

# 키워드기반 특허 네트워크 진화에 따른 동종성 분석

## Analysis of Assortativity in the Keyword-based Patent Network Evolution

최진호                      김정욱\*  
Jinho Choi                Junguk Kim

### 요약

우리가 살고 있는 세계에는 다양한 네트워크들이 발견된다. 특히, 기술 및 학문과 밀접하게 관련 있는 지식 네트워크는 지식이 생산되는 방식을 이해하는데 도움을 주기 때문에 큰 의미를 갖는다. 이러한 중요성을 바탕으로 지금까지 지식 네트워크를 대상으로 한 많은 네트워크 분석들이 이루어져 오고 있다. 그 중에서 동종성 계수는 네트워크 내의 노드들이 비슷한 성향을 가진 노드들과 연결을 맺으려는 경향 수치로 나타낸다. 동종성 계수가 가지는 이러한 특성은 지식 네트워크로 간주 될 수 있는 키워드기반 특허 네트워크에서 기술이 어떻게 진화하는지 확인 하는데 도움을 줄 수 있다. 왜냐하면 지식 네트워크내 노드로 표현되는 키워드들 간의 관계들이 기술이 만들어지는 구조를 나타내기 때문이다. 본 연구에서는 키워드 네트워크에는 핵심 노드가 존재한다는 기존 연구 결과를 기반으로 두 가지 가설을 세우고 이에 대한 검증으로 동종성 분석을 수행 하였다. 첫 번째 가설은 키워드 기반 특허 네트워크는 시간 흐름에 따라 비동종성을 띠는 것으로 예측 하며, 동종성 분석을 통해 특허 네트워크가 진화함에 따라 비동종성을 보이는 것을 확인 하였다. 다음으로, 키워드 기반 특허 네트워크가 비동종성을 보일수록 클러스터링 계수 또한 낮아 질 것으로 예측하는 두 번째 가설에 대한 동종성 분석 결과, 네트워크의 동종성 계수가 낮아질수록 클러스터링 계수 또한 낮아진다는 사실을 확인 할 수 있었다. 또한, 두 번째 가설의 검증과정에서 확인 한 흥미로웠던 결과로써, 동종성 계수가 감소함에 따라 클러스터링 계수가 낮아지는 정도는 네트워크가 동종성을 보일 때 보다 비동종성을 보일 때가 훨씬 높았다.

주제어: 키워드 기반 특허 네트워크, 네트워크 진화, 동종성, 클러스터링 계수

### ABSTRACT

Various networks can be observed in the world. Knowledge networks which are closely related with technology and research are especially important because these networks help us understand how knowledge is produced. Therefore, many studies regarding knowledge networks have been conducted. The assortativity coefficient represents the tendency of connections between nodes having a similar property as figures. The relevant characteristics of the assortativity coefficient help us understand how corresponding technologies have evolved in the keyword-based patent network which is considered to be a knowledge network. The relationships of keywords in a knowledge network where a node is depicted as a keyword show the structure of the technology development process. In this paper, we suggest two hypotheses based on the previous research indicating that there exist core nodes in the keyword network and we conduct assortativity analysis to verify the hypotheses. First, the patents network based on the keyword represents disassortativity over time. Through our assortativity analysis, it is confirmed that the knowledge network shows disassortativity as the network evolves. Second, as the keyword-based patents network becomes disassortative, clustering coefficients become lower. As the result of this hypothesis, we confirm the clustering coefficient also becomes lower as the assortative coefficient of the network gets lower. Another interesting result concerning the second hypothesis is that, when the knowledge network is disassortative, the tendency of decreasing of the clustering coefficient is much higher than when the network is assortative.

☞ keyword : Keyword-based patent network, network evolution, assortativity, clustering coefficient.

## 1. 서론

우리가 살아가고 있는 세계는 셀 수 없을 정도로 다채로운 네트워크들을 가지고 있다. 네트워크란 행위자와 그

들간의 관계가 관찰되는 구조적 환경 말하는데[1], 구성 요소로는 노드(node or vertex)와 연결(link or edge)이 있으며 대표적인 예로 사회 네트워크[8,21], 전력망 네트워크 [5,10] 단백질 네트워크[11]를 들 수 있다. 다양한 네트워크 중에서 지식 네트워크는 대학교, 전문가 집단과 같은 사회적 구조와 연관되어 공유되고 구성된 지식 구조를 뜻한다[2]. 지식네트워크에서 노드는 사람이 아니라 사람이 만들어낸 논문[15], 특허[17,23,25], 키워드[24]와 같은

1 School of Business, Sejong University, Seoul, 143-747, Korea.

\* Corresponding author (jukim@sejong.ac.kr)

[Received 30 July 2013, Reviewed 1 August 2013, Accepted 9 November 2013]

무형가치들로 대표된다. 그렇기 때문에 지식네트워크가 어떻게 형성되고 어떠한 구조적 특성을 지니는지 확인할 수 있다면, 인간이 만들어내는 지식이 어떠한 방식으로 만들어지는지 알게 되므로 유용한 지식 혹은 기술을 개발 하는데 상당한 도움을 얻을 수 있다.

