다 높은 것으로 나타났다.

이상의 분석 결과를 정리해보면 첫째 조건에 따라 이웃중심성 지수를 측정할 때 이웃중심성보다 최근접이웃중심성에 더 가깝게 하려면 파라미터 α 를 1.5 이상으로 설정해야 하고, 둘째 조건에 따라 최근접이웃중심성보다 변별력이 높게 하려면 파라미터 α 를 2.0 이하로 설정해야 하는 것으로 나타났다. 따라서 종합적으로 파라미터 α 의 적정 수준은 1.5에서 2.0 사이라는 결론을 얻을 수 있다. 동일한 조건이라면 소숫점 이하 자리를 사용하지 않고 $\alpha=2$ 로 설정하는 것이 무난하다고 판단된다.

5. 결론

점차 확산되고 있는 가중 네트워크를 위한 일반화된 지역중심성 지수로 이웃중심성 지수

를 개발하였다. 제안된 지수는 방향성, 비방향성 여부에 상관없이 적용할 수 있고, 이진 네트워크와 가중 네트워크를 모두 포괄하는 확장성을 가지고 있다. 또한 순위를 이용하여 측정하는 지수이므로 대상이 되는 네트워크의 특성이나 링크 강도의 수준에 영향을 받지 않는다는 강점이 있다. 어떤 종류의 네트워크에도 적용할 수 있으며, 값이 매우 크거나 매우 작은 링크 강도라도 차별화하여 측정할 수 있는 것이다.

이웃중심성 지수를 6가지 실제 자료에 대해서 적용해본 결과, 가중 네트워크에서 지역중심성을 측정하면서도 높은 변별력을 보이는 것으로 나타났다. 링크 강도를 반영하는 수준을 결정하는 파라미터 α 는 일반적인 경우 2로 설정하면 가중 네트워크의 특성을 충분히 반영하면서도 변별력을 유지할 수 있는 것으로 판단된다.

이웃중심성 지수 산출을 위한 프로그램은 Python으로 구현한 독립실행 버전과 R의 함수 버전을 개발하였으며 두 가지 모두 공개할 예정이다.¹⁾ 이웃중심성 지수는 순위를 산출하여 계산하는 간단한 방식이므로 전용 프로그램을 사용하지 않더라도 작은 규모의 네트워크에서는 누구라도 Excel과 같은 표계산 소프트웨어를 사용해서 손쉽게 계산할 수 있을 것이다. 이를 통해 가중 네트워크 분석이 더욱 심화 발전할 것으로 기대된다.

1) 현재는 저자에게 이메일로 요청하면 제공 가능

참 고 문 헌

- 강범일, 박지홍 (2013). 프로파일링 분석과 동시출현단어 분석을 이용한 한국어교육학의 정체성 분석. *정보관리학회지*, 30(4), 195-213. <http://dx.doi.org/10.3743/KOSIM.2013.30.4.195>
- 강범일, 이재운 (2014). 트위터 관련 연구에 대한 계량정보학적 분석. *정보관리학회지*, 31(3), 293-311. <http://dx.doi.org/10.3743/KOSIM.2014.31.3.293>
- 김원진, 정영미 (2010). 과학기술분야 국제협력 증진을 위한 아시아 국가 간 공동연구 현황 분석. *정보관리학회지*, 27(3), 103-123. <http://dx.doi.org/10.3743/KOSIM.2010.27.3.103>
- 김관준 (2015). 디지털 큐레이션 연구동향 분석과 과제: 문헌정보학 분야를 중심으로. *정보관리학회지*, 32(1), 265-295. <http://dx.doi.org/10.3743/KOSIM.2015.32.1.265>
- 서선경, 정은경 (2013). 동시출현단어 분석 기반 오픈 액세스 분야 지적구조에 관한 연구. *한국비블리아학회지*, 24(1), 207-228.
- 유중덕 (2013). 경기도 공공도서관 상호대차 네트워크 분석. *정보관리학회지*, 30(2), 83-99. <http://dx.doi.org/10.3743/KOSIM.2013.30.2.083>
- 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. <http://dx.doi.org/10.4275/KSLIS.2006.40.3.191>
- 이재운 (2012). 폭소노미 연구 문헌에 대한 자아 중심 주제 인용 분석. *정보관리학회지*, 29(4), 295-312. <http://dx.doi.org/10.3743/KOSIM.2012.29.4.295>
- 이재운 (2013). tnet과 WNET의 가중 네트워크 중심성 지수 비교 연구. *정보관리학회지*, 30(4), 241-264. <http://dx.doi.org/10.3743/KOSIM.2013.30.4.241>
- 이재운 (2014). 공동연구 네트워크 분석을 위한 중심성 지수에 대한 비교 연구. *정보관리학회지*, 31(3), 153-179. <http://dx.doi.org/10.3743/KOSIM.2014.31.3.153>
- 조선례, 이재운 (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 (2014). 동시출현단어분석을 이용한 연관영화정보 분석 연구. *정보관리학회지*, 31(4), 161-178. <http://dx.doi.org/10.3743/KOSIM.2014.31.4.161>
- Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11), 3747-3752. <http://dx.doi.org/10.1073/pnas.0400087101>
- Fagiolo, G., Reyes, J., & Schiavo, S. (2010). The evolution of the world trade web: A weighted-network analysis. *Journal of Evolutionary Economics*, 20(4), 479-514.

