

연구자의 투고 학술지 현황에 근거한 국내 학문분야 네트워크 분석*

Analyzing the Network of Academic Disciplines with Journal Contributions of Korean Researchers

이재윤(Jae Yun Lee)**

초 록

이 연구는 국내 연구자의 학술지 논문 발표 자료를 활용하여 학문분야간 학술지 공유도를 산출하고, 이로부터 국내 학문분야의 구조를 나타내는 네트워크를 생성하였다. 생성된 패스파인더 네트워크는 '생물학' 분야를 핵심으로 하는 생명과학 분야가 중양을 차지하고 있었으며, 인문학과 의약학, 공학에 속한 학문끼리는 학문간 연계가 매우 강하게 나타났다. 가중 네트워크로부터 각 학문분야의 중심성과 학제성을 파악하기 위해서 엔트로피 공식과 가중 네트워크 중심성 척도를 적용한 결과 전역 중심 학문, 지역 중심 학문, 전역 연계 학문, 기타 일반 학문의 네 가지 유형을 식별할 수 있었다. 가중 네트워크를 이진 네트워크로 변환한 패스파인더 네트워크에서는 다수의 약한 링크가 모인 데이트 허브가 드러나지 않았으나, 가중 네트워크에서의 중심성 지수인 삼각매개중심성의 측정 범위를 지역에서부터 전역까지 달리하며 측정된 결과로부터 '인지과학'분야와 같은 학제성이 높은 데이트 허브를 식별할 수 있었다.

ABSTRACT

The main purposes of this study are to construct a Korean science network from journal contributions data of Korean researchers, and to analyze the structure and characteristics of the network. First of all, the association matrix of 140 scholarly domains are calculated based on the number of contributions in common journals, and then the Pathfinder network algorithm is applied to those matrix. The resulting network has several hubs such as 'Biology', 'Korean Language & Linguistics', 'Physics', etc. The entropy formula and several centrality measures for the weighted networks are adopted to identify the centralities and interdisciplinarity of each scholarly domain. In particular, the date hubs, which have several weak links, are successively distinguished by local and global triangle betweenness centrality measures.

키워드: 계량서지학, 학문 네트워크, 과학지도학, 가중 네트워크, 중심성 척도, 데이트 허브
bibliometrics, science network, scientography, weighted network, centrality
measures, date hub

* 이 논문은 제3회 복잡계 컨퍼런스(2008년 11월 29일, 연세대학교 새천년관)에서 발표한 것을 수정·보완한 것임.

이 논문에서 사용한 자료는 한국학술진흥재단 정책과제의 지원을 받은 것임.

** 경기대학교 문헌정보학과 조교수(memexlee@kgu.ac.kr)

■ 논문접수일자: 2008년 11월 19일 ■ 최초심사일자: 2008년 11월 20일 ■ 게재확정일자: 2008년 12월 5일
■ 情報管理學會誌, 25(4): 327-345, 2008. [DOI:10.3743/KOSIM.2008.25.4.327]

1. 서론

전통적으로 계량서지학 분야에서는 특정 학문 분야의 연구 동향을 파악하기 위해서 인용분석을 주요 수단으로 활용해왔다. 학문 연구가 복잡해지고 개별 학문분야가 심층적으로 분화됨에 따라서 이와 같은 개별 분야 연구동향에 대한 계량서지학 연구는 더욱 필요성을 높여왔다. 그런데 학제적인 연구나 융합 학문이라는 표현이 등장하고 심지어 통섭과 같은 용어까지 등장하기에 이른 최근의 학술활동 경향은, 여러 학문분야의 학술활동을 개별적으로 파악하는 것에서 벗어나 전체 학문분야의 활동과 연계 상황 등을 종합적으로 파악할 필요가 있다.

이와 같은 전체 학문 분야 연구 동향의 분석은 오래전 Price(1966)에 의해 제안된 바 있으나, 구체적으로 이를 실현한 사람은 Small과 Garfield이다. 이들은 계량서지적 정보로부터 지도를 만드는 것을 scientography라고 부르기까지 하였다(Garfield 1994; Small & Garfield 1985). 최근에는 스페인의 Vargas-Quesada와 Moya-Anegón, 미국 인디애나 대학의 Boyack 등이 학문분야 사이의 인용 네트워크 구조를 분석하여 활발히 연구를 진행하고 있다.

이와 같은 광역 학문 지도(global map of science), 혹은 광역 학문 네트워크는 전체 학문 분야 사이의 구조적인 관계를 시각화하여 분석하는데 유용하며, 특히 네트워크 구조를 분석하기에 도움이 되는 도구이다. Small과 Price의 연구를 비롯하여 최근까지 학문 네트워크와 관련된 모든 연구는 인용을 근거로 수행되고 있다(이재운 2007).

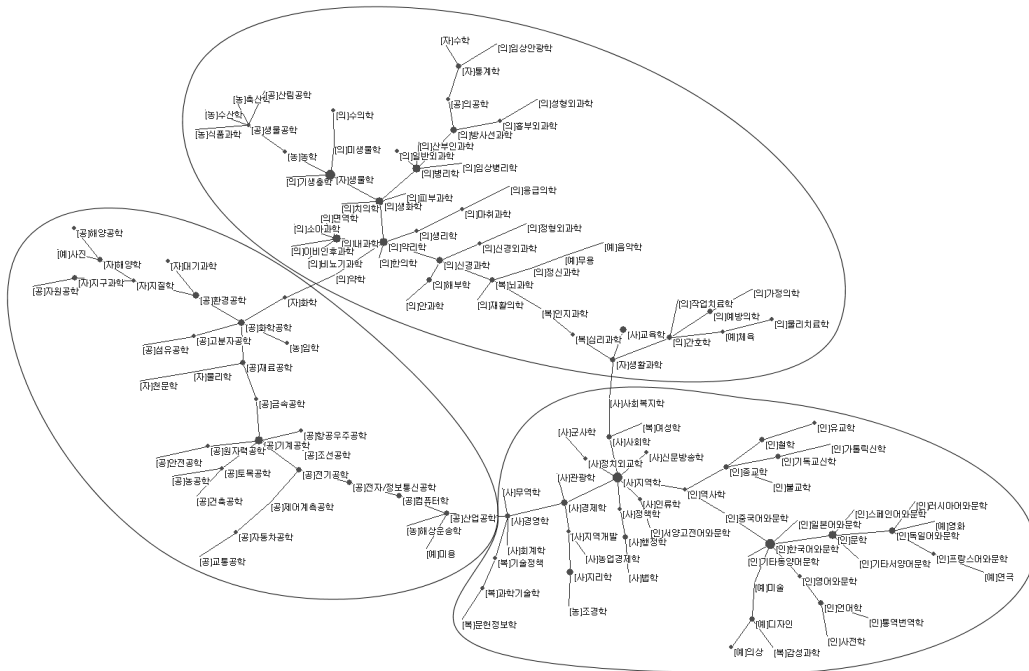
한편 국내 연구 동향에 대한 학문 네트워크

를 생성하는데 있어서 가장 큰 걸림돌은 국내 연구 성과물에 대한 인용 데이터베이스가 완성되어 서비스되고 있지 않은 점이다. 개별 학문 분야에 대한 분석을 위해 연구자가 개인적으로 인용 자료를 구축하는 것은 제한적으로는 가능하다. 기존의 국내 연구동향에 대한 인용분석 연구에서는 대부분 연구자가 직접 인용 자료를 수집하여 데이터베이스화하고 분석하였다. 그러나 수십 개 이상의 학문 분야를 다루는 광역 학문 지도를 생성하기 위한 인용 데이터베이스를 개인적 차원에서 구축하기란 거의 불가능하다 할 수 있다.

