

국가연구개발사업과 논문성과간 네트워크 분석을 통한 사업별 유사성 분석

A Similarity Analysis on National R&D Programs and Paper Outputs using Network Analysis

장한수(Hansoo CHANG)*, 홍정석(HONG)**, 최주석(CHOI)****

목 차	
I. 서론	III. 네트워크 분석
II. 데이터 및 네트워크의 규정	IV. 결론

논문 요약

정부의 연구개발투자로 수행되는 국가연구개발사업은 세부과제 단위로 매년 도출되는 6대 성과(논문, 특허, 기술료, 사업화, 인력양성 지원, 연수 지원)를 집계한다. 성과 중 하나인 SCI급 논문은 한 논문을 다른 연구과제의 성과로 제출하는 경우도 많아서 동일 논문성과를 다수의 과제에서 공유하게 된다. 본 논문에서는 국가연구개발사업과 그 성과로 도출된 SCI급 논문간의 네트워크 관계를 분석하고자 한다. 대상 데이터는 2015년도에 수행된 연구개발과제의 성과로 제출된 11,179편의 SCI급 논문 중 2개 이상의 과제 결과로 도출된 논문이다. 또한 국가연구개발사업의 SCI급 논문 성과는 세부과제 단위로 집계되므로 SCI급 논문-세부과제간 네트워크는 정보량이 과다하다. 따라서 SCI급 논문에 대한 네트워크 분석은 세부과제를 포함하는 국가연구개발사업간의 네트워크로 한 단계 높여 분석하였다. 구체적으로 논문-사업간 2모드 네트워크를 1모드 네트워크로 전환하여 네트워크 분석을 시도하였다.

Keyword : 논문성과, 국가연구개발사업, 네트워크 분석, 사업성과

* 국가핵융합연구소 책임연구원, jjang@nfri.kr, 042-879-6111

** 한국과학기술기획평가원, 연구위원, jshong@kistep.re.kr, 02-589-2296

*** 한국과학기술기획평가원, 연구원, jschoi@kistep.re.kr, 02-589-5259

I. 서론

국가연구개발사업은 연간 대략 600여개 사업에서 5만4천여 개의 과제를 통해 19조 원 가량의 예산이 집행된다. 이러한 투자로 2015년 기준 대략 35,849 편의 SCI급 논문과 14,957 건의 국내등록 특허, 20,088 건의 사업화, 7,372 건의 기술료 징수 등의 성과가 도출되었다. 이러한 성과들은 개별 세부과제별 직접적 연구개발 활동의 성과로 도출된다. 다른 한편으로는 연구개발 활동이라는 특성 상 각 세부과제들의 연구결과의 다양한 방식의 결합을 통해 도출될 수도 있다. 실제로 SCI급 논문과 특허의 상당수는 단일 과제가 아닌 다수 과제의 성과로 도출되고 있다.

이러한 정보를 분석하면 연구개발 활동의 결과가 논문 등의 성과로 도출되는 다양한 유형을 파악할 수 있게 된다. 이는 연구개발 활동이 성과로 전환되는 경로, 방식, 과정 등의 이해에 도움이 될 수 있다.

본 고는 이러한 활동의 탐색적 시도로 2015년 국가연구개발사업 성과 데이터 중 논문 성과가 공유되는 형태에 대한 네트워크 분석을 수행하였다. 주요 내용은 데이터의 구조와 그에 따라 규정되는 네트워크를 소개한다. 대상이 되는 네트워크에 대한 특성과 구조를 밝힘으로써 국가연구개발사업간의 성과 발생 형태에 대한 특성을 확인할 수 있는지 검토하는 내용으로 전개한다.

II. 데이터 및 네트워크의 규정

1. 데이터

대상 데이터는 2015년도에 수행된 연구개발과제의 성과로 제출된 11,179편의 SCI급 논문 중 2개 이상의 과제에 나타난 논문이다.¹⁾ 이를 <표 1>을 바탕으로 논문 성과의 구성에 대한 예시를 들면 다음과 같다. 국가연구개발사업의 성과는 PJT 단위로 수집되고 그 중 본 고에서 활용하는 성과는 논문성과이다. 이 논문성과는 각 PJT에서 중복될 경우 과제별 기여율을 별도로 구분한다. 예를 들어 paper1부터 paper4는 두 개 이상의 PJT에서 공통된 성과로 나타나고 이는 과제별로 기여율이 다르다. paper5의 경우는 PJT4-1, PJT4-2에서 공유되는 성과로 제출되나 PJT4-1에서만 유효하다. paper6은 PJT4-3의 단독 성과로 제출되었다. 결과적으로 순번 1~9만 분석 대상 데이터이다.