지금까지 제시되어온 네트워크 특성을 밝히는 분석 지표들 중에서[20] 이 제시한 동종성 분석이 있다. 동종성은 노드의 연결 선수를 기준으로 비슷한 수준의 연결 선수를 가지는 노드들끼리 연결을 맺는 경향을 의미한다 [20,21]. 반대로, 많은 연결 선수를 가지는 노드가 적은 연결 선수를 가진 노드와 연결 맺는 경향은 비동종성으로 정의한다. 앞서 말했듯이 세상에 존재하는 네트워크는 무궁무진하다. 각 네트워크마다 지니고 있는 특성 역시 다채롭다. 그 중에서도 우리는 동종성을 활용해 지식 네트워크가 어떤 특징을 보이며 연결되어 있는지를 계량적으로 표현할 수 있다면 지식 네트워크가 지니고 있는 본질적인 성질을 수치적으로 표현할 수 있으므로 지식이 만들어 지는 원리를 깊게 이해할 수 있다.

본 연구에서는 지식네트워크로 간주할 수 있는 키워드 기반 특허 네트워크에서 동종성이 어떤 패턴(동종성 혹은 비동종성)을 가지고 있는지 확인하기 위해, 두 가지 가설을 세우고 검증을 위한 분석을 실시한다. 첫 번째는 가설은 지식네트워크는 시간이 흘러감(evolution)에 따라 점점 더 비동종성 될 것을 보일 것으로 예측한다. 왜냐하면 특허 키워드 네트워크는 시간이 흘러감에 따라 지속적으로 핵심 역할을 하는 키워드들이 존재하기 때문이다[9]. 이어서 두 번째 가설은 지식네트워크가 비동종성을 띄게 될수록 클러스터링 계수 또한 점점 더 낮아지는 경향을 보일 것으로 추측한다. 그 이유는 핵심 역할을 하는 키워드들이 네트워크내 하부 집합의 노드들과 연결을 많이 맺기 때문이다.

상기 연구내용을 위한 본 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련 있는 선행 연구를 소개한다. 3장에서는 키워드기반 지식네트워크 내에서의 동종성 관련 가설 2가지를 제시한다. 4장에서는 본 연구에서 사용되는 특허 데이터의 정제 방식과 가설을 검증하기 위한 분석 방식을 소개한다. 5장에서는 분석결과를 제시하여 가설에 대한 타당성을 검증한다. 마지막으로 6장에서는 본 연구가 의미 바와 한계점 그리고 본 연구의 향후 연구 방향을 제시한다.

## 2. 문헌연구

### 2.1 특허 네트워크 분석

특허는 기술구현 내용, 기술 분류코드, 인용정보, 소유자 정보 등으로 구성되어 있다. 이러한 특허를 분석한다는 것은 그 기술의 변화 및 트렌드, 수준, 상업적 가치를 파악할 수 있다는 의미이므로 매우 중요하다. 이러한 특허를 분석하는 방법으론 네트워크 기반과 키워드 기반으로 구분 할 수 있다[9]. 네트워크란 사람 혹은 조직으로 이루어진 집합이 서로간에 연결을 맺고 있는 사회 구조를 의미한다. 이러한 네트워크를 기반으로 특허분석을 하는 것은 특허의 인용정보, 유사도 같은 특징을 통해 네트워크를 구성하고 이를 네트워크 관점으로 분석한다는 것을 의미한다[23]. Yoon과 Park[24]은 특허간의 유클리디안 거리를 측정하여 네트워크를 구성한 다음 해당 기술분야에서 중요도가 높은 특허를 파악하였다. von Wartburg 외 [23]은 특허 인용 네트워크를 구성해 기술 클러스터 그룹을 파악하였다. 네트워크 기반 특허 분석은 특허들간의 상호연관성을 기반으로 분석함으로써 어떤 특허가 특정 기술 분야에서 영향력을 지니고 있는지에 대한 정보를 제공하고 기술분야의 거시적인 흐름을 파악 할 수 있다. 반면, 키워드 기반 특허 분석은 기술의 핵심정보를 파악 하는 것을 의미한다[9]. 텍스트 마이닝을 통해 특허의 키워드를 추출하고 그 연관관계를 분석함으로써, 키워드들이 어떤 기술요소들로 구성되어 있는지를 볼 수 있으며, 이를 통해 해당, 기술 분야에서 활용되는 기술을 효과적으로 파악 할 수 있다.

### 2.2 동종성 연구

네트워크 내에서 많은 연결을 맺은 노드들이 또 다른 많은 연결을 맺고 있는 노드들과 연결을 맺으려는 성향을 동종성이라 정의한다[20,21]. 반대로 많은 연결을 맺고 있는 노드가 적은 연결을 맺고 있는 노드와 연결을 맺으려는 성향을 비동종성이라 정의한다.

동종성 분석에 관련된 대표적인 연구는 다음과 같다. 지진 네트워크 연관성 및 동종성에 대한 연구[3], P2P 네트워크에서의 동종성 및 시간의 추이에 따른 동종성 출현의 동적과정에 대한 연구[19], 기업간의 동종성과 M&A 성과 간의 상관관계에 관한[12], 인적 네트워크 내의 응집도(클러스터링 계수)와 동종성이 전염병의 확산에 미치는 영향에 대한 연구[6], 동종성에서의 허브 역할에 대한

연구[18], 온라인 소셜네트워크와 비동종성간의 상관관계에 관한 연구[14] 등 다양한 학문분야에 걸쳐 다수의 연구들이 수행되었다.

