

# 간호학 학술논문의 주제 분석을 위한 텍스트네트워크분석방법 활용

박 찬 숙

건국대학교 간호학과 시간강사

## Using Text Network Analysis for Analyzing Academic Papers in Nursing

Park, Chan Sook

Instructor, Department of Nursing, Konkuk University, Chungju, Korea

**Purpose:** This study examined the suitability of using text network analysis (TNA) methodology for topic analysis of academic papers related to nursing. **Methods:** TNA background theories, software programs, and research processes have been described in this paper. Additionally, the research methodology that applied TNA to the topic analysis of the academic nursing papers was analyzed. **Results:** As background theories for the study, we explained information theory, word co-occurrence analysis, graph theory, network theory, and social network analysis. The TNA procedure was described as follows: 1) collection of academic articles, 2) text extraction, 3) preprocessing, 4) generation of word co-occurrence matrices, 5) social network analysis, and 6) interpretation and discussion. **Conclusion:** TNA using author-keywords has several advantages. It can utilize recognized terms such as MeSH headings or terms chosen by professionals, and it saves time and effort. Additionally, the study emphasizes the necessity of developing a sophisticated research design that explores nursing research trends in a multidimensional method by applying TNA methodology.

**Key Words:** Text mining, Nursing methodology research, Social networking

### 서 론

#### 1. 연구의 필요성

정보통신혁명에서 4차산업혁명시대로 변화의 흐름이 가속화되면서 다양한 형태의 텍스트가 만들어지고 웹사이트를 통해 전세계로 전파되고 있다. 텍스트는 문장이 모여서 만들어진 하나의 글 덩어리이며, 언어에 따라 다른 문법구조를 가지고 있다. 텍스트 내에서는 동일한 단어도 사용된 맥락에 따라 상이한 의미를 나타낸다.<sup>1)</sup> 텍스트의 종류는 인터넷과 모바일의 발달로 형태가 더욱 다양해지고 있다. 학술지에 발표된

학술논문부터 뉴스기사, 인터뷰 글, 다양한 장르의 잡지 기고문과 블로그, 유튜브, 페이스북, 트위터, 카카오톡 같은 소셜미디어까지 광범위하다. 최근에는 소셜미디어 중에서도 이용자들의 사용시간이 늘고 있는 소셜네트워킹서비스(Social Networking Service, SNS)인 트위터, 페이스북과 유튜브에서 텍스트를 수집하여 분석하는 텍스트네트워크분석 연구가 활발해지고 있다.<sup>2)</sup>

텍스트네트워크분석 연구설계는 “계량적 내용분석”에 해당된다.<sup>3)</sup> 텍스트를 수집하여 계량적 방법을 적용한 내용분석을 수행하므로 양적 연구로 분류된다. 내용분석은 텍스트를 분석하는 방법 중 하나인데 크게 정성적 방법과 양적 방법으로

**주요어:** 텍스트 마이닝, 간호 방법론 연구, 사회연결망

**Corresponding author:** Park, Chan Sook

Department of Nursing, Konkuk University, 268 Chungwon-daero Chungju 27478, Korea.  
Tel: +82-43-840-3953, Fax: +82-43-840-3958, E-mail: sook706@hanmail.net