최근 한국학술진흥재단에서 인용 데이터베이스를 구축하고 있으나, 구축된 데이터의 수록기간이 짧고 대상 학술지가 국내 학술지로 제한되며 인용정보가 부정확하게 구축되고 있다는 점에서 광역 학문 네트워크를 구축하여 분석하기에는 미흡한 점이 많다. 특히 국내 이공계 연구자는 연구 성과의 상당한 부분을 외국 학술지에 발표하고 있다는 점을 감안하면 국내 연구활동 전반의 네트워크를 분석하기 위한 수단으로 국내 학술지만을 대상으로 하는 것은 충분하지 못하다.

이런 한계점을 극복하고 국내 학문 연구 동향을 파악하기 위해 이재운(2007)은 한국학술진흥재단에 제출된 과제신청서의 요약문으로부터 추출한 색인어를 기반으로 학문간 텍스트 유사성을 측정하고 국내 학문 네트워크를 생성한 바 있다(그림 1 참조). 그러나 이와 같은 텍스트 유사성에 근거한 방식은 각 학문분야에서 사용하는 어휘가 유사한 경우를 찾는 것이므로 학문간의 실제적인 연계를 나타낸다고 보기는 어렵다.

따라서 이 연구에서는 국내 학술활동을 반영



〈그림 1〉 국내 학문 영역간 텍스트 유사성 네트워크(이재윤 2007)

하는 학술지 논문 게재 자료를 활용하는 시도를 하였다. 이를 통해서 학문분야간 연계 네트워크를 구축하고 이를 시각화하여 살펴본 후 중심성 분석을 통해서 각 학문분야의 학제적 수준을 파악해보았다. 네트워크 생성과 중심성 측정은 직접 작성한 프로그램으로 수행하였고 시각화를 위해서는 Pajek(de Nooy et al. 2005)을 사용하였다.

2. 학문분야간 논문 발표 학술지 연계성 분석

2.1 학문분야간 학술지 공유도 산출

이 연구에서 학문간의 연계 정보로 활용한

것은 각 분야의 연구자가 논문을 발표하는 학술지의 일치 정도이다. 즉 어떤 두 학문 분야의 연구자들이 투고하는 학술지가 상당히 일치한다면 두 학문 분야는 학술적으로 연결되어 있을 가능성이 높다고 간주하는 것이다.

국내 연구자의 소속 학문 분야 정보와 논문 게재 학술지 정보를 완벽하게 파악하는 것은 불가능하지만, 한국학술진흥재단의 연구인력정보인 KRI DB에는 상당한 분량의 데이터가 축적되어 있다. 이 연구에서는 KRI DB의 데이터를 활용하여 국내 학문 분야 사이의 연계 네트워크를 구축하여 분석을 수행하였다.

분석 대상은 KRI DB에 2008년 2월 19일까지 입력된 각 연구자의 KCI 등재/등재후보지 및 SCI급 논문 발표실적이며 전체 입력 자료의 수는 468,163건이다. 이 중에는 공저자의 논문

발표실적이 개별적으로 입력됨에 따라서 중복된 사례가 다수 있으므로 실제 논문의 수는 이보다 상당히 적을 것으로 추정된다.

KRI DB에 나타난 학술지는 총 17,594종이었으며 이 중에서 KRI DB에 입력된 건수가 3회 이상인 학술지는 15,782종이었다. 동일한 학술지를 상이한 표현으로 입력한 경우가 많았으나 2회 이상 나타난 학술지의 명칭은 모두 수작업으로 통일시켰다.

한국학술진흥재단의 학문분야분류표는 여덟 범주로 나뉘는 대분류 구분에서부터 시작하여 중분류, 소분류, 세분류의 네 단계로 구성된 계층적 분류표이다. 세분류 항목의 수는 분류코드가 A000000인 인문학(대분류, 중분류, 소분류, 세분류 모두 인문학)에서부터 분류코드가 H990000인 학제간연구(대분류는 복합학, 중분류, 소분류, 세분류 모두 학제간연구)에 이르기까지 총 3,374범주에 달한다. 이 연구에서는 ‘물리학’, ‘경제학’ 등과 같이 일반적으로 구분하는 학문분야에 해당하는 중분류 수준을 분석 대상으로 하였다. 원래 중분류 수준 항목은 159개이지만 이 중에서 ‘인문학’, ‘사회과학’처럼 대분류 명칭과 같은 항목이나 ‘기타의약학’과 같은 ‘기타’가 포함된 항목, 그리고 ‘예술일반’과 같은 ‘일반’이 포함된 항목은 주제 범위가 특정적이지 않으므로 제외하고 나머지 140개 항목만을 분석대상 학문분야로 삼았다.

각 학술지에 대한 저자의 소속 학문분야는 KRI DB에 입력된 연구자의 소속분야 필드와 심사가능분야 필드의 중분류 수준을 근거로 하

여 판정하였다. 만약 한 연구자에 대해서 복수의 중분류가 지정되는 경우에는 모두를 인정하여 한 차례의 논문 발표 행위가 연구자와 관련된 각 분야에 모두 반영되도록 하였다. 예를 들어 *Nature*에 논문을 한 편 게재한 어떤 연구자의 소속분야와 심사가능분야에 나타난 중분류 수준 항목이 ‘물리학’과 ‘재료공학’의 두 가지였다면, 두 분야에서 모두 *Nature*에 논문을 한 편씩 발표한 것으로 처리하였다.

이와 같은 연구자의 학술지 논문 발표 데이터로부터 학문분야간 관계를 파악하기 위해서 학문분야간 학술지 공유도를 산출하였다. 학문분야 d_a 와 학문분야 d_b 사이의 학술지 공유도 $COMJ(d_a, d_b)$ 를 구하는 공식은 다음과 같이 고안하였다

$$COMJ(d_a, d_b) = \sum_{i=1}^n \min(f(d_a, j_i), f(d_b, j_i))$$

- n: 전체 학술지 종수
- d_a : 학문분야 a
- j_i : 학술지 i
- $f(d_a, j_i)$: 학문분야 a의 연구자가 학술지 i에 투고한 건수

이 공식은 널리 쓰이는 거리 함수인 도시블록 거리(또는 맨해튼 거리) 공식을 거리가 아닌 유사성을 나타내도록 변환한 것이다.¹⁾ 다시 말하면 학문분야간 학술지 공유도 COMJ는, 두 학문분야의 연구자가 동일한 학술지에 각각 발표한 논문 건수 중에서 작은 값을 취한 후 이를

1) 도시블록 거리 공식은 $\sum |x_i - y_i| = \sum \{\max(x_i, y_i) - \min(x_i, y_i)\}$ 이다. 흔히 거리 공식을 유사도 공식으로 변환하기 위해서는 최댓값에서 빼는 방식을 취하므로, 최댓값인 $\sum \max(x_i, y_i)$ 에서 $\sum \{\max(x_i, y_i) - \min(x_i, y_i)\}$ 를 빼면 $\sum \min(x_i, y_i)$ 만이 남는다.

모든 학술지에 대해 구하여 합산한 값이다. 예를 들어 <그림 2>에서 물리학과 경제학 분야의 학술지 공유도 COMJ(물리학, 경제학)은 각 학술지에 대해서 두 분야의 연구자들이 게재한 횟수의 최저값인 0, 1, 2, 0을 합한 3이 된다.

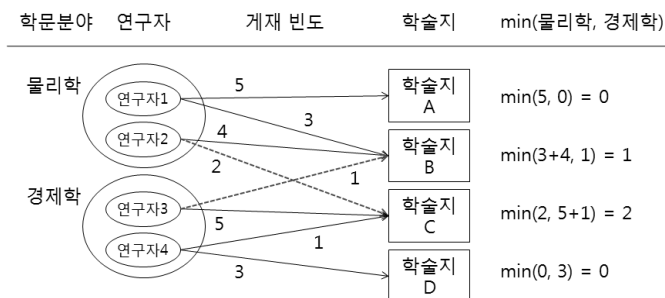
140개 학문분야 사이의 학술지 공유도를 모두 산출하면 크기가 140×140인 연관성 행렬이 도출된다(표 1 참조). 행렬에서 가장 공유도가 높게 나타난 두 학문분야는 ‘물리학’과 ‘재료공학’이었으며 그 다음은 ‘물리학’과 ‘전자/정보통신공학’이었다.