### 3. 가 설

본 장에서는 지식 네트워크가 진화해 가는 과정에서 동종성과 관련해서 두 가지 가설을 제시하고자 한다. 제시되는 가설을 검증하기 위한 분석 방법론 및 분석 결과는 4장과 5장에 각각 제시하였다.

#### 3.1 가설 1: 시간적 추이에 따른 네트워크내 동종성의 변화

*지식 네트워크는 시간이 흘러감에 따라 점점 더 비동종성이 높아질 것이다.*

어떤 기술 분야가 새롭게 소개되면, 해당 분야에는 수많은 특허들이 개발된다. 그로 인해 그 분야는 점점 더 기술적으로 성숙해져 간다. 즉, 시간이 흘러감에 따라 새로운 기술들은 지속적으로 소개되고 축적된다. 키워드기반 특허 네트워크로 비유해보면, 초기에 형성된 네트워크는 시간이 흘러감에 따라 지속적으로 키워드들을 축적 시켜 나간다. 이러한 과정 속에서 특정 기술분야에는 핵심 역할을 하는 기술 개념군들의 역할이 뚜렷해 진다[2]. 그러므로 네트워크가 진화해나가는 과정 속에서 핵심 역할을 하는 허브 노드들은 기존 네트워크 내에서 중요도가 낮은 노드들과 새로이 관계 맺거나 새로 유입된 키워드들과 연결을 맺으면서 지속적으로 새로운 기술 개념을 만들어 낼 수 있도록 한다. 즉, 시간이 흘러감에 따라 연결 선 수가 높은 핵심 노드들이 연결선 수가 낮은 노드들과 연결을 맺을 경향이 높아질 것이며, 이는 비동종성의 증가로 이어지게 될 것임을 추측할 수 있다.

#### 3.2 가설 2: 네트워크내 동종성의 변화 및 클러스터링 계수 간의 상관관계

*지식 네트워크가 비동종성이 높아질 수록, 네트워크의 클러스터링 계수는 낮아지는 경향을 보일 것이다.*

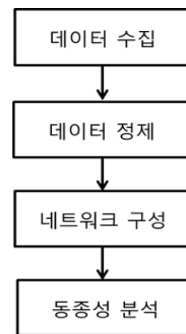
비동종성은 허브 역할을 하는 노드가 허브가 아닌 노드들과 연결을 맺는 사실을 수치로써 나타낸 자료다 [20,21]. 그렇기 때문에 네트워크가 점점 더 비동종성을

떨 수록 네트워크 내의 허브의 수가 증가 하거나 혹은 허브들이 계속해서 네트워크내의 하부집합에 존재하는 다양한 노드들과 연결을 맺어 나간다. 결국 허브로 인해 증가되는 연결들은 네트워크의 클러스터링 계수를 낮아지게 만들 것이다. 클러스터링 계수란 네트워크 내에서 노드들이 자신의 이웃들과 얼마나 잘 연결되는지 수량화한 값으로서, 네트워크 내에 직접적 연결 또는 중간 노드를 매개로 연결이 되어 있는 세 노드의 수 대비 직접적으로 연결이 된 닫힌 트리플렛(closed triplet)의 수로 계산된다[22]. 그러므로 허브역할을 하는 노드는 네트워크 내의 닫힌 트리플렛보다는 중개자의 역할로서, 간접적으로 연결을 맺는 세 노드의 수를 증가시킨다. 다시 말해서 허브 역할을 하는 노드는 교류가 없는 두 노드집합 간에 브릿지 역할을 할 가능성이 높다. 이로 인해 허브 노드는 자연스럽게 네트워크 내의 닫힌 트리플렛들의 비율을 떨어뜨리며, 네트워크의 클러스터링 계수가 낮아질 것으로 추측할 수 있다. 반대의 경우를 생각해 보더라도 네트워크가 진화함에 따라 점점 더 동종성을 띤다는 점은 비슷한 노드들 간의 연결성이 높아진다는 것을 뜻한다. 이러한 사실은 네트워크 내에 연결선 수가 비슷한 노드들끼리 집합을 이룰 확률이 높아진다는 점을 뜻하며, 이는 곧 네트워크가 동종성을 띄어 클러스터링 계수가 높아 질 수 있다는 점을 시사한다.

### 4. 분석방법론

#### 4.1 방법론

본 연구에서 제시하는 특허 키워드 네트워크 기반 기술예측은 아래 그림 1에서 제시된 바와 같이 네 단계로



(그림 1) 본 연구의 분석 과정  
(Figure 1) Analysis process of the research

나뉜다. 첫 번째 단계에서는 본 연구에 활용될 데이터인 특허들의 수집 방법을 소개한다. 두 번째 단계에서는 수집된 특허를 네트워크로 구축하기 위한 데이터를 정제 방법을 소개한다. 셋째 단계에서는, 둘째 단계를 거치고 난 정제된 데이터를 어떻게 네트워크로 구축하였는지에 대한 기준 및 방식을 설명한다. 마지막 넷째 단계에서는 본 연구에서 제시된 2가지 가설을 검증하기 위한 동종성 분석을 실시하고, 유의한 연구결과를 제시한다.