<http://dx.doi.org/10.1007/s00191-009-0160-x>

Goldenberg, J., Libai, B., Muller, E., & Stremersch, S. (2010). Database submission: The evolving social network of marketing scholars. *Marketing Science*, 29(3), 561-567.

<http://dx.doi.org/10.1287/mksc.1090.0539>

Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, 70(5), 056131.

<http://dx.doi.org/10.1103/PhysRevE.70.056131>

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3), 245-251.

<http://dx.doi.org/10.1016/j.socnet.2010.03.006>

Shim, Jiyoung (2013). An investigation on using TV contents for scholarly purposes. Unpublished doctoral dissertation, Yonsei University, Seoul, Korea.

• 국문 참고문헌에 대한 영문 표기

(English translation of references written in Korean)

Choi, Sanghee (2014). An analysis of related movie information using the co-word method. *Journal of the Korean Society for Information Management*, 31(4), 161-178.

<http://dx.doi.org/10.3743/KOSIM.2014.31.4.161>

Jo, Seon-Rye, & Lee, Jae Yun (2012). Journal co-citation analysis for library services in pharmaceuticals. *Journal of information management*, 43(1), 159-185.

<http://dx.doi.org/10.1633/JIM.2012.43.1.159>

Kang, Beomil, & Lee, Jae Yun (2014). A bibliometric analysis on Twitter research. *Journal of the Korean Society for Information Management*, 31(3), 293-311.

<http://dx.doi.org/10.3743/KOSIM.2014.31.3.293>

Kang, Beomil, & Park, Ji-Hong (2013). Profiling and co-word analysis of teaching Korean as a foreign language domain. *Journal of the Korean Society for Information Management*,

30(4), 195-213. <http://dx.doi.org/10.3743/KOSIM.2013.30.4.195>

Kim, Pan Jun (2015). An analytical study on research trends of digital curation: Focused on library and information science. *Journal of the Korean Society for Information Management*,

32(1), 265-295. <http://dx.doi.org/10.3743/KOSIM.2015.32.1.265>

Kim, Won-Jin, & Chung, Young-Mee (2010). A study on research collaboration among Asian countries in science and technology. *Journal of the Korean Society for Information Management*,

27(3), 103-123. <http://dx.doi.org/10.3743/KOSIM.2010.27.3.103>

- Lee, Eun Sook, & Chung, Young Mee (2002). A co-citation analysis of multiple authorship in the subject field of information science and computer science. *Journal of Knowledge Processing and Management*, 3(2), 1-26.
- 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>
- Lee, Jae Yun (2012). Ego-centered topic citation analysis on folksonomy research documents. *Journal of the Korean Society for Information Management*, 29(4), 295-312.
<http://dx.doi.org/10.3743/KOSIM.2012.29.4.295>
- 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://dx.doi.org/10.3743/KOSIM.2013.30.4.241>
- Lee, Jae Yun (2014). A comparative study on the centrality measures for analyzing research collaboration networks. *Journal of the Korean Society for Information Management*, 31(3), 153-179, <http://dx.doi.org/10.3743/KOSIM.2014.31.3.153>
- Ryoo, Jong-Duk (2013). An analysis on interlibrary loan network of public libraries in Gyeonggi Province. *Journal of the Korean Society for Information Management*, 30(2), 83-99.
<http://dx.doi.org/10.3743/KOSIM.2013.30.2.083>
- Seo, SunKyung, & Chung, EunKyung (2013). Domain analysis on the field of open access by co-word analysis. *Journal of the Korean Biblia Society for Library and Information Science*, 24(1), 207-228.