2.2 학술지 공유도 패스파인더 네트워크 생성

학술지 공유도를 산출한 결과 140개 학문분야 사이에 가능한 9,730개의 관계 중에서 학술지 공유도가 0보다 커서 실제로 링크가 형성된 경우는 8,946개로 91.9%에 달하였다. 이처럼 대부분의 칸이 0이 아닌 가중치 행렬이므로 이를 그대로 네트워크로 표현하여 분석하기에는 적절하지 않다. 0이 아닌 값을 링크로 표현할 경우에는 육안으로 식별하기 어렵게 얽히고 덩어리진 그림만이 생성될 뿐이다.

가중치 행렬로부터 일부 링크를 선택하거나 제거하여 시각적으로 표현하기 위한 네트워크를 만들기 위한 알고리즘으로는 여러 가지가 제시되어 있는데(이재윤 2006a), 이 연구에서는 전체 노드가 하나의 네트워크를 형성하는 것이 보장되는 패스파인더 네트워크(PFNet: Schvaneveldt 1990) 알고리즘을 사용하였다. 이 연구에서 구현한 PFNet은 가장 엄격한 조건인 $r=n-1$, $p=\infty$ 으로 설정하였으므로 연결강도가 동등한 링크가 없을 경우에는 MST(minimum spanning tree)와 같다.

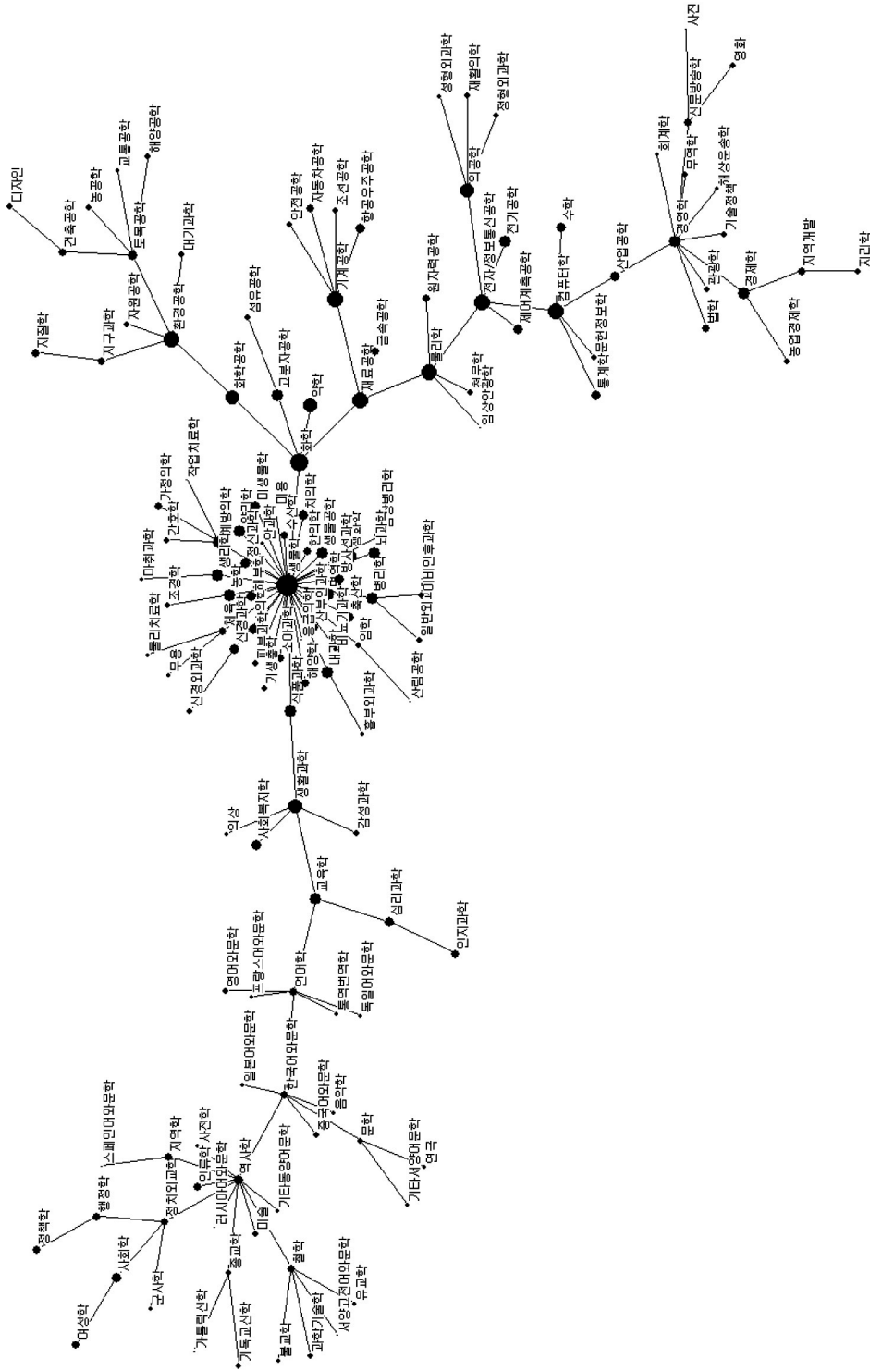
140개 학문분야간 학술지 공유도 행렬로부터 패스파인더 네트워크를 생성하면 <그림 3>과 같다. 이 그림에서 노드의 크기는 각 학문분야 연구자들이 논문을 발표한 학술지의 종수에 비례하도록 표현하였다. 즉 노드 크기가 클수록 해당 학문 연구자들이 많은 학술지에 논문을 발표하는 분야이다. <그림 3>에서는 ‘생물학’ 분야가 여러 의학의 세부 분야를 비롯한 다양한 생명과학 관련 분야의 중심 허브로 확연히 나타나고 있으며, 왼쪽의 ‘역사학’, ‘한국어와문학’, 오른쪽의 ‘환경공학’, ‘기계공학’, ‘전자/정보통신공학’, ‘경영학’ 등이 주요 허브 노드가 되었다.



<그림 2> 학문분야간 학술지 공유도 산출의 예

〈표 1〉 학문분야간 학술지 공유도 행렬(140 × 140 행렬 중에서 일부)