1) 2015년 기준 전체 논문건수 35,849 편 중 연구과제의 단독성과로 확인된 24,670건을 제외하고 2개 이상의 연구과제에서 공유되는 성과로 확인된 논문의 편수로 전체 논문의 32.1%임

실제 데이터는 순번 10, 11과 같이 기여율 0 또는 단독인 논문을 제외한 총 27,795편의 논문을 공유하는 10,110개의 연구과제, 309건의 PGM간 관계를 대상으로 한다. 네트워크 분석 도구는 Gephi 0.92를 활용하였다.

<표 1> 데이터의 구조

순번	국가연구개발사업(PGM)	세부과제(PJT)	논문성과(paper)	기여율(weight)
1	PGM1	PJT1-1	paper1	0.6
2		PJT1-2	paper2	0.7
3	PGM2	PJT2-1	paper1	0.4
4		PJT2-2	paper2	0.3
5		PJT2-2	paper3	0.9
6	PGM3	PJT3-1	paper3	0.1
7		PJT3-2	paper4	0.5
8		PJT3-3	paper4	0.5
9	PGM4	PJT4-1	paper5	1
10		PJT4-2	paper5	0
11		PJT4-3	paper6	1

2. 네트워크의 규정

논문과 연구과제간 네트워크는 과제수나 논문의 수가 다수이면서 연결도 다양하기 때문에 특정하게 유의미한 노드를 발견하기가 용이하지 않다. 따라서 본 고에서는 연구과제를 프로그램(PGM) 단위로 묶고 논문-PGM간 2모드 네트워크를 먼저 구성한다. 여기서 PGM간 직접적 연관성이 없는 것을 1모드로 전환하여 PGM간 연관을 나타내는 네트워크를 구성한다.

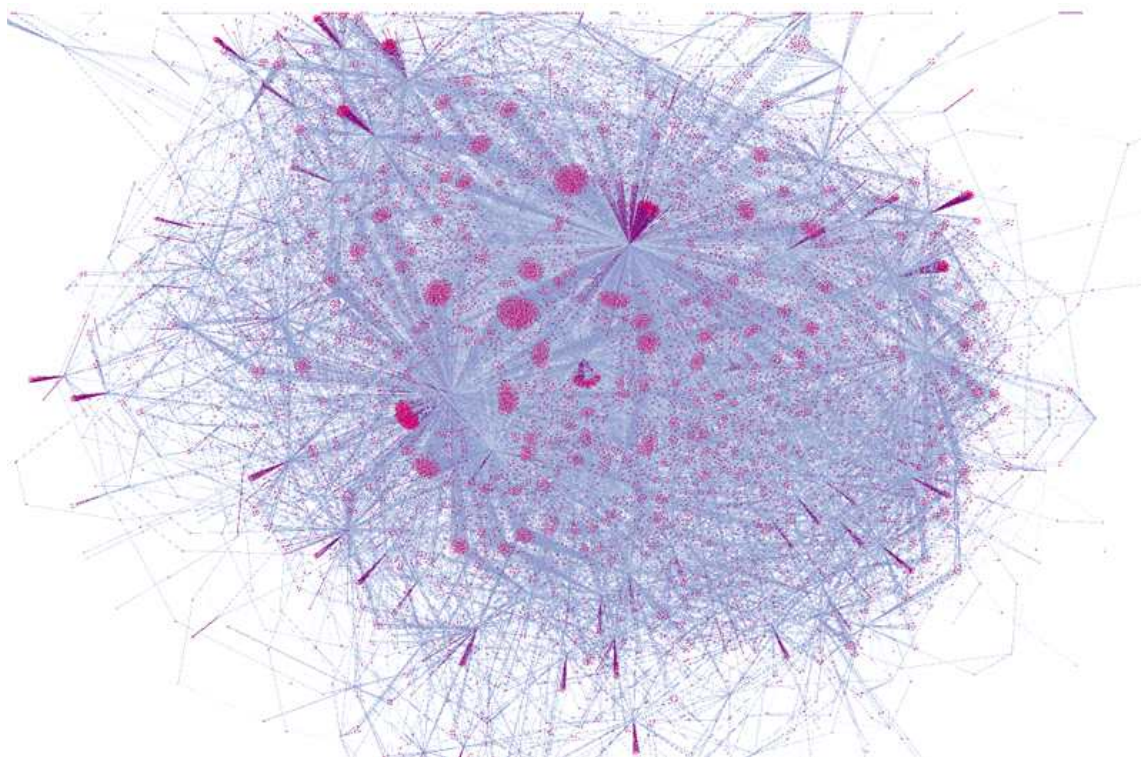
논문-PGM 간 네트워크의 일부를 나타내면 <그림 1>과 같다. 논문과 PGM간 2모드 네트워크로 논문-논문 또는 PGM-PGM간 직접 연결은 없다. 다만 공통된 논문 성과가 두드러진 PGM들이 드러남을 알 수 있다.

