

계량서지학적 분석을 통한 약물연구분야 특허출원인 간 영향력 비교*

A Comparative Study of the Impacts among Patent Assignees in Pharmaceutical Research based on Bibliometric Analyses

김희영 (Heeyoung Kim)**

박지홍 (Ji-Hong Park)***

초 록

본 연구는 약물 연구 분야에 속하는 특허 사이에 나타나는 지식의 흐름을 살펴보고 이들 간의 영향력을 파악해보기 위해 특허데이터에서 나타나는 인용 관계를 분석하였다. 특허데이터의 수집은 Google Patents에서 진행하였다. 약물 연구와 관련된 특허 문서를 검색하여 상위 25개의 출원인을 선정하였고, 이를 바탕으로 출원인 사이의 인용 관계를 알아보고 각 출원인의 각 문서에 대한 피인용빈도와 순위를 활용하여 h-지수와 h_s-지수의 파생지표들의 값을 계산하여 비교하였다. 분석 결과를 종합하면, 'Pfizer, MIT, Abbott' 등의 출원인이 약물 연구 분야에서 영향력이 높은 출원인으로 드러났다. 5개의 계량서지학적 지표 중에서 g-지수와 h_S-지수가 서로 유사한 결과를 보여주었고, 총인용빈도, 최대인용빈도, CPP의 순위를 가장 잘 반영하는 지표로 나타났다. 또한, 총인용빈도, CPP, 최대인용빈도 순으로 5개의 계량서지학적 지표와의 상관관계가 높았다. 한편, 기존의 특허 출원인의 기술적 영향력을 나타내는 것으로 알려진 지표인 CPP만으로는 정확한 비교가 어려운 경우도 나타났다.

ABSTRACT

This study analyzes the relationship of citations appearing in the patent data to understand knowledge transfers and impacts between patent documents in the field of pharmaceutical research. Patent data were collected from a website, Google Patents. The top 25 assignees were selected by searching for patent documents related to pharmaceutical research. We identify the citation relationships between assignees, then calculate and compare the values of h-index and derived indicators by using the number of citations and rank for each document of each assignee. As a result, in the case of pharmaceutical research, the assignee, such as 'Pfizer, MIT, and Abbott' shows a high impact. Among the five bibliometric indicators, the g-index and h_S-index show similar results, and the indicators are the most related to the rankings of Total Citation Frequency, Cites per Patents, and Maximum Citation Frequency. In addition, it is highly related to the five indicators in the order of Total Citation Frequency, Cites per Patents, and Maximum Citation Frequency. In some cases, it is difficult to make an accurate comparison with Cites per Patents alone, which is previously known to indicate the technological influence of patent assignees.

키워드: h-지수, 인용, 네트워크, 특허, 약물연구
h-index, citation, network, patent, pharmaceutical research

* 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2019S1A5C2A03083499).

** 연세대학교 일반대학원 문헌정보학과 석사과정(kimhy91@yonsei.ac.kr) (제1저자)

*** 연세대학교 문헌정보학과 교수(jihongpark@yonsei.ac.kr) (교신저자)

■ 논문접수일자: 2022년 2월 9일 ■ 최초심사일자: 2022년 3월 7일 ■ 게재확정일자: 2022년 3월 14일
■ 정보관리학회지, 39(1), 1-15, 2022. <http://dx.doi.org/10.3743/KOSIM.2022.39.1.001>

※ Copyright © 2022 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

특허는 지적 자산의 유형 중 하나로, 일정 기간동안 특허의 소유자가 자신의 발명에 대해 법적 권리를 가질 수 있게 하는 역할을 한다. 특허데이터의 디지털화와 컴퓨터 성능의 증가로 특허에 대한 활용이 늘고 있으며, 지난 20년간 사회과학 연구에서 특허 인용 데이터의 사용이 증가하였다(Jaffe & De Rassenfosse, 2019). 특허는 상업적인 측면, 기술적인 측면 모두에서 중요한 역할을 하는데, 상업적으로는 경쟁자로부터 발명에 대한 권리를 지켜주는 역할을 하고, 기술적으로는 신기술을 접할 기회를 주고 기술의 발전을 돕는 역할을 한다(Walter, 2014).

기존의 인용 색인에서 나타나는 인용 정보로 기반으로 하는 계량서지학적 방법은 일반적으로 학술 출판물의 연구 영향력을 평가하기 위한 수단으로 이용된다(Kousha & Thelwall, 2015). 그러나 일부 학술 출판물은 학술 인용 정보로는 직접적으로 드러나지 않는 상업적 유용성을 가지고 있다. 이러한 상업적 유용성은 특허 문서에 의한 인용을 통해 드러날 수 있다. 한편, 특허 문서는 학술 출판물에 대한 인용과 마찬가지로 선행된 기술과 관련된 기존의 특허 문서를 인용하기도 한다(Walter, 2014). 그러므로 특허데이터에서 나타나는 인용 관계를 분석해 본다면 특허 문서 사이에서 나타나는 지식의 흐름을 살펴볼 수 있을 것이고, 이들 간의 영향력을 확인해 볼 수 있을 것이다.

본 연구는 특허데이터의 인용 관계를 반영할 수 있는 지표를 찾아보고, 약물 연구 분야의 기업 간 영향력을 비교하는 것을 목적으로 한다.