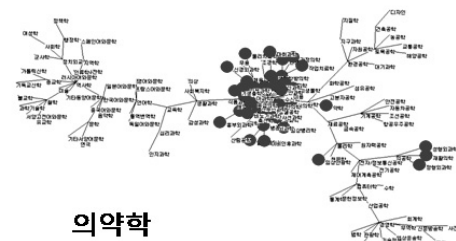
사건화	역사화	철학	종교학	기독교 신학	기독교 신학	불교학	유교학	불교학	언어학	문학	한국어 외문학	중국어 외문학	일본어 외문학	기타문학 어문학	영어와 문학	프랑스어 외문학	독일어 외문학	스페인어 외문학	러시아어 외문학	서양고전 어문학	기타서양 어문학	통역번역학
295	237	214	150	33	6	53	78	235	177	218	119	56	67	100	94	94	38	74	19	78	133	
237	26252	3786	2264	608	189	728	926	2146	2445	4073	1486	1163	602	615	496	386	268	539	239	352	502	
214	3786	21964	2147	770	199	1369	1318	1414	1663	2676	1104	522	288	589	478	631	107	193	285	154	448	
150	2264	2147	3977	1044	206	575	807	715	689	854	425	198	496	245	165	125	112	78	200	112	223	
33	608	770	1044	2426	177	203	221	93	139	152	84	44	23	80	58	52	6	39	127	39	34	
6	189	199	206	177	232	41	86	27	38	42	23	8	9	26	16	12	10	7	66	20	15	
53	728	1369	575	203	41	1425	333	122	253	521	306	123	36	93	65	71	7	18	79	33	37	
78	926	1318	807	221	86	333	2040	166	208	387	271	94	80	87	42	51	17	22	79	47	48	
235	2146	1414	715	93	27	122	166	13777	2652	5013	694	909	527	4810	1136	1608	343	526	90	344	608	
177	2445	1663	689	139	38	253	208	2652	8356	6310	786	688	342	990	526	935	329	512	81	380	483	
218	4073	2676	854	152	42	521	387	5013	6310	34335	1680	1201	277	2254	595	909	163	247	87	221	468	
119	1486	1104	425	84	23	306	271	694	786	1680	12122	362	134	338	153	264	106	135	43	89	214	
56	1163	522	198	44	8	123	94	909	688	1201	362	9575	128	309	176	204	86	129	30	87	153	
67	602	298	496	23	9	36	80	527	342	277	134	128	734	122	110	92	102	57	25	71	104	
100	615	589	245	80	26	93	87	4810	990	2254	338	309	122	13160	439	594	181	208	77	136	348	
94	496	478	165	58	16	65	42	1136	526	595	153	176	110	439	3243	247	90	136	43	116	181	
94	386	631	125	52	12	71	51	1608	935	909	264	204	92	594	247	7130	142	205	34	106	131	
38	268	107	112	6	10	7	17	343	329	163	106	86	102	181	90	142	1202	79	31	75	100	
74	539	193	78	39	7	18	22	526	512	247	135	129	57	208	136	205	79	2360	22	201	105	
19	239	295	200	127	66	79	79	90	81	87	43	30	25	77	43	34	31	22	353	64	45	
78	352	154	112	39	20	33	47	344	380	221	89	87	71	136	116	106	75	201	64	442	89	
133	502	448	223	34	15	37	48	608	483	468	214	153	104	348	181	131	100	105	45	89	808	
137	3846	1871	1216	298	58	542	405	976	801	1025	504	386	361	299	186	136	196	390	153	122	254	
82	1848	751	483	145	34	215	195	460	441	414	331	126	138	129	114	89	94	164	82	59	151	
13	152	72	45	12	3	24	18	32	40	37	26	30	12	17	26	18	7	2	3	12	8	
55	957	526	236	119	21	116	99	402	229	243	120	114	45	109	124	59	13	45	16	36	101	
6	83	42	13	8	0	6	10	23	18	12	10	16	7	4	10	7	5	2	0	6	3	
38	417	121	107	17	3	31	29	90	70	53	43	37	27	20	45	24	13	19	4	21	40	
117	3044	1692	981	377	98	556	411	686	744	1037	375	301	144	245	209	192	80	108	139	116	224	
33	773	536	280	181	44	144	79	315	229	236	115	93	30	122	76	72	15	24	44	23	76	
183	3311	1304	1127	183	84	251	344	1427	1282	1035	526	436	534	336	313	287	405	375	95	268	401	
161	2246	1089	774	160	74	243	237	988	1179	1255	389	367	249	391	295	245	134	151	81	167	382	
188	1777	2323	829	735	130	516	380	3384	1371	3042	531	410	158	1934	445	430	86	146	138	133	333	
47	845	515	285	117	36	158	127	189	204	334	109	57	31	106	115	69	6	21	56	36	55	
31	585	342	164	77	16	131	74	117	105	134	81	52	29	75	45	13	16	16	8	12	24	
39	602	397	141	79	17	124	60	156	117	121	67	55	21	68	83	53	9	24	5	16	46	
42	563	202	162	33	11	66	118	139	116	174	92	70	35	34	38	18	20	13	22	26	26	



〈그림 3〉 학술지 공유도에 근거한 국내 학문분야 패스파인더 네트워크

이와 같은 학문분야간 연계성이 대분류별로 어떻게 나타나는가를 알아보기 위해서 <그림 3>

의 각 학문분야를 소속 대분류별로 PFnet에 따로 표시한 것이 <그림 4>이다.



<그림 4> 국내 학문분야 패스파인더 네트워크의 8개 대분류 범주별 구분

〈그림 4〉를 보면 한국학술진흥재단 학문분류표를 기준으로 하였을 때 '인문학'과 '의약학', '공학'에 속한 학문끼리는 학술지 공유를 통한 학문간 연계가 매우 강한 것을 알 수 있으며, 이보다는 덜하지만 '사회과학'에 속한 학문들도 크게 두 집단으로 구분되어 상호 연계가 이루어지고 있는 것으로 보인다. 반면에 '예술체육'이나 '복합학'에 속한 학문들은 네트워크의 여러 곳에 산재되어 있어서 학술지 공유도를 통한 상호연계가 상당히 저조하다는 것을 알 수 있다. 또한 흠어짐의 정도가 이보다는 덜하지만 '자연과학'에 속한 학문분야들도 '자연과학' 내에서의 상호연계보다는 '의약학'이나 '공학'에 속한 학문들과의 연계가 더욱 활발한 것으로 나타났다.

3. 엔트로피로 측정된 학문분야별 학제성

앞에서 패스파인더 네트워크로 표현된 학문분야간 학술지공유 지도는 기대에 비해서 각 학문분야의 학제성이나 타 학문과의 연계 정도를 충분히 표현해주지 못한다. 이를 더면 대표적인 학제적인 분야로 흔히 언급되는 '인지과학' 분야는 〈그림 3〉의 네트워크의 왼쪽 중간 아래 가지 끝에 붙어있으므로 다양한 학문분야와의 연계를 전혀 보여주지 못하고 있다. 그러나 또다른 학제적인 분야인 '환경공학'은 〈그림 3〉의 오른쪽 위에 허브 노드로 위치하면서 '지구과학', '자원공학', '토목공학', '대기과학', '화학공학'과 같은 학문들과 연계된 학제적인 분야임이 시각적으로 표현된다.

이와 같이 학술지 공유도에 근거한 학문분야 패스파인더 네트워크에서 학제성이 시각적으로 충분히 표현되지 못하는 이유는, 패스파인더 네트워크가 상대적으로 강한 링크만을 살려서 지도를 표현하기 때문에 약한 링크가 생략되는 것이 주된 원인이다. 널리 알려져 있듯이 약한 링크가 네트워크의 안정성과 개별 노드에 미치는 영향은 무시할 수 없다(Csermely 2006; Granovetter 1973). 이는 Mantegna(1999)의 연구와 같이 연관성 행렬로부터 MST를 도출하여 분석하는 여타 연구에서도 나타날 수 있는 한계점이다.