Received: Feb 7, 2019 / Revised: Apr 11, 2019 / Accepted: Apr 12, 2019

구분한다.<sup>4)</sup> 전통적으로 내용분석은 대부분 정성적 방법이 적용되었으나, 컴퓨터 기술의 발달로 많은 양의 텍스트를 이용한 양적 내용분석이 가능해졌다. 양적 내용분석방법에 속하는 텍스트네트워크분석을 통해 연구자는 텍스트를 구성하고 있는 단어(words)들을 해체하여 의미 있는 개념을 추출하고 단어들이 어떤 관계를 형성하고 있는지 네트워크(연결망) 형태로 표현한다. 이를 통해 텍스트에 나타난 주요개념, 개념간의 관계, 주요개념을 포함한 경계 등을 파악하고 텍스트에서 명백히 표현된 맥락과 유추 가능한 맥락을 모두 드러낼 수 있다.<sup>5)</sup> 또한 양적으로 분석하여 발견된 지식이나 현상을 과학적 근거로서 제시하는데 유용하다. 따라서 텍스트네트워크분석은 텍스트에 명백하게 드러나 있는 의미구조뿐만 아니라 맥락 속에 내재되어 있는 잠재적 의미구조를 파악하기 위해 과학적인 방법을 엄격하게 적용하여 분석하는 방법론<sup>6)</sup>이라고 할 수 있다. 이를 통해 발견된 맥락적으로 통합된 지식을 선행연구자들은 “지식구조(knowledge structure)” 또는 “지식지도(knowledge map)”라고 부른다. 한편 텍스트 수집에 이용되는 웹 기반의 오픈 소스 라이브러리 패키지 프로그램들<sup>7)</sup>의 개발과 텍스트 자료 분석에 필요한 동시단어분석(co-word analysis)<sup>8)</sup>과 사회네트워크분석<sup>9)</sup> 방법을 수행할 수 있는 컴퓨터 프로그램들이 개발되고 발전되어 왔다. 이러한 프로그램들의 이용으로 과학적 객관성을 높이고 분석에 소요되는 시간, 노력, 비용을 단축할 수 있게 되면서 텍스트네트워크분석의 효용성이 높아지고 있다.

간호학에서 텍스트네트워크분석 초기연구에 해당되는 2003년 Baldwin 등<sup>10)</sup>의 “윤리와 치매 문헌 지도”, 2010년 Scott 등<sup>11)</sup>의 “간호에서 지식이용 연구 문헌 지도”를 살펴보면 계량서지학적 분석(bibliometric analysis) 방법을 적용하여 텍스트를 분석하였다. 계량서지학적 분석은 논문들의 서지정보를 양적으로 분석하는 방법이다. 오늘날 학술논문 검색을 위한 웹 데이터베이스에서 제공하는 서지 상세정보에는 논문의 제목, 저자이름, 출판 년도, 학술지 이름, 저자키워드, 초록 등이 있고 데이터베이스에 따라서는 검색된 많은 양의 논문서지정보에 대하여 “서지정보 상세 내보내기”가 가능하도록 기술적으로 지원하고 있어 수백에서 수만 편의 서지정보 수집이 가능해졌다. 이러한 기술적 진보로 과거에는 할 수 없었던 방대한 학술논문의 서지정보를 연구자료로 이용할 수 있게 되었고, 논문의 제목, 초록, 저자키워드, 저자이름, 연도, 인용문헌 등을 분리하여 계량함으로써 공저자, 동시인용, 핵심주제와 같은 여러 측면에서의 연구동향을 분석할 수 있다.

국내에서 간호학 논문에 대하여 텍스트네트워크분석방법을 적용한 초기연구로는 Lee 등<sup>12)</sup>의 “한국 간호학 연구주제

의 사회연결망분석”이 있었는데, 계량서지학적 방법을 이용하여 논문의 제목과 초록에 출현한 용어를 추출하여 10회 이상 출현한 용어(term)들을 선정하고 용어 간의 동시출현 행렬을 개발한 후 사회연결망분석을 적용하였다. 이처럼 계량서지학적 방법과 사회연결망분석방법을 혼합한 텍스트네트워크분석방법을 적용한 연구들이 발표되면서 여러 가지 사회네트워크분석 지표들이 이용되었다.<sup>3,12-16)</sup> 가령 연결정도(degree)와 Freeman의 3가지 중심성 지표<sup>17)</sup>인 연결중심성(degree centrality), 매개중심성(betweenness centrality), 근접중심성(closeness centrality)이나 컴포넌트(component), 커뮤니티(community)와 같은 응집분석(cohesion analysis)을 이용하였다. 그리고 단어의 출현빈도, 연결정도, 중심성 값 등을 근거로 핵심주제를 발견하여 순위에 따라 나열한 표뿐만 아니라, 직관적 탐색이 가능한 네트워크지도(sociogram, knowledge map)나 단어구름(word cloud) 등으로 시각화한 그림을 제시하였다.