본 연구를 통해 알아보고자 하는 연구 질문은 다음과 같다. 우선, 기존에 알려져 있는 여러 계량서지학적 지표들의 특징을 살펴보고, 이러한 지표들을 특허데이터에 적용하여 서로 다른 지표를 적용하였을 때 특허 출원인들의 순위 변화를 살펴보고자 한다. 또한, 특허데이터의 영향력을 잘 반영하는 지표를 선정하여 해당 데이터 내에서 가장 영향력이 있는 특허 출원인을 확인한다. 마지막으로 인용 관계를 바탕으로 하는 네트워크를 통해 특허출원인 간의 인용 관계를 알아보고자 한다. 최근 건강에 대한 관심과 신약 개발에 대한 관심이 높아짐에 따라 약물 연구 분야 기업의 경쟁력을 알아보는 것이 의미가 있을 것이라 생각되어, 본 연구에서는 특허 데이터 중에서도 약물 연구 분야의 특허 문서를 대상으로 사례 연구를 진행하였다.

2. 이론적 배경

2.1 평가를 위한 계량서지학적 지표

학술 출판물이나 저자, 또는 저널 등을 평가하기 위해 사용되는 계량서지학적 지표에는 대표적으로 h-지수가 있다. 예를 들어, h-지수를 이용하여 특정 연구자를 평가하고자 한다면 그 연구자의 논문을 피인용빈도 기준으로 내림차순 정렬을 한 후에 '피인용빈도보다 순위값이 작거나 같은' 조건을 만족하는 최대의 순위값을 h-지수로 한다. h-지수의 파생지표로 g-지수, hT-지수 등이 있는데, g-지수는 '해당 순위까지의 피인용빈도의 합이 순위값의 제곱보다 더 이상 크거나 같지 않아질 때'

의 순위값을 의미하고(Egghe, 2006), hT-지수는 h-지수에서 핵심 범위(h-core)를 벗어난 논문들이 반영되지 않는 한계점을 보완하기 위해 정수론의 Durfee square 개념을 적용한 지표이다(Anderson, Hankin, & Killworth, 2008). 본 연구에서는 특허데이터를 평가하기 위한 방법으로 h-지수, g-지수, hT-지수를 활용하였다.

여기에 추가로, 이재운(2006)이 제안한 hS-지수와 hA-지수를 특허데이터에 적용하고자 하였다. hS-지수는 h-지수의 h 순위 이상의 피인용빈도가 지표 값에 잘 반영되지 않는 단점과 g-지수의 지나치게 최상위 논문의 인용빈도에 좌우되는 단점을 보완하기 위한 것으로, 최상위 논문의 인용빈도를 제한적으로 반영할 수 있게 하는 지표이다. hS-지수를 구하는 공식은 다음과 같다(수식 1).

$$hS\text{-index} = \sum_{r=1}^h \sqrt{C(r)} \quad (1)$$

hA-지수는 각도와 코사인 계수를 이용하는 지표로, 각 문서를 순위와 피인용빈도를 좌표로 하는 점으로 나타낸다고 가정하고, 원점으로부터 각 문서를 나타내는 점을 잇는 직선과 x축 사이의 각도에 대한 코사인 값을 구한 값을 활용한다(이재운, 2006). hA-지수를 구하는 공식은 다음과 같다(수식 2).

$$hA\text{-index} = h - \sum_{r=1}^h \frac{r}{\sqrt{r^2 + C(r)^2}} \quad (2)$$

마지막으로 CPP는 특허데이터의 인용도지수(Cites per Patents)를 의미하며, 해당 특허가 얼마나 인용이 되었는지를 등록되어 있는 전

체 특허 건수로 나누어 주는 식으로 계산하여 대상 특허들의 평균 피인용빈도를 나타낼 수 있다(서진이 외, 2006). CPP값이 높을수록 주요한 특허일 가능성이 높아, 특허데이터에서 질적 측면과 기술적 영향력을 나타내는 지표로 쓰인다.

2.2 인용 네트워크에서의 중심성

인용 네트워크 내에서 각 노드의 영향력을 알 수 있는 척도로 중심성(Centrality)에 대한 개념을 적용하였다. 중심성은 특정 노드가 전체 네트워크 내에서 얼마나 중요한 위치에 있는지를 나타낼 수 있는 중요한 지표이다(Zhang & Luo, 2017). 중심성이 높을수록 해당 노드가 네트워크의 중심에 가까움을 의미하고 영향력, 편리성의 측면에서 유리할 수 있음을 나타낸다. 사회과학의 영역에서 주로 사용하는 중심성의 종류로는 연결중심성(Degree centrality), 근접중심성(Closeness centrality), 매개중심성(Betweenness centrality), 고유벡터 중심성(Eigenvector centrality)이 있다. 특히, 고유벡터 중심성은 연결된 노드가 단순히 많은 것 외에도 얼마나 중요한 노드와 연결되었는지도 함께 반영한 중심성 지표라고 할 수 있다(Ruhnau, 2000). 상대적 중요도에 따라 가중치를 부여한다는 관점에서 페이지랭크(PageRank) 알고리즘이 이와 유사하다. 페이지랭크 알고리즘은 웹 페이지의 상대적 중요도를 계산하기 위해 고안된 것으로, 페이지랭크 알고리즘에서도 영향력이 높은 노드로부터 인용을 많이 받을수록 대상 노드의 영향력이 높아지는 효과가 있다(Yan & Ding, 2011). 본 연구에서는 25개의 출원인 사이에서 나타나는 직접적인 인용횟수를 통해

네트워크를 나타내고, 서로 의존적인 관계 안에서 각 출원인의 영향력을 알아볼 수 있도록 페이지랭크 알고리즘을 활용하여 중심성을 계산하였다.