## 4.2 데이터 수집 단계

본 연구에서 사용된 분석 대상 자료는 LED(Light Emitting Diode) 분야의 특허로서, 미국특허청 (U.S. Patent and Trademark Office, USPTO)에서 출원된 국제특허들을 대상으로 하였다. 추출된 특허는 WIPS에서 검색어 ((light adj emi\* adj diode) LED) and (fluorescen\* adj material\*)를 통해 검색된 2000년부터 2012년까지 출원된 LED 관련 347개의 국제 특허이다. LED분야를 선택한 이유는 LED 분야가 지금까지 상당히 발전해왔으며, 이로 인해 새로운 특허가 많이 생성되고 있으며, 특허 정보가 결합되는 과정을 효과적으로 볼 수 있기 때문이다. 첫 번째 단계는 데이터를 수집하는 과정으로서, 먼저 기술을 예측하고자 하는 특허분야를 정하고, 해당 분야 내 특허데이터를 수집한다. 앞서 설명한 것처럼 본 연구에서는 특허 정보의 결합 과정을 효과적으로 파악하기 위해 새로운 특허가 많이 도출되는 LED 분야를 분석 대상으로 선정하고, 2000년부터 2012년까지 출원된 LED분야의 특허 347개의 정보를 수집하였다.

## 4.3 데이터 정제 단계

두 번째 단계로서 자연어로 구성된 347개의 LED분야 특허초록에서 키워드를 추출하고, 이를 표준화 과정을 통해 정제하는 과정이다. 특허 초록은 자연어로 구성되어 있기 때문에 먼저 텍스트 마이닝을 통해 특허 별로 중요한 키워드들을 추출한다. 다음으로 추출된 키워드 데이터들에 대해 해당 산업의 전문가의 도움을 받아 키워드들에 대한 표준화 작업을 수행하였다. 표준화 작업을 하는 이유는 텍스트 마이닝을 통해 추출된 키워드가 서로 동일한 뜻을 의미함에도 불구하고, 다른 단어로 쓰여져 있는 경우가 있기 때문이다. 표준화 규칙은 크게 아래의 세 가지 규칙을 기준으로 변경하였다.

- 복수형을 단수형으로 바꿈: 예) DISPLAY SUBFIELDS → DISPLAY SUBFIELD
- 하이픈을 제거함: 예) SECOND-HARMONIC LIGHT → SECOND HARMONIC LIGHT
- 동의어를 하나로 통일함: 예) WHITE LED, WHITE LIGHT-EMIT DIODE → WHITE LED

## 4.4 네트워크 구성 단계

세 번째 단계에는 정제된 키워드들을 대상으로 네트워크를 구성한다. 앞서 언급한 것처럼 특허에서 추출된 각각의 키워드를 노드로 두고 키워드 간의 관계를 링크로 맺는 네트워크를 구축하였다. 이렇게 네트워크를 형성한 이유는 텍스트 마이닝을 통해서 추출된 키워드는 특허에서 빈도수나 문맥에 따라서 핵심이 되는 키워드이고[13]. 이 키워드들이 한 특허에서 도출되었다는 것은 이 키워드들이 핵심 특허 개념요소로 작용하여 하나의 특허, 즉 기술로 만들어 졌다는 것을 의미하기 때문이다. 이러한 네트워크 구성 방식에 기반하여, 지난 13년(2000-2012년) 동안 출원된 LED분야 특허들을 대상으로 가중치 네트워크[7]를 만들었다.

## 4.5 동종성 분석 단계

네 번째 단계는 가설 1 및 가설 2에 대한 검증을 위한 동종성 분석 작업을 수행한다.

가설 1의 검증을 위해서는 시간 흐름에 따른 동종성 계수를 계산해야 한다. 이를 위해 분석 대상이 되는 지식 네트워크를 시간 흐름에 따라 진화 시킬 필요가 있다. 왜냐하면 네트워크가 진화하는 과정 속에서 동종성 계수 변화를 포착할 수 있기 때문이다. 네트워크가 진화하는 방식은 다음과 같다. 먼저 1년 간격으로 특허 네트워크를 진화 시키기로 한다. 왜냐하면 분석을 활용되는 특허의 개수가 1년 평균 30개보다 약간 못 미치기 때문에 1년 이하로 주기로 설정하면 변화가 미미할 수도 있다. 그러므로 먼저 2000년도에 출원된 LED특허를 키워드 네트워크로 구축한다. 해당 네트워크는 2000년도의 LED 특허 키워드 네트워크를 나타내며, 이 시점의 동종성 계수를 기록한다. 이어서 1년만큼 네트워크를 진화시키기 위해 2001년도에 출원된 특허를 2000년도에 출원된 특허에 더해서 다음 네트워크로 구축한다. 이러한 방식으로 2012년까지 총 13회의 동종성 계수를 측정할 수 있으며 가설 1의 검증을 위한 기반으로 활용한다.