간호학문의 지식체 형성에 필요한 근거자료를 생산하는데 있어 텍스트네트워크분석방법론이 유용한 하나의 연구방법으로 활용될 수 있다.<sup>12)</sup> 과학적 방법으로 얻어진 객관적 사실들을 토대로 귀납적 추론과정<sup>18)</sup>을 통해 정립한 간호이론은 다양한 반복 연구들에 의해 검증되면서 간호지식체가 완성<sup>19)</sup>된다. 텍스트네트워크분석방법을 적용한 양적 내용분석은 많은 양의 논문을 정량적으로 분석하고 발견된 지식구조를 정성적으로 해석함으로써 흩어져 있던 자료에서 일정한 패턴을 파악하여 간호지식체 형성 과정에 활용할 수 있으므로 방법론적 가치가 있다고 본다. 본 연구는 간호학 학술논문의 주제분석방법의 하나로 텍스트네트워크분석방법을 활용할 수 있도록 국내에서 간호학술논문에 대하여 텍스트네트워크분석방법을 적용한 연구들을 중심으로 분석방법, 분석지표, 연구결과 해석 등의 전반적인 연구방법론을 전달하고자 시도하였다. 이를 통해 더 많은 연구자들이 텍스트네트워크분석방법론에 관심을 가지고 연구를 수행한다면 간호학의 다양한 주제영역에서 지식구조를 파악하여 간호지식체의 발전에 필요한 함의를 제공하는데 기여할 것으로 기대한다.

## 본 론

텍스트네트워크분석에 대한 용어가 통일되지 않아 “개념네트워크, 주제 연결망, 네트워크 텍스트 분석, 언어네트워크, 단어네트워크, 의미연결망” 등의 여러 명칭이 혼용되고 있다.<sup>6)</sup> 그러나 분석과정은 유사한 절차를 보이는데 텍스트 자료를 수집하고, 수집된 텍스트를 형태소 단위로 해체하고 의미 있는

단어를 선정한 후 동시출현 행렬을 개발하여 사회연결망 분석을 수행한다. 텍스트를 해체하고 최종 선정한 단어를 “키워드”라고 부른다.<sup>20)</sup> 본 연구에서는 의미상 혼란을 피하기 위해 키워드와 저자키워드(주요어)를 구분하였다.

텍스트네트워크분석방법의 장점을 요약하면 다음과 같다. 첫째, 오랫동안 축적된 많은 양의 관련논문을 수집하여 모두 연구자료에 포함하므로 연구결과에 대한 과학적 근거로서 강력한 힘을 가진다. 둘째, 시간 흐름에 따른 연구주제의 변화, 학문분야에서의 핵심연구주제, 국내 현황, 세계적 추세와 같은 거시적 조망이 가능하다. 셋째, 핵심주제를 중심으로 하위주제들과의 관련성, 주제의 확장 영역과 같은 미시적 조망이 가능하다. 넷째, 소시오그램(sociogram)으로 연구결과를 시각화하여 보여줌으로써 연구자와 독자가 직관적으로 지식구조를 탐색할 수 있다. 이러한 장점을 활용하여 간호학의 특정주제 또는 특정 학술지에서 서지정보 텍스트를 수집하여 다양한 측면에서 연구동향을 탐색할 수 있다.

본 연구에서는 텍스트네트워크분석방법론에 대한 고찰과 활용을 위해 필요하다고 생각되는 배경 이론, 소프트웨어 프로그램 특성을 간략히 살펴보고, 국내 간호학 학술지를 대상으로 텍스트네트워크분석을 적용한 연구들에 대한 방법론적 분석을 바탕으로 텍스트네트워크분석방법의 연구과정을 기술하였다.

## 1. 배경 이론

배경이론으로 정보이론, 동시단어분석, 그래프 이론, 네트워크 이론, 사회네트워크분석의 개요를 다루고, 사회네트워크 분석 지표들은 연구과정 단락에서 자세히 설명하고자 한다.