2.3 선행연구

특허데이터를 활용한 선행연구 중 하나로, 한국 신약개발의 기술적 역량 진단과 개선방안에 대한 연구가 있었는데, 이는 특허명세서에 나타난 정보들과 세부 지표를 활용하여 기술적 역량의 일반적 흐름을 파악하고, 정량적 특허평가시스템인 K-PEG를 활용하여 한국 신약개발의 질적 수준을 평가한 연구이다(엄익천, 박진서, 김봉진, 2019). 해당 연구에서는 특허명세서의 출원인 정보를 활용하면 협력 구조의 파악이 가능함을 제시하였다.

한편, 시대별 주요 특허 기술군의 변화를 파악하고 유사한 기술들의 군집을 찾아내는 것을 목적으로 하는 연구가 있었다(이민정, 김용대, 장원철, 2016). 해당 연구에서는 특허 인용 네트워크를 통해 특허의 가치를 적절히 평가할 수 있는 지표로서 페이지 랭크 알고리즘을 가정하고 기존의 연결중심성, 근접중심성, 매개중심성, 고유벡터 중심성과의 차이를 살펴 보았다. 특허 인용 네트워크에 확률적 블록 모델(Stochastic block model), 잠재 공간 모델(Latent space model)을 적합하여 각 모델이 특허 인용 네트워크를 분석하는 데 적절한지 분석하였으며, 기술군 그룹의 특징 및 각 그룹에 속한 기술군들의 연관성을 해석하고자 하였다.

다른 연구로는, 특허 인용 분석과 관련된 연구로, 여기서는 특허 분야(Family)의 개념을

소개하며, 여러 기관의 특허데이터를 특허 분야와 함께 제공하면 기존의 특허 인용 분석의 적용 범위와 실행 가능성을 크게 향상시킬 수 있음을 제시하였다(Nakamura et al., 2015). 해당 연구에서는 인용에 대한 정보를 얻기 위해 DWPI(Derwent World Patent Index)와 DPCI(Derwent Patent Citation Index)를 이용하여 데이터를 얻었다. 인용 정보 유형 중, 두 특허 간 직접적인 연결 외에도 '직접 연결된 특허의 파생 특허문서와의 연결' 또는 '직접 연결된 두 특허의 파생 특허문서 간의 연결' 등을 모두 이용하여 분석을 하게 되면 특허 간 연결의 밀도를 더 잘 반영할 수 있고, 특정 분야의 현재 기술 경향을 더 잘 설명할 수 있다고 제시하였다.

3. 연구방법

3.1 조사 대상과 자료 수집

특허데이터의 수집을 위한 방법으로 Google에서 제공하는 특허문서 데이터베이스인 Google Patents(<https://patents.google.com>)를 활용하였다. 약물 연구와 관련된 특허 문서를 검색하기 위해 '(drug) (A61K) (A61P) language: ENGLISH type: PATENT'와 같이 검색어를 입력하여 검색된 결과들을 python을 통해 작성한 크롤링(Crawling) 코드를 활용하여 수집하였다. A61K와 A61P는 CPC(Cooperative Patent Classification)를 의미하는데, 이는 미국 특허청과 유럽 특허청이 선행기술조사의 효율성을 위해 개발한 특허분류코드(United States

Patent and Trademark Office & European Patent Office, 2015)로, 본 연구에서는 주제 분야를 의미하는 역할로 사용하였다. A61K와 A61P 코드는 각각 '의료, 치과 또는 화장실과 관련된 약품(Preparations for medical, dental, or toilet purposes)', '화학 화합물 또는 의약품에 대한 특정 치료 활동(Specific therapeutic activity of chemical compounds or medicinal preparations)'에 해당하는 항목을 포함하게 된다.

사전 수집 단계에서는 특허데이터의 인용관계에 대해 대략적으로 알아보는 목적으로, 본 수집에 앞서 시범적으로 데이터를 수집하였다. 입력한 검색어와 관련도가 높은 순으로 총 1,000개의 특허 문서에 대한 정보를 수집하였다. 논문에서의 저자에 해당하는 개념으로 특허에서는 발명자(Inventor) 또는 출원인(Assignee)을 활용할 수 있는데, 발명자의 경우 동명이인의 문제가 발생하거나, 같은 사람의 이름이 여러 방식으로 표기될 가능성이 있기 때문에 데이터의 처리에 있어 어려움이 발생할 수 있다.

또한, 개개인 발명자의 특허문서의 수는 논문 데이터에서 적용하였던 것처럼 h-지수 및 관련 파생지표들을 적용하기에는 데이터의 개수가 적어 인용 관계를 분석하기에 적절하지 않을 수 있다. 따라서 발명자 대신 출원인을 기준으로 분석을 실시하였다. 출원인의 경우 (계약) 회사명 또는 연구기관명이 주로 나타나기 때문에 발명자의 경우보다 혼동될 가능성이 낮으며, 계량서지학적 지표를 적용할 수 있는 만큼의 특허 문서가 존재하므로 이를 활용하여 분석하였다.

본 수집 단계에서는, 사전 수집 단계에서 파악한 출원인 중에서 출원한 특허의 수가 많았던 상위 28개의 출원인을 선정하였다. 이를 통해, Google Patents에서 각 출원인에 대해 최대 100개의 특허문서를 수집하였다. 수집한 결과 문서의 수가 적어 분석이 어려운 일부 출원인을 제외하고 최종적으로 25개의 출원인에 대해 데이터를 확보하여 분석에 활용하였다. 25개의 출원인에 대한 목록은 <표 1>과 같다.