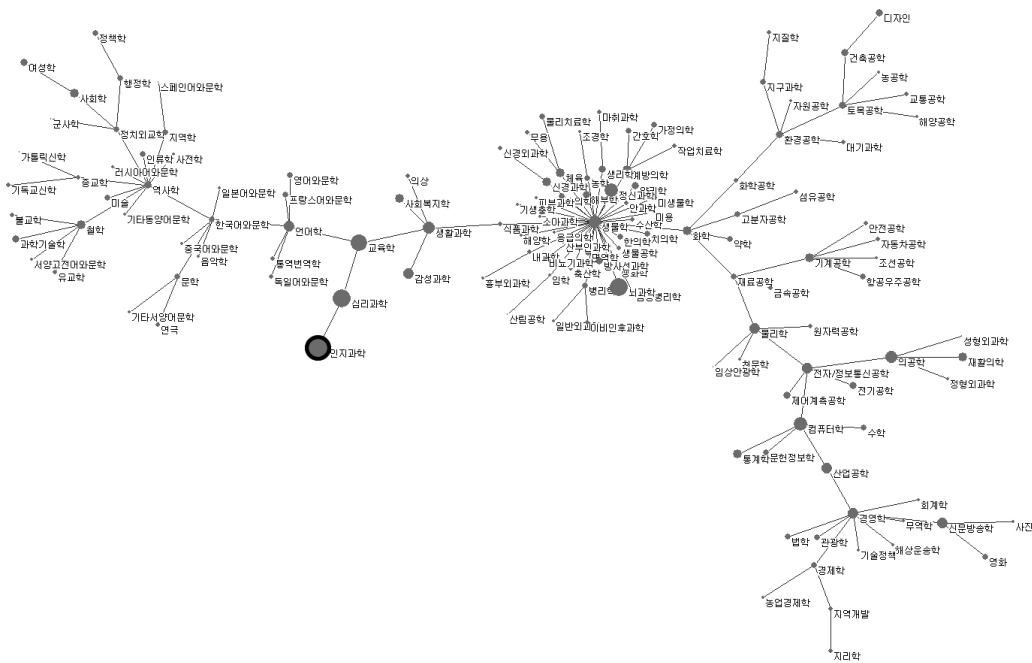
개별 학문분야의 학제성이 얼마나 되는가를 알아보기 위해서는 다른 학문분야와 어느 정도로 연계되어 있는가를 살펴보는 것이 한 방법이다. 이를 위해서 특정 학문분야가 다른 각 학문분야와 가지는 학술지 공유도를 확률값으로 변환한 후 이의 엔트로피를 산출하였다. 종종 다양성 지표로도 사용되는 엔트로피는 확률값이 고르게 분포되어 있을수록 높게 나타나며, 여러 학문분야와 학술지 공유도가 고르게 높은 분야는 엔트로피가 높은 학제적인 학문일 가능성이 높다. 학술지 공유도 엔트로피가 높은 상위 10개 학문분야와 엔트로피가 낮은 하위 10개 학문분야는 〈표 2〉와 같다.

학술지 공유도 엔트로피가 높아서 학제성도 높다고 추정되는 '인지과학' 분야와의 학술지 공유도를 개별 학문분야별로 산출하여 패스파인더 네트워크에 표현하면 〈그림 5〉와 같다. 이 그림에서 네트워크 전반에 걸쳐 연계되는 학문이 존재하므로 '인지과학' 분야는 포괄적인 학제성을 가지고 있는 것으로 해석할 수 있다.

반면에 학술지 공유도 엔트로피가 낮은 '물리

〈표 2〉 학술지 공유도 엔트로피

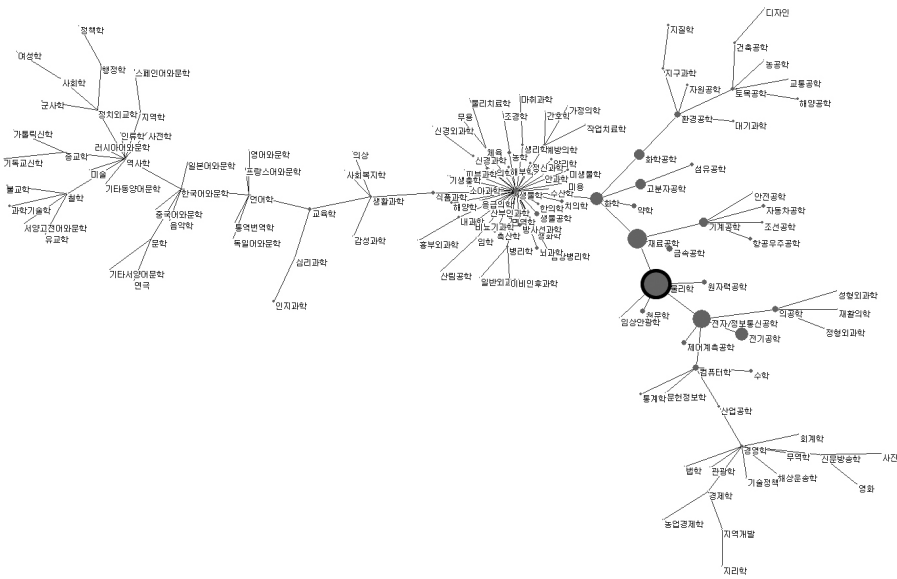
순위	학술지 공유도 엔트로피	분야명
1	6,349	과학기술학
2	6,326	인지과학
3	6,201	감성과학
4	6,189	미용
5	6,138	기술정책
6	6,067	임상안광학
7	6,036	뇌과학
8	5,938	안전공학
9	5,900	의공학
10	5,885	작업치료학
∴	∴	∴
131	4,393	러시아어외문학
132	4,293	전자/정보통신공학
133	4,209	법학
134	4,159	재료공학
135	4,039	독일어외문학
136	3,951	영어외문학
137	3,951	중국어외문학
138	3,875	물리학
139	3,865	한국어외문학
140	3,808	일본어외문학



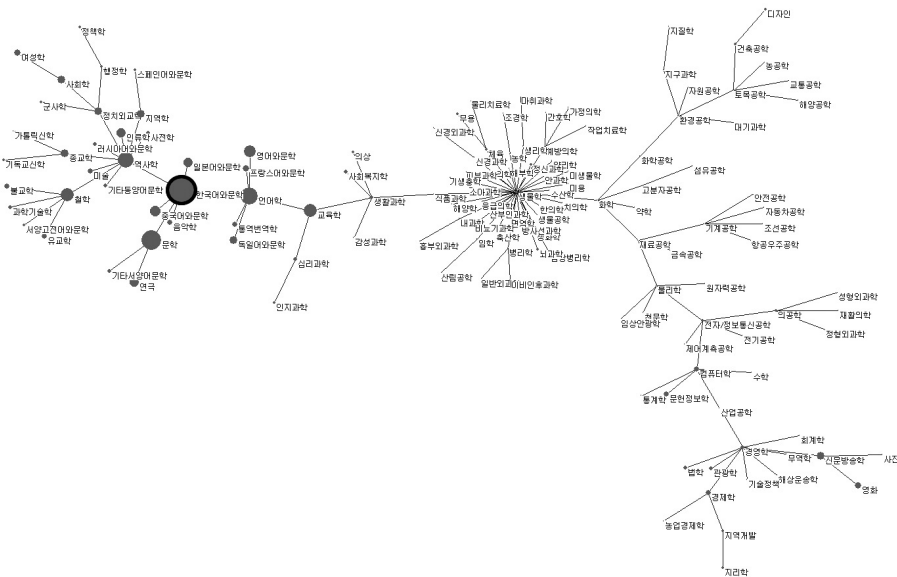
〈그림 5〉 패스파인더 네트워크에 표현한 '인지과학' 분야의 학술지공유도 표현

학'이나 '한국어외문학' 분야와의 학술지 공유도를 개별 학문분야별로 산출하여 패스파인더 네트워크에 표현하면 <그림 6>이나 <그림 7>처럼

주변 일부 학문 노드만 크게 나타나므로, 상대적으로 학제성이 낮거나 국지적인 학제성만 가지고 있음을 시각적으로 확인할 수 있다.



<그림 6> 패스파인더 네트워크에 표현한 '물리학' 분야와의 학술지공유도



<그림 7> 패스파인더 네트워크에 표현한 '한국어외문학' 분야와의 학술지공유도

4. 학문분야 중심성 분석

4.1 학술지 공유도 행렬의 중심성 분석

어떤 학문분야가 여러 학문들 가운데 중심성이 높다는 것은 여러 측면으로 해석할 수 있다. 특히 학술지를 공유하는 정도로 측정된 결과에서 중심성이 높은 경우는, 앞서 살펴본 '생물학' 분야처럼 생명과학 분야 전반에 걸쳐서 폭넓게 연계되어 있거나 '물리학'이나 '한국어와문학'처럼 제한된 주변 학문에 대해서 지배적인 결속력을 발휘하는 학문인 경우가 많다. 반면에 '인지과학'처럼 주변 학문과의 연계가 뚜렷하게 강하지는 않으나 학제적인 성격으로 인하여 다양한 학문과 광범위하게 연계되는 경우도 나타난다. 이처럼 다른 노드와 폭넓게 연결되는 노드나 주변 노드와의 연결이 집중되어 있는 노드를 허브 노드라고 부른다. 특히 Han 등(2004)의 주장처럼 허브를 파티 허브와 데이트 허브로 구분할 때 '생물학'과 같은 분야는 파티 허브, '인지과학'과 같은 분야는 데이트 허브라고 할 수 있다.