정보이론(information theory)은 추상적인 정보를 수학적 모형으로 계산하여 정보의 정량화(quantification)를 가능하게 하였고 디지털 통신기술의 발전에 수리적 토대가 된 이론이다. 정보이론은 동시단어분석법의 발전과 더불어 1990년대부터 여러 형태의 복잡한 텍스트에서 주제 탐색, 주제영역 근접화 등의 이론적 근거가 되었다.<sup>21)</sup> 예를 들어 반복적으로 되풀이되는 자료보다는 불규칙하고 특징적인 자료가 정보로서 의미가 있다는 정보이론의 가정에 따라 텍스트에서 불규칙하지만 일정한 유사성을 가진 키워드를 발견하는 방법을 적용하여 주제 패턴을 발견하였다.<sup>22)</sup>

동시단어분석법은 1980년대 프랑스에서 “LEXIMAPPE” 프로그램의 개발로 발전하였는데, 초기에는 두 개의 단어가 같은 말뭉치(corpus)에 자주 등장할수록 유사성이 높고 의미적

연관성이 있다는 전제하에 근접성 지표(proximity indexes)와 포괄성 지표(inclusion indexes)를 구하고 이를 근거로 나타난 연관관계를 지도(map)로 표현하였다.<sup>8)</sup> 컴퓨터 기술의 발전에 힘입어 동시단어분석법은 대규모의 비정형(unstructured) 텍스트를 기계적으로 처리하여 유용한 정보를 발굴하는 텍스트 마이닝(text mining)에 활용되고 있다. 즉 일정한 규격이나 형식을 갖춘 정형(structured) 텍스트가 아닌 문서나 자연어 형태의 비정형텍스트를 대상으로 자연어처리과정(Natural Language Process, NLP)을 수행할 수 있는 컴퓨터 프로그램을 이용하여 텍스트를 의미형태소로 정제한 후 많은 문서에서 동시에 출현한 의미형태소가 무엇인지 계량할 수 있다.

그래프 이론(graph theory)은 1736년 오일러의 “쾨니히스베르크 다리 건너기” 논문이 시초이며, 점과 선을 이용하여 관계를 수학적 그림으로 나타내는 이론으로서 널리 활용되고 있다.<sup>23)</sup> 그래프에서 점과 선은 여러 가지 용어로 표현된다. 점(point or node), 교차점(junction), 꼭지점(vertex)과 선(line), 변(edge), 호(arc), 연결(link), 분지(branch)처럼 필요에 따라 다르게 사용된다. 그래프에서 한 점의 연결정도(연결차수, degree)는 그 점과 연결된 선의 수를 말한다. 선의 방향에 따라 점에 들어오는 선의 수를 뜻하는 내향연결정도(in-degree)와 점에서 나가는 선의 수를 뜻하는 외향연결정도(out-degree)를 구분한다. 그래프는 수학적으로 행렬(matrix) 형식으로 표현된다. 그래프에서 점의 개수가  $n$ 인 행렬은 ‘ $n \times n$ ’으로 나타낸다. 정 사각형(정방형) ‘ $n \times n$ ’행렬은 행과 열의 개수가 같고, 직사각형 ‘ $m \times n$ ’행렬은 행과 열의 개수가 다르다. 일원 모드 행렬(1-mode matrix)은 행과 열의 속성이 같고, 이원 모드 행렬(2-mode matrix)은 행과 열의 속성이 다르다. 가령 텍스트네트워크에서 ‘키워드\*키워드’는 1원 모드 행렬, ‘키워드\*논문’은 2원 모드 행렬이다.

네트워크 이론(network theory)은 네트워크를 이루는 구성요소와 네트워크 특성의 두 측면으로 구분된다.<sup>24)</sup> 먼저 네트워크 구성요소에 대한 이론을 설명하면, 네트워크에서는 그래프의 점과 선을 노드와 링크로 부른다. 노드의 속성은 매우 다양하다. 노드 속성은 개별 행위자(actor), 회사 내 여러 조직체계, 다양한 기업, 국가, 바이러스를 비롯한 생물체, 단어 형태의 키워드, 논문의 저자 등 현실세계에 존재하는 다양한 개체의 특성이 반영된다. 또한 역할, 지위, 기능과 같이 추상성이 부여된 속성도 있다. 노드 간에 발생한 관계는 링크가 되며 노드와 링크를 네트워크 그래프로 표현한다. 관계의 유형도 노드 속성처럼 다양한데, 크게 사회적 관계, 상호작용, 유사성, 흐름으로 분류한다.<sup>23)</sup> 각 예를 들면 사회적 관계는 서로 알고 지내