<표 1> 약물 연구 관련 특허 출원인의 목록

순번	특허 출원인	순번	특허 출원인
1	Abbott	14	Intarcia Therapeutics
2	AbbVie	15	Johnson and Johnson
3	Allergan	16	Mannkind
4	Alza	17	Massachusetts Institute of Technology
5	Antecip Bioventures II	18	Novartis
6	Baxter	19	Noven Pharmaceuticals
7	Bayer	20	Pfizer
8	Bend Research	21	pSivida
9	Boehringer Ingelheim	22	Samyang
10	Daiichi Sankyo	23	Seattle Genetics
11	Euro Celtique	24	The Johns Hopkins University
12	Genentech	25	University of California
13	Hanmi		

3.2 데이터 분석

각 특허문서에서는 '출원인, 특허문서의 ID, 피인용 수, 분야 내 또는 분야 간 해당 문서를 인용한 특허문서의 ID'와 같은 정보들을 중점적으로 수집하였다. 피인용 수는 인용 관련 색인을 적용하는데 활용하였고, 해당 문서를 인용한 특허문서의 ID에 대한 정보들은 종합하여 25개의 출원인 사이에서의 인용관계를 알아보고자 하였다. 데이터에 대한 분석에서는 R 4.0.2(R Development Core Team, 2011)를 주로 사용하여 분석하였다. 각 출원인의 각 문서에 대한 피인용빈도와 순위를 활용하여 h-지수, g-지수, hS-지수, hA-지수 및 hT-지수 값을 계산하였다. 25개 출원인 사이에서의 인용관계를 알아보기 위해서는, 분야 내 인용관계와 분야 간 인용관계로 나누어, 각각 인용-피인용 관계를 정방형 매트릭스로 정리한 다음, R의 igraph 패키지를 통해 네트워크 분석을 하였다.

4. 분석 결과

4.1 계량서지학적 지표를 활용한 분석

분석을 위한 특허데이터를 확보할 수 있었던 상위 25개의 출원인에 대해 h-지수 및 h-지수의 파생지표들에 대한 값을 계산하였고, 각 출원인의 기술적 영향력을 나타낼 수 있는 지표인 CPP를 구하였다. 또한, 각 출원인의 특허 문서 중 가장 피인용이 많이 된 문서의 피인용빈도 값을 구한 후 이를 Max라 지칭하였다. 총 특허수(N), 총 인용횟수(TC)를 포함하여 각 특허

출원인의 계량서지학적 지표를 구한 값과 각각의 순위는 <표 2>에 제시하였다.

h-지수의 경우, 다른 지표에 비해 동일 순위를 가지는 출원인이 다수 존재하였고, 점수의 폭이 다소 좁은 특징을 가졌다. 반면, g-지수의 경우 동점자의 수가 많이 줄어들었고 h-지수에서 상위권에 위치하지 않았던 출원인이 상위권에 올라오는 등 순위에 차이가 발생하였다.

1번 출원인(Abbott)의 경우, 가장 높은 h-지수 값으로 나타났는데, 해당 출원인은 다른 지표에서도 상위권에는 위치하였지만 h-지수 외의 지표에서는 가장 높은 순위에 있지 않았다. 17번 출원인(MIT)의 경우, h-지수의 경우 외는 달리 g-지수에서는 가장 높은 값을 가지게 되었는데, 이는 h-지수를 계산하였을 때 피인용빈도가 최상위인 논문들의 영향력이 가려졌던 것이 g-지수를 계산하면서 드러난 것으로 보인다.

h-지수의 파생지표인 hT-지수, hS-지수, hA-지수는 서로 유사한 결과로 나타났다. hT-지수와 hS-지수의 경우, 가장 높은 값을 가지는 출원인은 g-지수의 경우와 마찬가지로 17번 출원인(MIT)으로 나타났다. 한편, hA-지수의 경우, 20번 출원인(Pfizer)이 가장 높은 값을 보였다. CPP의 경우, 21번 출원인(pSivida)이 모든 지표에서 낮은 값을 가짐에도 불구하고 CPP에서는 상위권에 위치하였는데, 이는 해당 특허출원인의 총 특허수(N)가 다른 출원인에 비해 상대적으로 적었던 점을 보아, CPP가 양적 수준 즉, 생산성을 반영하는데 한계가 존재한다는 것을 보여준다.

<표 2> 각 특허 출원인의 계량서지학적 지표의 값과 순위(괄호 안은 순위를 의미)