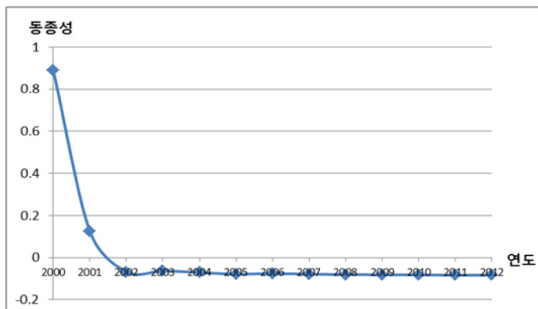
가설 2의 검증을 위해서는 동종성 계수의 다른 클러스터링 계수를 확인해야 한다. 이를 위해선 가설 1의 과정과 같이 시간 흐름에 따른 동종성 계수의 변화를 확인하고, 동시에 그에 따른 클러스터링 계수를 확인해야 한다. 위의 가설1 과정에서 설명한 바와 같이, 변화를 제대로 포착하기 위해 1년을 주기로 네트워크를 생성한다. 가설 1의 과정을 따라 생성된 네트워크에서 클러스터링 계수를 추가로 파악한다. 다시 말해서, 시간의 흐름에 따라 변화하는 동종성 계수와 클러스터링 계수와의 상관관계를 파악한다. 이를 통해, 가설 2의 검증을 위한 분석기반으로 활용한다. 이러한 과정을 거쳐 2012년까지 매년 출원된 특허데이터를 기존의 특허네트워크에 축적해서 네트워크 계수(동종성 계수, 클러스터링 계수)의 추이를 기록한다. 이를 통해, 가설검증에 필요한 네트워크가 시간 흐름에 따라 진화하는 과정을 포착할 수 있다.

## 5. 분석 결과

본 장에서는 앞서 4장에서 제시한 가설의 타당성을 검증하기 위해 시간이 흘러감에 따른 동종성 계수 변화 추이와 네트워크의 동종성 계수 변화에 따른 클러스터링 계수변화를 관찰한다.

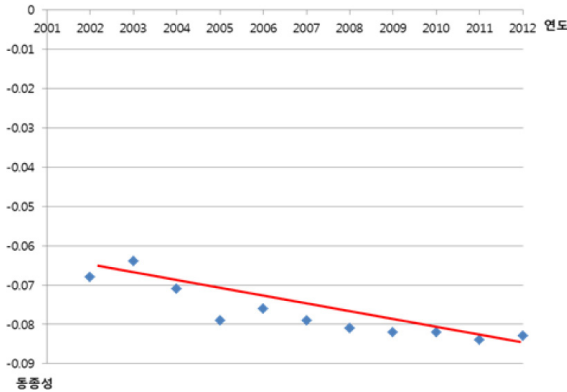
### 5.1 시간적 추이에 따른 네트워크내 동종성의 변화

(그림 2)와 (그림 3)은 가설 1을 검증하기 위한 분석 결과로서, 연도별 LED 특허 키워드 네트워크의 동종성 계수를 측정하였다. 이를 통해 동종성 관점에서 시간이 흘러감에 따라 특허키워드 네트워크가 어떻게 변해가는지 관측할 수 있다.



(그림 2) 시간 흐름에 따른 동종성 변화(2000~2012년)  
(Figure 2) The change of assortativity over time (2000~2012)

(그림 2)에서 알 수 있듯이, 초반(2000년, 2001년)에는 동종성 계수가 양수이다. 이는 초기 2년동안 네트워크 내의 노드들이 자신과 비슷한 연결선 수를 가지는 노드들과 연결을 맺는 경향이 강하다는 사실을 의미한다. 그러나 2002년부터는 동종성 계수가 비동종성으로 역전되며, 2012년까지 동종성 계수는 지속적으로 감소한다. 동종성 계수가 초반 2년 동안에는 동종성을 띄고 2002년부터는 비동종성을 띄는 이유는 본 연구에서 LED분야에 있는 특허 전체를 활용하지 못했기 때문이다. LED는 2000년 이전부터 개발되고 발전되어 온 기술이다. 하지만 (그림 2)의 2000년도에 기록된 동종성 수치는 2000년도 이전에 출원된 특허를 제외하고 당해 1년동안 출원된 특허만을 네트워크로 구성해 분석한 수치이다. 그러므로 1999년 이전부터 LED분야에서 핵심 역할을 하는 허브 노드들과 그들이 맺는 연결을 포착할 수 없다. 이는 곧 첫 두 해의 네트워크가 동종성을 띄게 만드는데 결정적인 역할을 한다. 그러나 2000년도 이후부터는 허브노드들이 만들어 내는 연결들을 추적할 수 있기 때문에, 이를 기반으로 분석한 결과 네트워크의 동종성 계수가 분석 기간 전반에 걸쳐 음의 값을 보임을 알 수 있다. 본 연구에서 제시된 가설1에 따르면 지식네트워크는 시간이 흘러감에 따라 점점 더 비동종성을 보일 것으로 예상하였으며, 다시 말해, 네트워크가 진화함에 따라 동종성 계수가 낮아져야 한다는 의미로서, 특정 시점에서 기록한 동종성 계수는 이전에 이전 시점에 기록된 동종성 계수보다 낮아야 가설 1을 만족할 수 있다. (그림2)에 나타나 있듯이 키워드기반 LED 특허 네트워크는 가설 1을 따르는 것으로 확인할 수 있다. 구체적으로 살펴보면, 2001년에 기록한 동종성 계수는 2000년도에 비해서 큰 폭으로 동종성 계수가 낮아졌다. 더 나아가, 2003년도 네트워크의 동종성 계수는 비동종성을 뿜 정도로 낮아진다. 그러므로 LED 특허 키워드 네트워크는 가설 1을 따른다고 볼 수 있다. 그렇지만 2002년 이후로는 (그림2)를 통해서 동종성 계수가 점점 더 낮아지는지 확인하기 쉽지 않다. 왜냐하면 초반 두 해의 계수 값이 나머지 해의 수치보다 상당히 높아 2002년 이후의 작은 변화를 확인하기 힘들기 때문이다. 그래서 우리는 시간 흐름에 따라 변하는 동종성 계수를 정밀하게 포착하기 위해 2000년과 2001년을 제외한 2002년부터 2012년까지의 동종성 계수를 분석하여 (그림 3)에 제시하였다.