는 사이, 친족, 채팅 친구 등이 있고, 상호작용은 좋아하거나 싫어하는 관계, 이메일을 주고 받는 관계, 거래 관계 등이 있고, 유사성은 같은 동호회 참석 유무, 같은 이벤트 참여 유무, 두 키워드의 동시 출현 유무 등이 있고, 흐름은 정보나 자금 흐름, 물류 등이 있다. 다음으로 네트워크 특성에 대한 이론을 살펴보면, 전체 네트워크(complete network or global network)와 하위 네트워크(sub-network), 개별 노드의 에고 네트워크(ego-network or personal network or local network), 두 노드 간의 양자 네트워크(dyad network), 세 노드 간의 삼자 네트워크(triad network), 약한 관계(Strength of Weak Ties, SWT) 네트워크, 구조적 공백(structural holes) 네트워크 등이 있다. 한편 네트워크 모양에 따라 별(star) 형, Y자 형, 선형, 원형으로 분류하여 그 특성이나 조직구조를 진단하고, 시·공간에 따른 변화 유무에 따라 정적(static) 또는 동적(dynamic) 네트워크로 구분하여 현상을 설명할 수 있다. 노드 수에 따라 소규모 네트워크(small network)는 노드 100개 미만, 중규모 네트워크(medium network)는 노드 100~1000개 미만, 대규모 네트워크(large network)는 노드 1000개 이상으로 구분<sup>23)</sup>한다.

사회네트워크분석은 1930년대 제이콥 모레노가 가출 청소년에게 영향을 미친 또래 집단의 사회적 관계를 소시오그램 그림으로 설명하면서 시작되었다. 1940~50년대 행렬 대수학의 개념을 바탕으로 그래프 이론이 발전하면서 사회네트워크분석은 노드를 행위자로, 링크를 행위자간의 다양한 관계로 표현하였다. 별 모양의 사회네트워크 조직은 중심행위자가 있어 원형 조직보다 문제해결을 빨리 하는 것으로 나타났고, 조직 구조에서 동일한 위치에 있는 행위자는 유사한 사회적 영향을 미친다. 또한 매개중심성이 높은 행위자는 조직 내에서 구조적 공백을 일으키는 매우 영향력 있는 사람이다. 사회네트워크의 특성을 수학적으로 계산하는 알고리즘 및 이를 활용한 지표들이 발표되면서 사회네트워크분석방법론이 더욱 발전하였다.

## 2. 소프트웨어 프로그램의 특성

텍스트네트워크분석은 텍스트를 추출하고 단어를 정제하는 전처리 과정(pre-process)을 거치고 선정된 단어 쌍의 동시 출현 관계 행렬(co-occurrence matrix)을 개발하는 일련의 기술적 과정(technical procedure)이 필요하다. 각 단계별로 필요한 소프트웨어 프로그램을 살펴보았다. 단어 추출과 전처리에 사용되는 대표적인 형태소분석 프로그램<sup>7)</sup>으로 KrKwick (Korean Key Words in Context)<sup>25)</sup> 프로그램(<https://www.leydesdorff.net/krkwick>)은 한글 텍스트를 해체하고 자연어 처리와 동시출현 행렬 개발, 그래프 시각화도 가능하다. Kim 등<sup>13)</sup>은 KrKwick을 이용하여 수집된 논문의 영문초록에서 주제를 추출하였다.

동시출현행렬 개발 및 사회네트워크분석에 사용되는 프로그램에는 NetMiner (<http://www.netminer.com>), R (<https://www.r-project.org/about.html>), UCINET (<http://sites.google.com/site/ucinetsoftware/home>), Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek>) 등이 있다. NetMiner는 유료 프로그램으로 사회네트워크분석과 시각화, 비정형텍스트 자료를 이용한 단어 전처리 과정까지 모두 가능하다. 최근에는 유튜브나 트위터 텍스트 수집기능이 추가되었다. R 패키지는 공개 프로그램으로 텍스트네트워크 분석패키지로는 SNA와 IGraph (<https://igraph.org>)가 주로 이용된다. UCINET은 Freeman이 개발하여 공개한 프로그램으로 사회네트워크분석에서 전통적으로 가장 많이 사용되고 있다. 사회네트워크분석과 그래프(소시오그램) 시각화가 가능하다. Pajek은 “거미”라는 뜻의 슬로베키아 이름을 가진 프로그램으로 UCINET 시각화 메뉴에 포함되어 있지만 Pajek 사이트에서 별도로 다운받아 사용할 수 있다. UCINET 프로그램에서 사회네트워크분석 결과를 시각화하려면 “Visualize” 메뉴에 있는 “NetDraw”와 “Pajek”을 사용하게 되는데, 500개 이상의 많은 노드를 가진 네트워크 시각화는 “Pajek”을 이용하는 것이 해상도가 높은 그림을 얻을 수 있다. 가령 Lee 등<sup>12)</sup>은 1995~2009년 동안 대한간호학회지에 발표되었던 5,936편 논문의 영문 제목과 초록에서 용어추출을 위해 텍스트분석기로 “BiKE Text Analyzer”를 이용하였고 네트워크분석 및 소시오그램을 위해 “Pajek”을 활용하였다.