순번	특허 출원인	N	TC	Max	CPP	h-지수	g-지수	hT-지수	hS-지수	hA-지수
1	Abbott	100	6310	611 (5)	63.10 (2)	48 (1)	76 (2)	72.62 (2)	444.84 (2)	30.65 (3)
2	AbbVie	100	5143	611 (5)	51.43 (7)	43 (5)	68 (6)	65.55 (5)	375.45 (5)	28.30 (4)
3	Allergan	100	4105	151 (18)	41.05 (13)	41 (10)	55 (12)	61.02 (10)	313.2 (12)	26.03 (11)
4	Alza	100	3388	111 (23)	33.88 (19)	39 (11)	53 (13)	55.95 (13)	295.08 (13)	24.29 (12)
5	Antecip Bioventures II	47	828	99 (24)	17.62 (25)	21 (21)	29 (25)	26.95 (24)	121.39 (24)	11.27 (23)
6	Baxter	100	2743	85 (25)	27.43 (21)	30 (17)	44 (18)	48.95 (16)	212.06 (18)	20.10 (17)
7	Bayer	100	3581	169 (17)	35.81 (17)	34 (14)	53 (13)	55.31 (14)	268.37 (14)	22.78 (14)
8	Bend Research	55	2041	197 (14)	37.11 (14)	28 (18)	45 (17)	41.75 (18)	215.39 (17)	17.71 (18)
9	Boehringer Ingelheim	100	4591	184 (16)	45.91 (10)	42 (7)	62 (10)	64.29 (6)	350.05 (9)	27.75 (6)
10	Daiichi Sankyo	100	1972	130 (21)	19.72 (23)	25 (19)	37 (21)	40.62 (19)	160.57 (19)	16.20 (19)
11	Euro Celtique	100	4153	206 (11)	41.53 (12)	38 (12)	63 (9)	59.77 (11)	343.80 (11)	26.99 (7)
12	Genentech	100	5887	678 (3)	58.87 (3)	44 (4)	74 (3)	68.95 (4)	415.68 (4)	27.76 (5)
13	Hanmi	100	1793	203 (13)	17.93 (24)	24 (20)	36 (22)	37.71 (20)	154.48 (21)	14.68 (20)
14	Intarcia Therapeutics	29	1031	149 (19)	35.55 (18)	18 (23)	33 (24)	28.12 (23)	125.69 (23)	13.25 (22)
15	Johnson and Johnson	100	4819	419 (8)	48.19 (9)	42 (7)	65 (8)	63.19 (9)	357.53 (8)	26.98 (8)
16	Mannkind	68	2493	135 (20)	36.66 (16)	34 (14)	48 (16)	48.13 (17)	248.38 (16)	22.18 (15)
17	Massachusetts Institute of Technology (MIT)	100	8023	1091 (1)	80.23 (1)	45 (3)	89 (1)	74.98 (1)	486.73 (1)	31.45 (2)
18	Novartis	100	4483	187 (15)	44.83 (11)	42 (7)	61 (11)	63.45 (8)	346.92 (10)	26.60 (9)
19	Noven Pharmaceuticals	31	1145	296 (9)	36.94 (15)	14 (25)	34 (23)	24.98 (25)	107.26 (25)	8.44 (24)
20	Pfizer	100	5789	204 (12)	57.89 (5)	47 (2)	72 (5)	72.23 (3)	435.61 (3)	32.03 (1)
21	pSivida	32	1877	510 (7)	58.66 (4)	17 (24)	44 (18)	28.70 (22)	148.57 (22)	7.03 (25)
22	Samyang	79	1753	222 (10)	22.19 (22)	20 (22)	42 (20)	35.60 (21)	157.70 (20)	13.66 (21)
23	Seattle Genetics	100	5315	760 (2)	53.15 (6)	34 (14)	73 (4)	58.80 (12)	366.96 (6)	20.82 (16)
24	The Johns Hopkins University	100	3384	124 (22)	33.84 (20)	36 (13)	50 (15)	55.22 (15)	266.87 (15)	23.47 (13)
25	University of California	100	4937	675 (4)	49.37 (8)	43 (5)	66 (7)	63.82 (7)	361.85 (7)	26.55 (10)

4.2 스피어만 등위상관계수(Spearman's rank correlation)를 통한 분석

h-지수 및 파생지표들 사이의 순위에 대한 상관관계를 파악하기 위해 스피어만 등위상관계수를 활용하여 분석을 진행하였다(〈표 3〉 참조). 모든 관계는 통계적으로 유의미한 관계를 나타내었다($p < 0.05$). 각 관계 쌍 중에서 g-지수와 hS-지수가 가장 유사한 순위를 나타내었으며(0.985), g-지수와 hA-지수의 경우 가장 유사하지 않은 결과가 나타났다(0.894). hT-지수는 다른 4개의 지표들과 평균 0.968의 값을 나타내어 관련성이 상대적으로 높았고, 이와 반대로 g-지수는 다른 4개의 지표들과 평균 0.933의 값을 나타내어 상대적으로 관련이 적었다.

또한, 각 출원인의 영향력을 나타낼 수 있는 Max, CPP 및 TC를 기준으로 5개의 지표에 대한 스피어만 등위상관계수를 계산하였다(〈표 4〉 참조). 모든 지표에 대해 TC, CPP,

Max 순으로 높은 상관 관계를 보였고, Max와 CPP, TC 모두 g-지수, hS-지수, hT-지수 순으로 상관 관계가 높게 나왔다. 특히, 특허문서 중 가장 피인용이 많이 된 문서의 피인용빈도 값을 의미하는 Max가 g-지수와 높은 상관 관계를 가진다는 것은 피인용빈도가 최상위인 논문들의 영향력을 잘 반영하는 g-지수의 특징을 보여준다. 한편, h-지수와 hA-지수는 Max에 대해서는 통계적으로 유의한 관계를 나타내지 않기 때문에 Max값의 순위를 반영하기 어려운 것으로 보인다($p > 0.05$).