앞에서 학술지 공유도 엔트로피 척도로 살펴 보았듯이, 노드간 연관성 가중치 행렬에서 약한 링크를 삭제하고 패스파인더 네트워크나 MST 처럼 강한 링크 위주로 네트워크를 구성하면 다수의 정보를 잃어버리게 된다. 비록 강한 링크 위주로 표현된 네트워크에서는 허브가 되지 못하지만 중요한 노드일 가능성이 있는 것이다. 약한 링크를 주로 모아서 형성되는 데이트 허브(Csermely 2006)는 '인지과학' 분야처럼 패스파인더 네트워크에서 드러나지 않을 가능성이 더 높다.

패스파인더 네트워크나 MST와 같이 가중

네트워크를 이진 네트워크로 변경시키는 것은 가중치를 0 또는 1로 변환하는 것으로서 이 과정에서 어느 정도의 정보가 상실되는 것은 피할 수 없다. 이는 특히 원래의 가중치 행렬이 값이 0인 원소가 매우 적은 경우에 더욱 심하다. 예를 들어 Newman(2004)이 제안한 바와 같이 가중 네트워크를 이진 링크가 중복된 다중그래프로 변환한 후 대표적인 지역 중심성 척도인 연결정도중심성을 측정하면, 인접한 노드만이 아닌 매우 낮은 가중치로 연결된 원거리 노드까지 고려하게 되므로 사실상 전역 중심성을 반영하는 결과가 된다(이재윤 2006b).

따라서 이 논문에서는 패스파인더 네트워크로 변환되기 이전의 가중치 행렬을 그대로 가중 네트워크로 보고, 이 가중 네트워크에서의 중심성을 아래의 다섯 가지 척도로 측정해보았다.

① C_N (nearest centrality: 최근접중심성): 다른 노드에 의해 최근접이웃으로 꼽히는 비율이다(이재윤 2006b). 해당 노드를 최근접 이웃으로 꼽을 만큼 최대한 인접한 노드만을 대상으로 분석하므로 지역 중심성 척도에 해당한다.

② C_M (mean associations: 평균연관성): 다른 노드들과의 연관성 값의 평균이다(이재윤 2006b). 이는 다른 노드들과 전체적으로 얼마나 근접한 위치에 있는가를 측정하는 것이므로 전역 중심성 척도에 해당한다.

③ C_{MP} (mean profile associations: 평균프로파일연관성): 연관성 행렬로부터 노드간 피어슨상관계수를 도출한 후 다른 노드와의 상관계수의 평균을 취한 것이다(이재윤 2006b). 이는 평균연관성에 비해서 노드 자체의 영향력보다는 네트워크 내에서의 구조적인 위치에 더 좌우되는 지표이다. 실제 계산할 때에는 최저

값이 일정하지 않은 것을 피하고 0이 되도록 하기 위해서 산출된 C_{MP} 값 중 최저값을 일괄적으로 빼서 보정한다.

④ C_{TB} (triangle betweenness centrality: 삼각매개중심성): 한 노드가 다른 임의의 두 노드 사이를 결속시켜주는 능력의 비율이다. 연관성 행렬에서 분석 기준이 되는 노드를 a 라고 할 때 a 를 제외한 다른 노드들의 조합쌍 중에서 다음 조건을 만족하는 노드 쌍(b, c)의 비율로 측정한다(이재윤 2006b).

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

삼각매개중심성은 노드 셋으로 구성된 삼각형을 기본 분석 단위로 하며 측정 범위를 대상 노드와 연관성이 가장 높은 두 노드에서부터 시작하여 다른 모든 노드까지 넓힐 수 있으므로 지역 중심성과 전역 중심성을 측정하는데 모두 쓰일 수 있다. 별도로 지정하지 않으면 다른 모든 노드의 조합쌍을 대상으로 측정하므로 전역 중심성 지표가 된다.

⑤ C_{TB-10} (최근접 10노드 삼각매개중심성): 삼각매개중심성을 분석 대상 노드와 연관성이 가장 높은 10개 노드 사이의 관계에 대해서만 측정하는 것으로서 지역 중심성 지표가 된다.

이상과 같은 여섯 가지 척도를 학술지 공유도 행렬에 적용하여 산출한 학문분야의 중심성은 <표 3>과 같다.

전역 중심성을 측정하는 C_{TB} 와 C_M 으로는 '생물학' 분야가 가장 중심성이 높으나, 구조적인 위치가 많이 반영되는 전역 중심성 척도인 C_{MP} 로는 생물학이나 생화학처럼 다른 전역 중

심성 척도값이 높은 분야 이외에 '미생물학', '약리학'과 같이 위치가 전체 네트워크의 가운데에 가까워서 중심성이 높게 나타나는 분야도 있다. 대체적으로 전역 중심성 지표값이 높게 나타나는 노드들은 패스파인더 네트워크의 중심과 가깝고 연결 정도가 조밀한 의약학이나 공학에 속한 분야가 많다.

반면 지역 중심성을 측정하는 C_N 과 C_{TB-10} 으로는 '생물학' 분야만 아니라 '경영학', '역사학', '한국어외문학'과 같이 패스파인더 네트워크의 양쪽 끝에 가까우면서도 링크가 많이 연결된 분야들이 상위를 차지한다.

4.2 중심성 표출 범위에 따른 학문 유형 분석

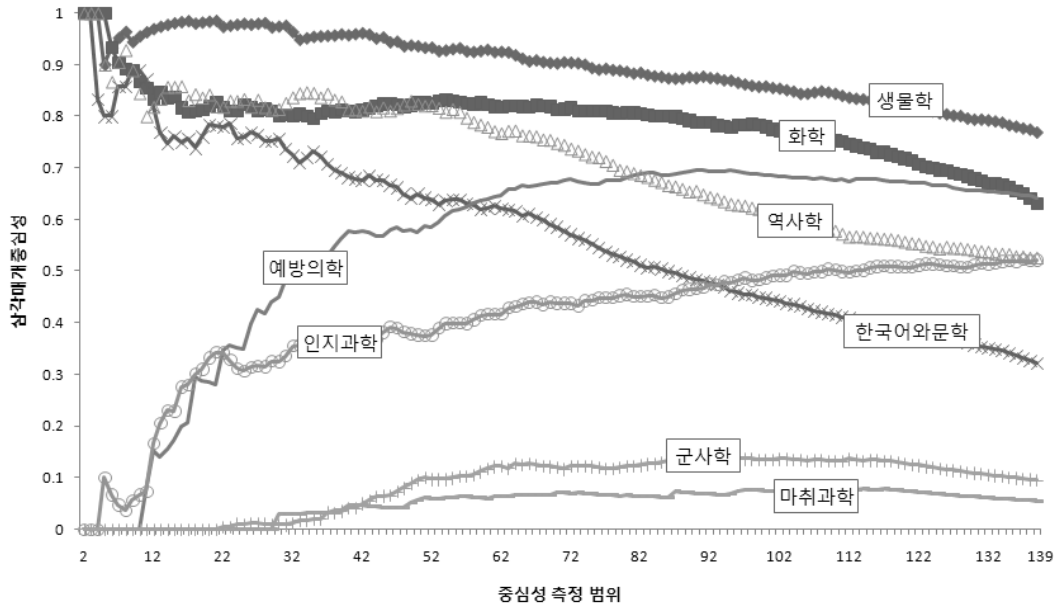
삼각매개중심성 C_{TB} 의 측정 범위를 주변 2개 노드에서부터 전체로 넓혀가면서 측정하여 지역 중심성으로부터 전역 중심성까지 각 노드의 중심성이 어떻게 변화하는가를 상세하게 살펴보면 네트워크를 구성하는 네 가지 노드 유형을 발견할 수 있다(이재윤 2006b). 학술지 공유도 네트워크에서 각 학문분야에 대해서 가장 지역적인 C_{TB-2} 에서부터 C_{TB-139} 까지 삼각매개중심성을 측정한 결과 중에서 주요 사례만 제시한 결과가 <그림 8>이다. <그림 8>로 식별할 수 있는 네 가지 학문(노드) 유형은 다음과 같다.