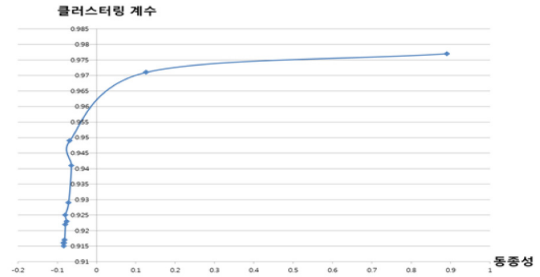
(그림 3) 시간 흐름에 따른 동종성 변화(2002~2012년)  
(Figure 3) The change of assortativity over time (2002~2012)

(그림 3)에서 초기 네트워크 형성 단계(2000~2001년)를 제외한 전체기간 (2002~2012년)에 대한 동종성 변화추이를 살펴 보면 LED 특히 키워드 네트워크가 가설1을 따른다는 사실을 좀더 명확하게 확인 할 수 있다. (그림 2)와는 달리, 동종성 계수가 시간의 흐름에 따라 뚜렷하게 비동종성을 나타낸다. 비동종성이란 연결선 수가 높은 노드가 연결선 수가 낮은 노드들과의 연결하는 경향을 의미하며, 이는 결국 지식 네트워크 내에 핵심 역할을 하는 노드가 시간이 흘러감에 따라 지속적으로 새로 유입된 키워드와 연결을 맺을 확률이 높다는 것을 시사한다. 여기서 연결선 수가 높은 노드는 자연스럽게 네트워크 내에서 핵심역할을 하는 키워드이며, 연결선 수가 낮은 노드들은 주로 매년 새롭게 유입된 키워드들일 가능성이 높다. 그러므로 키워드 기반 지식네트워크의 핵심 노드들은 시간이 흘러감에 따라 허브역할을 강화해 나간다는 사실을 확인할 수 있다.

## 5.2 네트워크내 동종성의 변화 및 클러스터링 계수 간의 상관관계

아래 제시된 (그림4), (그림 5)는 가설 2를 검증하기 위하여 그래프의 가로축을 동종성 계수로 세로축을 클러스터링 계수로 나타낸다. 이를 통해 네트워크가 비동종성로 변화함에 따라 클러스터링 계수가 어떻게 변하는지 추적할 수 있다. 가설 2에 대한 검증에 들어가기에 앞서 (그림 4)에 주목해 보면, 2000년부터 2012년도까지 출원된 LED 특히 키워드 네트워크는 항상 0.9 이상의 클러스터링 계수를 보인다. 이렇게 높은 클러스터링 계수는 인용네트워

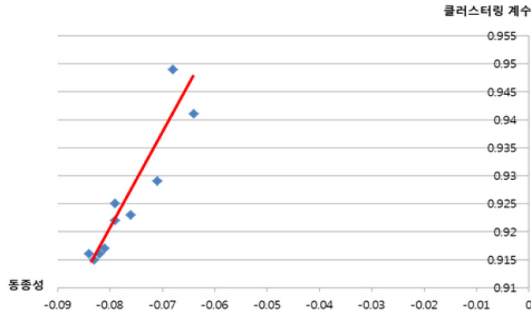
크[16], 키워드네트워크[9], 공동저자네트워크[4]등의 지식 또는 지식기반 네트워크가 가지는 공통적 특성임을 알 수 있다.



(그림 4) 동종성 계수 대비 클러스터링 계수(2000~2012년)  
(Figure 4) Assortativity coefficients vs. Clustering coefficients (2000~2012)

(그림 4)에 나타나 바와 같이, 동종성 계수가 낮아 질수록 클러스터링 계수도 낮아진다. 이는 가설 2에서 추측했듯이 허브 역할을 하는 노드가 많아지거나 그 역할이 강해져서 이러한 수치를 보이는 것으로 예상된다. 그 중에서 특이한 점은 동종성 계수가 양수 일 때는 클러스터링 계수가 1에 거의 수렴하고 양의 범위 내에서는 동종성 계수가 낮아지더라도 클러스터링 계수의 감소 폭이 크지 않다. 반면 동종성이 음수가 되면, 다시 말해 네트워크가 비동종성 일수록, 클러스터링 계수가 감소하는 폭이 매우 크다. 이는 비동종성 경향이 높을수록 허브 노드의 역할이 강화되기 때문에, 키워드 기반 특허 네트워크에서의 닫힌 트리플렛의 비율이 줄어들게 되며 결과적으로 클러스터링 계수가 급격히 낮아진다. 우리는 가설1을 명확하게 검증하기 위해 (그림 3)을 제시하여 2002년부터 2012년까지의 동종성 계수와 시간간의 상관관계를 정밀하게 살펴 보았다. 마찬가지로 동종성 계수와 클러스터링 계수 간의 상관관계 또한 세밀하게 확인해야 할 필요가 있으므로, 동종성 계수가 양수인 경우를 제외하고 음수인 데이터만으로 (그림 5)를 재구성 하였다.