4.3 인용 네트워크를 통한 중심성 분석

특허데이터 내에서 25개 출원인의 인용 관계를 파악하여 네트워크 행렬을 작성하였고(〈표 5〉, 〈표 6〉 참조), 이를 토대로 인용 네트워크 그래프를 생성하였다(〈그림 1〉, 〈그림 2〉 참조). 개별 문서 단위가 아닌 각 출원인의 전체 특허데이터를 기준으로 하여 서로 몇 번

〈표 3〉 h-지수 및 파생 지표들 간의 상관관계 분석

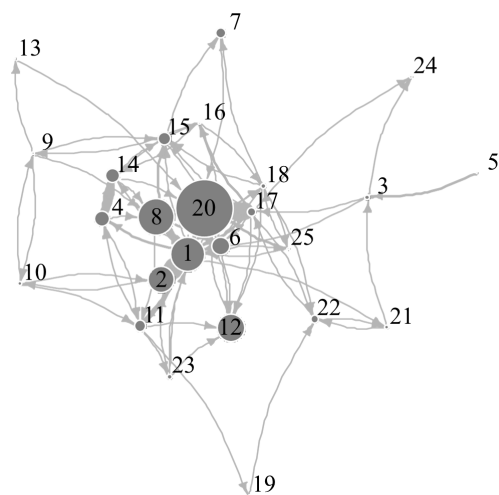
	h-지수	g-지수	hT-지수	hS-지수	hA-지수
h-지수		0.906	0.980	0.950	0.972
g-지수	0.906		0.946	0.985	0.894
hT-지수	0.980	0.946		0.976	0.972
hS-지수	0.950	0.985	0.976		0.936
hA-지수	0.972	0.894	0.972	0.936	

〈표 4〉 h-지수 및 파생지표와 Max, CPP 및 TC에 대한 상관관계 분석

	h-지수	g-지수	hT-지수	hS-지수	hA-지수
Max	-	0.646	0.480	0.567	-
CPP	0.672	0.838	0.740	0.776	0.654
TC	0.898	0.987	0.950	0.982	0.890

<표 5> 약물연구 분야 내 인용 네트워크 행렬

		인용된 특허 출원인																								
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
인용한 특허 출원인	1	0	30	0	5	0	0	0	12	0	0	9	0	0	3	2	0	1	1	0	6	1	0	0	0	1
	2	13	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1	0
	4	0	0	0	0	0	0	0	1	0	0	1	0	0	8	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	8	0	0	0	0	0	5	0	0	0	0	2	0	0	0	4	0	0	0	0	5	0	0	0	0	0
	9	0	0	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
	10	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	2	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	14	0	0	0	15	0	0	0	1	0	0	0	0	0	0	2	1	0	0	0	1	0	0	0	0	0
	15	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	0
	17	5	0	0	0	0	3	0	0	0	0	0	1	0	0	2	4	0	0	0	29	0	3	0	1	0
	18	0	1	0	0	0	0	1	0	0	0	0	2	0	0	1	0	0	0	0	0	0	1	0	0	1
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	20	2	2	0	0	0	0	0	6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	21	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0
	23	1	3	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	1	0	0	0	0	0
	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	0	0	0	0	0



<그림 1> 약물연구 분야 내 인용 네트워크

인용하였는지를 정리하였으며, 인용 네트워크는 약물연구 분야 내 인용(Family-to-family citation)과 약물연구 분야와 기타 분야 간 인용(Non-family-to-family citation)으로 구분하여 각각의 분석을 수행하였다. <표 5>와 <그림 1>은 같은 약물연구분야와 관련된 범위 내에서의 인용관계를 나타내고, <표 6>과 <그림 2>는 다른 분야에서의 인용관계를 나타낸다. 각 노드의 크기는 페이지랭크 값을 반영하여 표현하였으며, 링크를 나타내는 화살표의 방향은 인용을 하는 출원인에서 피인용된 출원인으로 향한다.

<그림 1>의 그래프에서 20번(Pfizer), 8번(Bend Research), 1번(Abbott) 출원인이 높은 영향력을 나타낸다. 특히, 8번 출원인의 경우 본 연구에서 다른 데이터에서 총 특허수가 55개로 상대적으로 적었고, 여러 계량서지학적 지표들의 순위도 14위에서 18위 사이로 나타나는 등 영향력이 낮게 측정이 되었는데도 불구하고, 인용 네트워크 상에서는 높은 영향력을 보였다. 한편, 높은 g-지수값을 보여준 17번 출원인(MIT)과 CPP에서 상위권을 차지한 21번 출원인(pSivida)과 같은 경우 25개 특허출원인 간 약물연구 분야 내 인용 네트워크 상에서는 높은 영향력을 보여주지 못하였다.

약물연구 분야와 기타 분야 간의 인용은 약물연구 분야 내 인용보다 현저히 연결 정도가 낮았다. 이는 25개의 출원인들이 주로 약물연구 분야에 대해서 특허를 출원하고, 이외의 다른 분야를 연구하는 경우는 상대적으로 적기 때문인 것으로 보인다. 인용이 되지 않고 고립된 노드들이 나타나며, 인용이 나타난 경우에도 두

그룹으로 단절되어 나타났다. 각 그룹에 속하는 노드들은 약물연구분야 외 다른 연구분야를 서로 공통적으로 가지는 노드들로 생각할 수 있다. 15번 출원인(Johnson and Johnson)과 4번 출원인(Alza)이 가장 높은 영향력을 보였고 그 뒤를 이어 20번(Pfizer), 22번(Samyang), 17번 출원인(MIT) 등이 높은 영향력을 보였다. 특히, 20번 출원인은 두 개의 인용 네트워크 모두에서 상위권의 영향력을 나타내었다는 점이 특징적이다. 15번 출원인의 계량서지학적 지표들의 순위는 7위에서 9위를 차지하였고, 17번 출원인의 경우 1위에서 3위를 차지하여 지표를 통해서도 높은 영향력을 보여주었다.