III. 네트워크 분석

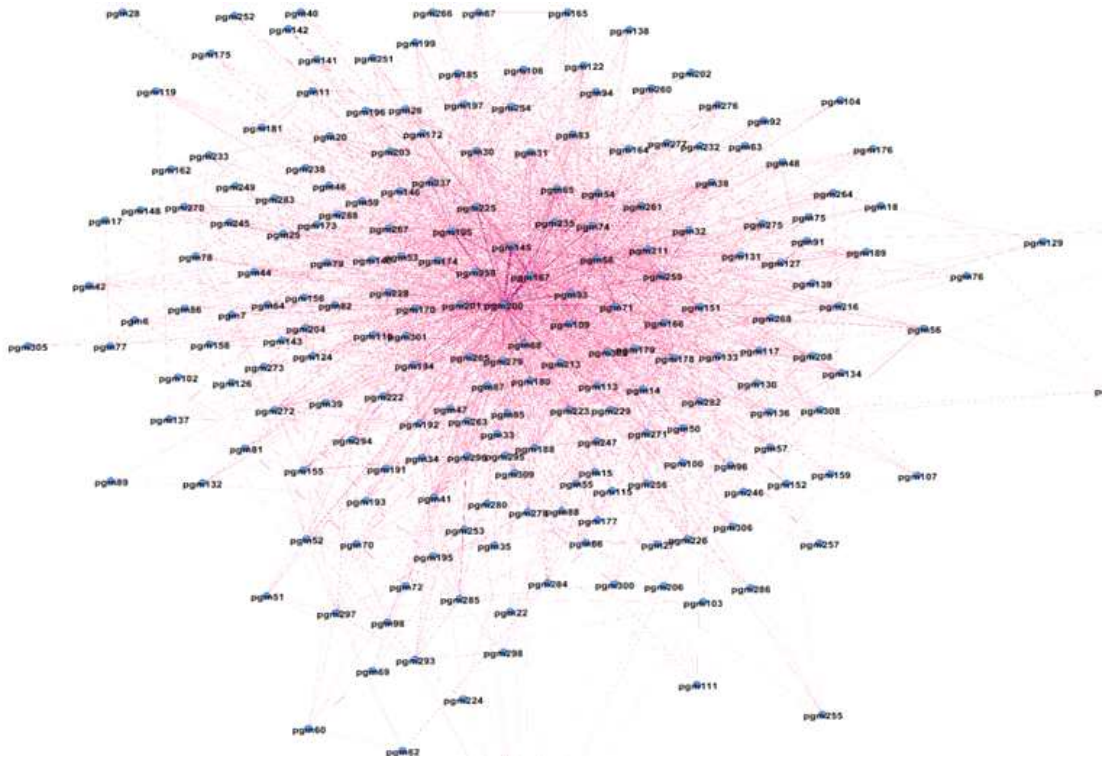
본 고에서 파악하고자 하는 PGM간 네트워크 분석을 위하여 <그림 1>의 네트워크를 PGM간 1모드 네트워크로 전환하여 나타내면 <그림 2>와 같다. 여기서 degree가 8 이하인 노드는 제외하고 나타내어 주요한 PGM간 네트워크만 파악하고자 하였다.