① 전역 중심 학문: 특정 분야에 대한 영향력이 매우 강력하여 국지적으로 뿐만 아니라 전체적으로도 중심 역할을 차지하는 학문이다. 측정 범위가 넓어지더라도 삼각매개중심성이 크게

〈표 3〉 학술지 공유도 행렬에서 측정된 다섯 가지 중심성(중간 순위 생략)

순위	분야명	C _{TB}	분야명	C _M	분야명	C _{MP}
1	생물학	0.7680	생물학	2512.1	생화학	0.1788
2	교육학	0.7561	화학	2374.9	뇌과학	0.1718
3	생활과학	0.7424	재료공학	1654.8	생물공학	0.1694
4	컴퓨터학	0.7303	물리학	1558.1	면역학	0.1681
5	뇌과학	0.6584	생화학	1507.4	미생물학	0.1675
6	경영학	0.6533	전자/정보통신공학	1503.3	약리학	0.1672
7	예방의학	0.6416	환경공학	1336.7	생물학	0.1671
8	환경공학	0.6406	약학	1302.7	생리학	0.1657
9	화학	0.6316	화학공학	1300.7	임상병리학	0.1623
10	의공학	0.6150	식품과학	1214.5	의공학	0.1588
∴	∴	∴	∴	∴	∴	∴
131	기타동양어문학	0.0560	군사학	36.1	프랑스어외문학	0.0154
132	마취과학	0.0547	기타서양어문학	34.0	영어외문학	0.0145
133	응급의학	0.0387	미용	33.4	법학	0.0140
134	가톨릭신학	0.0293	사전학	33.4	무용	0.0112
135	입상안광학	0.0275	스페인어외문학	27.9	음악학	0.0058
136	러시아어외문학	0.0269	서양고전어외문학	24.4	중국어외문학	0.0056
137	서양고전어외문학	0.0242	입상안광학	21.0	러시아어외문학	0.0037
138	흉부외과학	0.0198	사진	18.2	스페인어외문학	0.0029
139	스페인어외문학	0.0082	가톨릭신학	17.5	독일어외문학	0.0007
140	산림공학	0.0063	산림공학	13.3	일본어외문학	0.0000

순위	분야명	C _N	분야명	C _{TB-10}
1	생물학	0.2446	생물학	0.9556
2	경영학	0.0647	경영학	0.8889
3	역사학	0.0647	한국어외문학	0.8889
4	한국어외문학	0.0432	화학	0.8667
5	물리학	0.0360	역사학	0.8444
6	기계공학	0.0288	재료공학	0.8222
7	생활과학	0.0288	교육학	0.8000
8	언어학	0.0288	경제학	0.7778
9	전자/정보통신공학	0.0288	전자/정보통신공학	0.7556
10	철학	0.0288	지역개발	0.7333
∴	∴	∴	∴	∴
131	프랑스어외문학	0.0000	응급의학	0.0000
132	피부과학	0.0000	이비인후과학	0.0000
133	한의학	0.0000	일반외과학	0.0000
134	항공우주공학	0.0000	일본어외문학	0.0000
135	해부학	0.0000	입상안광학	0.0000
136	해상운송학	0.0000	정형외과학	0.0000
137	해양공학	0.0000	중국어외문학	0.0000
138	해양학	0.0000	피부과학	0.0000
139	회계학	0.0000	한의학	0.0000
140	흉부외과학	0.0000	흉부외과학	0.0000



〈그림 8〉 측정 범위에 따른 삼각매개중심성의 변화

낮아지지 않고 완만하게 하락하는 경향을 보인다. 대개 노드가 밀집된 지역의 중심 노드가 동시에 전역 중심성도 높게 된다. '생물학'이 대표적인 사례로서 강한 링크를 집중시키는 전형적인 파티 허브(party hub; Han et al. 2004)라고 할 수 있으며 패스파인더 네트워크에서 뚜렷하게 부각된다.

② 지역 중심 학문: 특정 분야에 대한 영향력만 강력하여 국지적으로 중심을 차지하는데 그치는 유형이다. 측정 범위가 넓어질수록 삼각매개중심성이 급격히 하락하는 경향을 보인다. 지역중심성인 C_N 이나 C_{TB-10} 이 높은 '역사학', '경영학', '한국어와문학', '화학' 분야가 이에 해당한다. 이들도 '생물학' 분야 만큼은 아니지만 다수의 강한 링크를 가지고 있으며 패스파인더 네트워크에서 중간 수준의 허브로 나타난다.

③ 전역 연계 학문: 특정 분야에 대한 영향

력은 미미하지만 관계의 범위가 넓기 때문에 전체적인 차원에서는 결속력을 발휘하는 드문 유형이다. 여타 중심 노드 유형의 경우에는 삼각매개중심성의 측정 범위가 넓어질수록 대체적으로 중심성이 하락하는데 반해서 이 유형은 오히려 상승하는 것이 특징이다. 전역중심성인 C_{TB} 에 비해서 지역중심성인 C_{TB-10} 이 매우 낮은 '예방의학' 분야와 '인지과학' 분야가 이에 해당한다. 이들은 타 분야와 공유하는 학술지가 적지 않으나 소수 분야에 집중되지 않고 여러 분야와 고르게 조금씩 공유하는 분야로서 다수의 약한 링크를 가지고 있는 데이트 허브에 해당한다. 강한 링크 위주로 표현되는 패스파인더 네트워크에서는 이들이 허브로 두드러지지 못할 가능성이 높다.

④ 일반 학문: 특정 분야로 관계의 폭이 제한되어 있으며 지역 중심 학문의 영향력 하에

있는 학문 유형이다. 타 분야와 공유하는 학술지가 많지 않은 상당수의 분야가 이에 해당한다. 모든 범위에서 삼각매개중심성이 매우 낮게 나타난다.

이상의 네 가지 유형 중에서 전역 중심 학문과 지역 중심 학문, 그리고 일반 학문의 세 가지 유형은 패스파인더 네트워크에서 시각적으로도 대부분 식별할 수 있다. 그러나 전역 연계 학문은 주로 약한 연결만 가지고 있으므로 패스파인더 네트워크에서는 일반 학문 유형과 구분되지 않는다. 또한 중심성 척도로 살펴볼 때에도 전역 중심성과 지역 중심성을 반영하는 척도 두 가지 이상을 함께 보지 않으면 각 유형의 구분이 어렵다.

이 연구에서는 지역 중심 학문이나 전역 연계 학문과 같이 중심성의 표출 구간이 상반된 경우를 별도로 확인하기 위해서 전역 중심성을 반영하는 삼각매개중심성 C_{TB} 값으로부터 지역 중심성을 반영하는 최근접 10노드 삼각매개중심성 C_{TB-10} 값을 빼어서 산출하는 '전역 연계성 (Global Association) 척도'를 만들어 보았다.

전역 연계성 GA는 네트워크에서 지역 중심성이 낮음에도 불구하고 전역 중심성이 높은 전역 연계 노드이면 1에 가까운 값이 되고, 반대로 지역 중심성은 높으나 전역 중심성은 낮은 지역 중심 노드이면 -1에 가까운 값이 된다. 140개 학문분야에 대해서 전역 연계성을 산출한 결과는 <표 4>와 같다.