인용 네트워크의 중심성 지표와 h-지수 및 파생지표들 사이의 상관관계를 분석하기 위해 스피어만 등위상관계수를 활용하였다(<표 7> 참조). 약물 연구분야와 기타 분야 간 인용네트워크를 통해 구한 페이지랭크 값의 순위와 hT-지수의 순위에 대한 상관관계 분석을 제외하고($p=0.052$) 통계적으로 유의한 관계를 나타냈다($p<0.05$). 유의미한 관계를 나타내는 나머지 관계 쌍은 전부 양의 상관관계를 보였고, 특히, hS-지수가 상대적으로 높은 값을 나타내어, 가장 영향력을 잘 반영하는 것으로 판단된다. 한편, h-지수 및 파생지표에 대한 값은 25개의 특허출원인 외의 다른 출원인의 인용과 피인용에 대한 데이터를 포함하여 계산한 값인 반면, 페이지랭크 값은 25개의 특허출원인 간의 관계만을 반영한 값이다. 그럼에도 불구하고 이들 사이에 상관관계가 나타난다는 것은 25개의 특허출원인이 전체 약물 연구분야의 특허출원인 중에서도 영향력 있는 출원인임을 암시한다.

〈표 7〉 네트워크 중심성 지표와 h-지수 및 파생지표에 대한 상관관계 분석

	h-지수	g-지수	hT-지수	hS-지수	hA-지수
PageRank (분야 내)	0.396	0.440	0.427	0.449	0.447
PageRank (분야 간)	0.415	0.497	-	0.495	0.438

5. 결론 및 토의

본 연구를 통해 25개의 특허출원인의 특허 문서에서 나타나는 인용 관계를 확인하고 그 관계 속에서 발생하는 지식의 흐름과 영향력을 알아보려고 하였다. 이를 위해 여러 계량서지학적 지표들의 특징과 순위 변화를 확인하여 높은 영향력을 가지는 특허출원인을 확인하였다. 결과를 크게 계량서지학적 지표에서의 영향력, 약물연구 분야 내 인용 네트워크에서의 영향력, 약물연구 분야와 기타 분야 간 인용 네트워크에서의 영향력으로 나누어 정리를 하고자 한다. 1번 출원인(Abbott)의 경우 h-지수가 1위를 차지하였고, 분야 내 네트워크에서도 상위권을 차지하였으나, 기타 분야와의 인용 네트워크에서는 하위권에 위치하였다. 17번 출원인(MIT)의 경우 g-지수, hT-지수, hS-지수에서 1위를 차지하였고, 기타 분야와의 인용 네트워크에서는 상위권에 위치하였으나, 분야 내 네트워크에서 낮은 영향력을 보여주었다. 20번 출원인(Pfizer)은 hA-지수에서 1위를 차지하였고, 이외의 모든 지수에서도 상위권을 차지하였다. 또한, 20번 출원인은 두 개의 네트워크 모두에서 높은 영향력을 나타냈다. 반면, 8번 출원인(Bend Research)은 5개의 계량서지학적 지표에서는 영향력이 낮게 판단되었으나

분야 내 네트워크 상에서는 상위권이였다. 기존에 특허 문서의 영향력을 판단하는 지표인 CPP가 대체로 다른 계량서지학적 지표의 순위를 반영하는 듯하나, 21번 출원인(pSivida)과 같이 다른 계량서지학적 지표의 순위와 네트워크에서 낮은 영향력을 보여준 사례도 나타나므로 양적 수준을 반영하지 못하는 CPP의 한계를 알 수 있었다.

지표간의 관계는 g-지수와 hS가 서로 유사하였고, TC, CPP, Max의 순위를 가장 잘 반영하는 지표로 나타났다. 한편, TC, CPP, Max 순으로 5개의 계량서지학적 지표와의 상관관계가 높게 나타났다. 이는 선행연구(이재윤, 2006)에서 h-지수 또는 hA-지수가 g-지수 및 hS-지수에 비해 최대인용빈도(Max)와의 상관관계가 낮게 나타났으며, g-지수가 총인용빈도(TC)보다 최대인용빈도(Max)와 더 높은 상관관계를 보였던 결과와도 동일하다. 인용 네트워크에서 활용한 페이지링크의 경우 hS-지수와 가장 높은 상관관계를 보였다.

특허 인용을 분석하기 위한 지표에 따라 특정 출원인의 영향력에 대한 분석이 달라질 수 있기 때문에, 분석 목적에 따라 지표를 선택해야 한다고 생각된다. 예를 들어, g-지수의 경우 다른 지표에 비해 상위권의 영향력이 지나치게 반영된다는 특징이 있지만, 해당 출원인

이 출원한 특허의 전반적인 인용빈도보다 최상위권의 특허가 더 중요한 분석에서는 이것이 장점이 될 수도 있을 것이다.

또 다른 유용점으로는, 네트워크 분석을 통해 출원인 간의 연관성을 직, 간접적으로 파악할 수 있고, 특허 family 간의 연결도 구할 수 있기 때문에 이를 확장한다면 계량서지학 지표만으로는 직관적으로 알기 어려운 출원인 간의 관계를 이해하는데 도움이 될 것이다.