(그림 5)는 앞에서 제시한 가설2를 강하게 뒷받침 해준다. 핵심 키워드들은 노드 집합들간에 허브 역할을 강화해 나가서 네트워크내의 노드 집합들 간에 다양한 네트워크적 경로를 제공한다. 그로 인해 밀집되어 닫힌 트리플렛을 보이는 노드 집합들은 허브들이 만들어낸 새로운 연결로 인해 적어진다. 지금까지 본 장에서는 앞서 제시한 가설을 검증하기 위해 두 가지 분석을 수행하였다. 이



(그림 5) 동종성 계수 대비 클러스터링 계수(2002~2012년)  
 (Figure 5) Assortativity coefficients vs. Clustering coefficients (2002~2012)

를 통해 우리는 키워드기반 특허 네트워크가 시간 흐름에 따라 비동종성을 띄는 것을 알 수 있었고, 이러한 현상이 네트워크의 클러스터링 계수를 지속적으로 낮아지도록 하였다. 여기에는 근본적으로 지식네트워크가 가지는 허브가 기인했기 때문으로 추측된다. 허브를 가지는 네트워크는 자연스럽게 핵심 노드가 상위 계층에 존재해 아래 하부 집합에 존재하는 노드들과 관계를 다양하게 맺기 때문이다. 그러므로 키워드기반 지식 네트워크는 자연스럽게 비동종성을 보이게 된다.

## 6. 결 론

본 연구는 키워드기반 LED 특허 네트워크를 통해, 지식네트워크의 연결 특성을 살펴보기 위해 두 가지 가설을 세우고 동종성 분석을 통해 이를 검증 하였다.

첫번째 가설은 특정 지식 네트워크에서 허브 역할을 하는 노드들이 지속적으로 중심적인 역할을 할 것으로 예상하여 네트워크가 진화함에 따라 점점 더 비동종성을 보일 것으로 예측하였다. 두번째 가설은 비동종성 증가 패턴은 허브노드의 역할을 증가시켜서, 네트워크 내의 닫힌 트리플렛 비율을 감소시킬 것으로 예상하였으며, 결국 네트워크가 비동종성을 나타낼수록 클러스터링 계수 또한 낮아 질 것으로 예측 하였다.

두 가정을 검증하기 위해서 우리는 1년을 주기로 특허 데이터를 축적시키는 방식으로 2000년부터 2012년까지의 특허 키워드 네트워크에 대해 동종성 분석을 수행하였다. 가설1에 대한 검증을 위해 시간적 추이에 따른 네트워크내 동종성의 변화를 분석하였다. 분석 결과 동종성 계수는 네트워크의 시간이 흘러감에 따라 낮아진다. 비록

모든 LED 분야 특허를 활용하지 못해서 초반 2년동안 동종성 계수가 양수이긴 했으나, 동종성 계수는 지속적으로 감소하여 비동종성을 나타냄을 확인 할 수 있었다. 가설 2에서는, 네트워크내 동종성의 변화 및 클러스터링 계수 간의 상관관계를 분석하였다. 분석 결과, 네트워크의 동종성 계수가 낮아질수록 클러스터링 계수 또한 낮아졌다. 이는 우리가 제시한 가설 2를 따른다고 볼 수 있다. 흥미로운 사실은 가설 2의 검증과정에서 동종성 계수가 양수 구간이나 음수구간 이나에 따라 클러스터링 계수가 감소하는 폭이 크게 차이 났음을 확인 할 수 있었다. 네트워크의 동종성 계수가 양수 구간의 경우에는 동종성 계수가 줄어 들더라도 클러스터링 계수에 대해 큰 변화가 없었으나, 동종성 계수가 음수 구간으로 접어들면서부터는 클러스터링 계수의 변화 수준이 동종성 계수 구간일 때에 비해 상대적으로 매우 높았다. 본 연구의 한계점으로는 데이터 수집 범위와, 특허 분야를 들 수 있다. 먼저 데이터 수집 범위 측면에서는, 2000년도부터 데이터를 모았기 때문에 이전에 형성 되어있던 네트워크가 어떤 특성을 보이는지는 확인이 불가능하다. 이로 인해 분석을 위해 구성한 LED특허 키워드 네트워크에서 동종성 계수가 초기 2년간은 양수로 계측 되었다. 그러나 본 사실은 특정 기술 분야의 원초적 시점을 파악하지 않고서는 제반 기술분야에서 나타날 수 있는 상황이며, 본 가설 1에서의 분석대상인, 시간적 추이에 따른 네트워크 동종성이 증가하는지, 줄어드는지 여부와는 관련성이 없다.

본 연구의 한계점 및 향후 연구방향은 다음과 같다. 본 연구에서는 시간 및 인력의 제약으로 인해 LED 한 분야만 분석하였다. 따라서 향후 연구에서는 보다 다양한 특허 분야에 대해 본 가설들을 적용하여 검증해볼 필요가 있다. 차기 연구에서는 성숙도가 다른 기술 분야도 다양하게 살펴봄으로써 특허 키워드 네트워크가 지니는 동종성 특성을 일반화 하도록 한다. 나아가 특허 외의 다른 지식 네트워크를 분석함으로써 본 연구에서 제시한 가설들이 모든 키워드 기반 지식네트워크에서 보이는 특성인지 여부를 분석함으로써, 지식 네트워크의 일반적 특성을 이해하는데 있어 중요한 연구 기반으로 활용될 수 있을 것으로 기대한다.