<표 4> 전역 연계성 GA 산출 결과(중간 순위 생략)

전역 연계성 ($C_{TB} - C_{TB-10}$)	전역 연계성 순위	분야명	전역 중심성 (C_{TB})	지역 중심성 (C_{TB-10})
0.6416	1	예방의학	0.6416	0.0000
0.4541	2	인지과학	0.5208	0.0667
0.4215	3	한의학	0.4215	0.0000
0.4086	4	정신과학	0.5197	0.1111
0.3186	5	체육	0.5186	0.2000
0.3159	6	재활의학	0.3381	0.0222
0.3085	7	수산학	0.4641	0.1556
0.3039	8	의공학	0.6150	0.3111
0.2830	9	해부학	0.3941	0.1111
0.2806	10	뇌과학	0.6584	0.3778
⋮	⋮	⋮	⋮	⋮
-0.2418	131	디자인	0.4249	0.6667
-0.2763	132	언어학	0.4348	0.7111
-0.2801	133	토목공학	0.4310	0.7111
-0.3077	134	지구과학	0.3145	0.6222
-0.3165	135	문학	0.2391	0.5556
-0.3188	136	역사학	0.5256	0.8444
-0.3289	137	지역개발	0.4044	0.7333
-0.3509	138	재료공학	0.4713	0.8222
-0.3528	139	정치외교학	0.3361	0.6889
-0.5671	140	한국어와문학	0.3218	0.8889

전역 연계성 산출 결과에 따르면 '예방의학'과 '인지과학', '한의학', '정신과학' 등이 전역 연계성이 0.4 이상으로 두드러지게 높은 학문으로 나타났다. 이들은 앞에서 엔트로피로 측정된 학제적인 학문과 마찬가지로 학제성이 높은 학문으로서, 학문내 연구자의 동질성이 상대적으로 낮다고 할 수 있다. 반면에 전역 연계성이 낮은 학문분야로는 '한국어외문학', '정치외교학', '재료공학' 등이 있으며, 해당 학문 연구자의 동질성이 매우 높은 분야라고 할 수 있다.

5. 결론

학술활동을 근거로 국내 학문분야의 구조를 파악하기 위해서 국내 연구자의 학술지 논문 발표 자료를 활용하여 학문분야간 학술지 공유도를 산출하고, 이로부터 국내 학문분야 패스파인더 네트워크를 생성하였다. 생성된 패스파인더 네트워크는 '생물학'을 핵심으로 하는 생명과학 분야가 중앙을 차지하고 있었으며, 각 학문분야를 학문분류표상의 8가지 대분류 영역으로 구분해본 결과 '인문학'과 '의약학', '공학'에 속한 학문끼리는 학술지 공유를 통한 학문간 연계가 매우 강하게 나타났고 '예술체육'이나 '복합학' 영역은 소속 학문분야 사이의 연계가 미미한 것으로 나타났다.

개별 학문분야가 다른 여러 분야와 학술지를 공유하는 정도를 엔트로피로 측정한 결과에서는 '과학기술학', '인지과학', '감성과학'과 같은 학제적 분야의 엔트로피가 높아서 다양성이 큼을 확인하였고, 반대로 '일본어외문학' 등 주요어문학들이 대부분 엔트로피가 낮아서 학술지

를 공유하는 다른 분야가 폭넓지 못한 것으로 나타났다.

중심성이 큰 학문 분야를 파악하기 위해서 가중 네트워크를 위한 중심성 척도를 적용하고 삼각매개중심성의 측정 범위를 달리하면서 살펴본 결과 '생물학'과 같은 전역 중심 학문, '한국어외문학'과 같은 지역중심 학문, '인지과학'과 같은 전역연계 학문, 그리고 기타 일반 학문의 네 가지 유형을 식별할 수 있었다.

이상의 분석을 수행한 결과, 가중 네트워크에서 강한 링크만을 남기는 패스파인더 네트워크는 데이트 허브를 제대로 표현하지 못하는 것으로 나타났다. 데이트 허브는 지역 중심성은 낮으나 전역 중심성이 높은 성향을 가지므로 이 두 가지 지표를 함께 활용하여 식별하는 것이 바람직하다고 판단된다.

연구자의 논문 발표 학술지 공유도에 근거해 확인된 각 학문분야의 특성은 향후 국내 학술지원 정책이나 연구정보 관리 활동에 참고가 될 것으로 기대된다. 학술지원 정책에 수립함에 있어서는 전역 연계 학문과 같이 소속 연구자가 이질적인 분야에 대한 평가 및 지원 전략은, 지역 중심 학문처럼 소속 연구자가 동질적인 분야와는 다르게 수립해야 할 것이다. 학술 및 연구 정보 관리 측면에서도 이와 같은 학문분야별 연구자의 특성을 고려할 필요가 있다.

이 연구에서는 분석 자료에 논문 발표 시기 정보가 없으므로 인하여 학문분야 연계 네트워크에 대한 역동적인 분석을 수행하지 못하였으나, 추후 정보를 확보하여 네트워크의 역동성을 확인하는 것이 매우 중요한 과제이다. 또한 후속 연구에서 국내 학문분야간 인용 네트워크가 확보된다면 텍스트 유사성, 저널 공유도, 저널

상호인용 등의 세 가지 방식에 의해 도출되는 학문간 관계와 네트워크를 상호 비교 분석함으

로써 각 자료와 그에 기반한 네트워크에 대해서 많은 것을 밝힐 수 있을 것으로 기대된다.

참 고 문 헌

- 이재윤. 2006a. 지적 구조의 규명을 위한 네트워크 형성 방식에 관한 연구. 『한국문헌정보학회지』, 40(2): 333-355.
- 이재윤. 2006b. 계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구. 『한국문헌정보학회지』, 40(3): 191-214.
- 이재윤. 2007. 국내 광역 과학 지도 생성 연구. 『정보관리학회지』, 24(3): 363-383.
- Boyack, K. W., R. Klavans, and K. Börner. 2005. "Mapping the backbone of science." *Scientometrics*, 64(3): 351-374.
- Csermely, P. 2006. *Weak Links: Stabilizers of Complex Systems from Proteins to Social Networks*. Berlin: Springer-Verlag.
- de Nooy, W., A. Mrvar, and V. Batagelj. 2005. *Exploratory Social Network Analysis with Pajek*. New York: Cambridge University Press.
- Garfield, E. 1994. "Scientography: Mapping the tracks of science." *Current Contents: Social & Behavioral Sciences*, 7: 5-10.
- Granovetter, M. 1973. "The strength of weak ties." *American Journal of Sociology*, 78: 1360-1380.
- Han, J. D. J., N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth, and M. Vidal. 2004. "Evidence for dynamically organized modularity in the yeast protein-protein interaction network." *Nature*, 430: 88-93.
- Mantegna, R. N. 1999. "Hierarchical structure in financial markets." *European Physical Journal B*, 11(1): 193-197.
- Moya-Anegón, F., B. Vargas-Quesada, V. Herrero-Solana, Z. Chinchilla-Rodríguez, E. Corera-Álvarez, and F. J. Muñoz-Fernández. 2004. "A new technique for building maps of large scientific domains based on the cocitation of classes and categories." *Scientometrics*, 61(1): 129-145.
- Newman, M. E. J. 2004. "Analysis of weighted networks." *Physical Review E*, 70: 056131.
- Price, D. J. DeSolla. 1966. "The science of scientists." *Medical Opinion and Review*, 1(10): 88-97.
- Schvaneveldt, R. W.(ed). 1990. *Pathfinder Associative Networks: Studies in Knowledge Organization*. Norwood, NJ: Ablex.
- Small, H. and E. Garfield. 1985. "The geog-

raphy of science: disciplinary and national mappings." *Journal of Information Science*, 11: 147-159.

Vargas-Quesada, B. and F. Moya-Anegón. 2007. *Visualizing the Structure of Science*. Berlin: Springer-Verlag.