1. 네트워크 특성

PGM으로만 구성된 노드는 215개로 2,601개의 링크로 연결되어 있다. 네트워크 밀도(density)는 0.113, 평균 링크(average degree)는 24.195, 군집계수(clustering coefficient)는 0.596으로 나타났다. degree 8 이상인 네트워크로 한정하였기 때문에 비교적 네트워크 연결은 높은 편이다.



<그림 1> 논문-PGM간 네트워크(논문을 매개로 연결된 PGM간 2모드 네트워크를 Gephi의 MultiGravity ForceAtlas 2 Layout으로 구현)



<그림 2> PGM간 1모드 네트워크(Gephi의 MultiGravity ForceAtlas 2 Layout으로 구현)

2. 주요 노드

<그림 2>에서 링크가 높은 상위 20개의 노드에 대한 특성치를 나타내면 <표 2>와 같다. 상위에 포진된 사업들은 주로 기초연구, 신진연구자지원, 선도연구센터, 학술연구기반 구축 등 과제별로 범위가 다양한 사업들이다. 또한 출연연 운영사업, 대형장비 운영 사업 등도 상위에 나타나는 것으로 확인되었다.

〈표 2〉 Degree 기준 상위 20개 node의 특성치

node	Degree	Weighted degree	Closeness centrality	Betweenness centrality	Eigenvector centrality	Clustering coefficient
pgm200	183	559.54	0.8765	5407.26	1.0000	0.1321
pgm167	160	312.25	0.8008	3382.24	0.9491	0.1604
pgm145	133	233.91	0.7245	1870.24	0.8642	0.1972
pgm201	108	131.21	0.6698	828.12	0.7891	0.2586
pgm58	100	148.43	0.6514	563.21	0.7725	0.2960
pgm105	89	136.09	0.6283	659.91	0.6540	0.2755
pgm259	86	100.96	0.6246	479.51	0.6858	0.3223
pgm265	80	47.36	0.6156	347.46	0.6536	0.3367
pgm235	80	103.98	0.6138	311.47	0.6721	0.3595
pgm71	78	74.63	0.6121	383.13	0.6385	0.3447
pgm225	74	89.89	0.6017	367.84	0.5844	0.3291
pgm109	73	63.94	0.6000	286.01	0.6276	0.3771
pgm65	71	61.09	0.5966	210.68	0.6064	0.3738
pgm261	67	40.34	0.5933	278.24	0.5910	0.4012
pgm151	67	41.50	0.5917	284.18	0.5621	0.3740
pgm93	67	61.81	0.5900	290.42	0.5740	0.3754
pgm68	65	56.38	0.5884	219.41	0.5829	0.4154
pgm54	55	36.03	0.5741	120.28	0.5082	0.4438
pgm74	56	53.52	0.5726	99.39	0.5273	0.4701
pgm166	55	39.45	0.5726	106.98	0.5119	0.4653

IV. 결론

국가연구개발사업에서 발생하는 논문성과는 단일한 연구과제에서 생성되기도 하지만, 다양한 과제로부터 발생하는 경우도 30% 이상으로 나타난다. 본 고에서는 연구개발사업 수준에서 어떻게 공통된 논문이 발생하는지 파악하고자 네트워크 분석을 시도하였다.

분석결과 공통된 논문이 20건 이상되는 사업으로 한정해도 85개로 나타났다. 이는 지식의 창출이 단일 과제에서 생성되기 보다는 다양한 연구활동에 따른 것이라는 반증이다.

다만, 본 연구는 시험적 성격으로 네트워크 분석을 통하여 각 사업의 세부적 상황에 대한 검토가 필요하나 이는 학술적 영역에서 다루기에는 논란의 소지가 있다. 따라서 국가연구개발사업 또는 세부 연구과제에 대한 공통적 학술성과 발생 특성을 논문에 의한 네트워크 분석으로 확인할 수 있다는 점에 의의가 있다.

참고 문헌

- 미래부 · 한국과학기술기획평가원 (2017), 「2015년도 국가연구개발사업 성과분석보고서」, <http://www.kistep.re.kr/>
- 임성민 · 김진흠 (2013), “사회 네트워크 분석으로 본 수강신청 행태”, 「응용통계연구」, 26(4) : 661-674.
- Bastian, M., Heymann, S., and Jacomy, M. (2009), “Gephi: an open source software for exploring and manipulating networks”, International AAAI Conference on Weblogs and Social Media