## 사사 (Acknowledgement)

이 논문은 2011년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2011-330-B00094)

## 참 고 문 헌 (Reference)

- [1] Abe, S., Suzuki, N., "Scaling relation for earthquake networks," *Phys. Rev. E*, 74, 026113, 2006
- [2] Albert, R., A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, Vol.74 No.1, pp. 47-97, 2002
- [3] Albert, R., I. Albert, G. Nakarado, "Structural vulnerability of the North American power grid," *Physical Review E*, Vol.69 No.2, 2004
- [4] Badham, J., R. Stocker, "The impact of network clustering and assortativity on epidemic behaviour," *Theor Popul Biol*, Vol.77 No.1, pp. 71-75, 2010
- [5] Bagler, G., "Analysis of the airport network of India as a complex weighted network," *Physica A: Statistical Mechanics and its Applications*, Vol.387 No.12, pp. 2972-2980, 2008
- [6] Batagelj, V., A. Mrvar, "Some analyses of Erdos collaboration graph", *Social Networks*, Vol.22 No.2: pp. 173-186, 2000
- [7] Choi, J., Kim, H., Im, N., "Keyword network analysis for technology forecasting", *Journal of Intelligence and Information Systems*, Vol.17 No.4, pp.227-240, 2011
- [8] Choi, J., N. Im, H. Kim, Y.-S. Hwang, "A study on the patent analysis for effective technology forecasting," *Information Science and Digital Content Technology (ICIDT)*, 2012 8th International Conference on, 2012
- [9] DJ, W., S. SH., "Collective dynamics of 'small-world' networks," *Nature*, Vol.393 No.6684, pp.440-442, 1998
- [10] Eisenberg, E., E. Levanon, "Preferential Attachment in the Protein Network Evolution," *Physical Review Letters*, Vol.91 No.13, 2003
- [11] Fabac, R., M. Schatten, T. Đuričin, "Social Network Mixing Patterns In Mergers & Acquisitions - A Simulation Experiment," *Business Systems Research*, Vol.2 No.1, pp. 36-44, 2011
- [12] Frantzi, K., S. Ananiadou, H. Mima, "Automatic recognition of multi-word terms: the C-value/NC-value method," *International Journal on Digital Libraries*, Vol.3 No.2, pp. 115-130, 2000
- [13] Hu, H.-B., X.-F. Wang, "Disassortative mixing in online social networks," *EPL (Europhysics Letters)*, Vol.86 No.1: 18003., 2009
- [14] Hummon, N. P., P. Dereian, "Connectivity in a citation network: The development of DNA theory," *Social Networks*, Vol.11 No.1, pp.39-63, 1989
- [15] Klemm, K., V. M. Eguiluz, "Highly clustered scale-free networks," *Physical Review E*, Vol.65 No.3: 036123, 2002
- [16] Li, X., H. Chen, Z. Huang, M. C. Roco, "Patent citation network in nanotechnology (1976 - 2004) ," *Journal of Nanoparticle Research*, Vol.9 No.3: pp. 337-352, 2007
- [17] Lusseau, D., M. E. Newman, "Identifying the role that animals play in their social networks," *Proc Biol Sci* 271 Suppl 6, pp. S477-481, 2004
- [18] Murai, F., D. R. Figueiredo., "Assortative Mixing in BitTorrent-Like Networks". *INFOCOM Workshops 2009, IEEE.*, 2009
- [19] Newman, M. E. J., "Assortative Mixing in Networks." *Physical Review Letters*, Vol.89 No.20: 208701, 2002
- [20] Newman, M. E., D. J. Watts, S. H. Strogatz, "Random graph models of social networks." *Proc Natl Acad Sci U S A* 99 Suppl 1, pp. 2566-2572, 2002
- [21] Soffer, S., A. Vázquez, "Network clustering coefficient without degree-correlation biases." *Physical Review E*, Vol.71 No.5, 2005
- [22] Sohn, D.W. *Social network analysis*, "Kyoungmoon publishing company, 2002
- [23] von Wartburg, I., T. Teichert, K. Rost, "Inventive progress measured by multi-stage patent citation analysis." *Research Policy*, Vol.34 No.10, pp. 1591-1607, 2005
- [24] Yi, S., J. Choi, "The organization of scientific knowledge: the structural characteristics of keyword networks." *Scientometrics*, Vol.90 No.3, pp. 1015-1026, 2011
- [25] Yoon, B., Y. Park, "A text-mining-based patent network: Analytical tool for high-technology trend." *The Journal of High Technology Management Research*, Vol.15 No.1, pp. 37-50, 2004



● 저 자 소 개 ●



**최 진 호(Jinho Choi)**

1996년 KAIST 산업경영학과 학사  
1998년 KAIST 테크노경영대학원 경영공학과 석사  
2006년 KAIST 테크노경영대학원 경영공학과 박사  
2008년~현재 세종대학교 경영학과 부교수  
관심분야 : 플랫폼 비즈니스, 지식네트워크



**김 정 욱(Jeongwook Kim)**

1980년 서강대학교 전자공학과 학사  
1982년 KAIST 경영과학과 석사  
1998년 KAIST 테크노경영대학원 경영공학과 박사  
1982년~1984년 한국개발연구원 연구원  
1984년~1996년 한국통신 선임연구원 /부장  
2000년~현재 세종대학교 경영학과 교수  
관심분야 : 인터넷마케팅, 전자상거래, 전자적자원관리