한편, 본 연구의 경우, 인용이 많이 되면 영향력이 높은 문서로 가정하여 진행이 되었지만, 실제로는 논문의 경우와 마찬가지로, 인용이 많이 된다고 해서 반드시 중요도가 높다고 할 수는 없다. 또한, 기술적인 면에서 특허문서는 논문에 비해 최신성이 더욱 강조되는 특징이 있으므로 최근의 특허일수록 더 많이 인용할 수도 있는데, 본 연구에서는 최신성을 고려하지는 않았다는 한계가 있다. 그리고, 산업분야 별 특허

의 출원등록형태의 다양성을 고려하지 않으면 정확한 분석이 어려울 수 있다. 한편, 본 연구에서 특허 주제분야를 CPC에 의존해서 검색어를 구성하였는데, 적절한 CPC가 없는 경우 데이터 수집에 어려움이 있을 수 있다.

한편, 특허 인용은 참고 문헌에 대한 중요한 데이터 소스임에도 불구하고 아직까지는 대규모 연구에 사용하기 어렵다는 단점이 있다(Kousha & Thelwall, 2015). 특허데이터에 대한 연구를 위해서는 Google Patents 또는 Derwent World Patents Index와 같은 특허 데이터베이스에 수동적으로 검색을 하게 되어 데이터 수집에 시간이 많이 걸리는데, 게다가 이러한 데이터베이스는 학술 인용 색인을 포함하지 않고 있기 때문에 다양한 연도 및 분야에 걸친 연구에는 한계가 존재한다. 또한, 결과를 최신 상태로 유지하기 위해 반복적으로 수집의 단계를 거쳐야 한다는 한계가 있다.

참 고 문 헌

- 서진이, 권오진, 노경란, 김원중 (2006). 문헌인용색인에 기반한 특허인용정보 활용에 관한 연구. 한국콘텐츠학회 종합학술대회 논문집, 4(2), 657-661.
- 엄익천, 박진서, 김봉진 (2019). 한국 신약개발의 기술적 역량 진단과 개선방안. 과학기술정책, 2(2), 107-137.
- 이민정, 김용대, 장원철 (2016). 특허 인용 네트워크 분석. 한국통계학회, 29(4), 613-625.
<http://doi.org/10.5351/KJAS.2016.29.4.613>
- 이재윤 (2006). 연구성과 측정을 위한 h-지수의 개량에 관한 연구. 정보관리학회지, 23(3), 167-186.
<https://doi.org/10.3743/KOSIM.2006.23.3.167>
- Anderson, T. R., Hankin, R. K., & Killworth, P. D. (2008). Beyond the durfee square: enhancing the h-index to score total publication output. *Scientometrics*, 76(3), 577-588.

- <https://doi.org/10.1007/s11192-007-2071-2>
- Egghe, L. (2006). Theory and practice of the g-index. *Scientometrics*, 69(1), 131-152.
<https://doi.org/10.1007/s11192-006-0144-7>
- Jaffe, A. B. & De Rassenfosse, G. (2019). Patent citation data in social science research: overview and best practices. In Depoorter, Ben & Menell, Peter S. eds. *Research handbook on the economics of intellectual property law*. Cheltenham: Edward Elgar Publishing, 20-46.
<https://doi.org/10.4337/9781789903997.00043>
- Kousha, K. & Thelwall, M. (2015). Patent citation analysis with Google. *Journal of the Association for Information Science and Technology*, 68(1), 48-61. <https://doi.org/10.1002/asi.23608>
- Nakamura, H., Suzuki, S., Kajikawa, Y., & Osawa, M. (2015). The effect of patent family information in patent citation network analysis: a comparative case study in the drivetrain domain. *Scientometrics*, 104(2), 437-452. <https://doi.org/10.1007/s11192-015-1626-2>
- R Development Core Team (2011). *R: a language and environment for statistical computing*. Vienna: R foundation for statistical computing. Available: <http://www.R-project.org/>
- Ruhnau, B. (2000). Eigenvector-centrality—a node-centrality? *Social networks*, 22(4), 357-365.
[https://doi.org/10.1016/S0378-8733\(00\)00031-9](https://doi.org/10.1016/S0378-8733(00)00031-9)
- U.S. Patent and Trademark Office & European Patent Office (2015). *Guide to the CPC*. Available: <http://www.cooperativepatentclassification.org/publications/GuideToTheCPC.pdf>
- Walter, D (2014). Patent-to-patent versus patent family-to-patent family citations and the impact of an invention. *Pharmaceutical patent analyst*, 3(3), 219-221.
<https://doi.org/10.4155/ppa.14.20>
- Yan, E. & Ding, Y. (2011). Discovering author impact: A pagerank perspective. *Information Processing & Management*, 47(1), 125-134. <https://doi.org/10.1016/j.ipm.2010.05.002>
- Zhang, J. & Luo, Y. (2017). Degree centrality, betweenness centrality, and closeness centrality in social network. In 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017). Atlantis Press, 300-303.

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Lee, Jae Yun (2006). Some improvements on h-index: measuring research outputs by citations. *Journal of the Korean Society for Information Management*, 23(3), 167-186.
<https://doi.org/10.3743/KOSIM.2006.23.3.167>

- Lee, Minjung, Kim, Yongdai, & Jang, Woncheol (2016). Patent citation network analysis. *The Korean Journal of Applied Statistics*, 29(4), 613-625.
<http://doi.org/10.5351/KJAS.2016.29.4.613>
- Seo, Jinny, Kwon, Oh-jin, Noh, Kyung-Ran, & Kim, Wan-jong (2006). A study on the application of the patent citation information based on the journal citation index. *The Korea Contents Society*, 4(2), 657-661.
- Um, Ikcheon, Park, Jinseo, & Kim, Bongjin (2019). Diagnosis and improvement of technical capabilities of Korean drug development. *Science and Technology Policy Institute*, 2(2), 107-137.