3. 연구과정

본 연구에서는 텍스트네트워크분석 연구과정(Fig. 1)을 1) 문헌수집, 2) 텍스트 추출, 3) 전처리 과정, 4) 동시출현 행렬개발, 5) 사회네트워크 분석 순으로 세분화 하였고, 6) 결과 해석 및 논의를 추가로 설명하였다. 또한 텍스트네트워크분석방법을 국내 간호학 학술논문을 대상으로 적용한 연구논문 6편<sup>3,12-15,26)</sup>에 대하여 텍스트경계, 동시출현 단위, 군집방법, 소프트웨어를 구분하여 정리하였다(Table 1).

## 3. 연구과정

1) 문헌수집

문헌 검색을 위한 검색식은 연구주제와 관련된 문헌을 모두

포함하면서 관련이 없는 문헌은 제외되어야 한다.<sup>27)</sup> 연구자는 데이터베이스의 특성을 잘 파악하여 검색영역을 어떻게 제한할 것인지 고려하고 불리언(boolean) 연산자를 적절히 사용하여야 한다. 검색영역을 “전체(all field or any field)”로 선택하는 것이 문헌검색에서 문헌누락을 방지하므로 검색영역은 넓

히고 검색어의 조합과 각 데이터베이스에서 제공하는 “제한 검색 기능”을 활용하여 정확도를 높이는 것이 바람직하다. 데이터베이스에서 제공하는 제한검색 기능을 이용하여 간호학, 동료평가 논문, 인간대상, 영어, 년도와 같은 항목을 선택하고 연구목적에 부합하는 문헌수집을 할 수 있다. 따라서 자료수

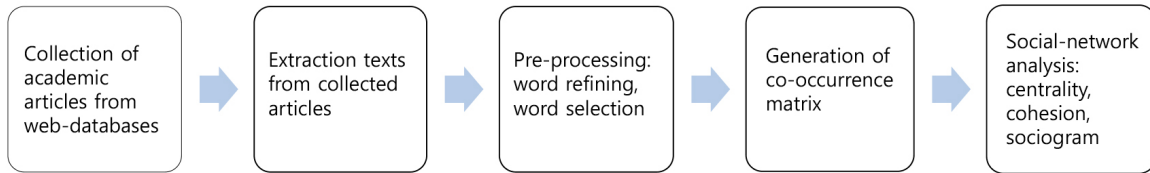


Fig. 1. The study flow of text-network analysis.

Table 1. Korean Researches applying Text Network Analysis in Nursing Discipline

Author (year)	Title	Boundary of text (measure)	Co-occurrence unit	Clustering method	Software
Lee et al. (2011)	A social network analysis of research topics in Korean nursing science	Titles and abstracts of the 5,936 papers (frequency, clustering)	Cosine similarity based on article	Kth nearest neighbors	BiKE Text Analyzer, Pajek
Kim et al. (2014)	Research trend analysis of Do-Not-Resuscitate decision: Based on text network analysis	Abstracts from 32 articles (frequency, degree centrality)	Article	Structural equivalence	KrKwic, Wordle, NetMiner 4
Kwon & Park (2015)	Knowledge structure of the Korean journal of occupational health nursing through network analysis	Author-keyword from 400 articles (frequency, graph drawing algorithm)	Cosine similarity based on article	NA	COOC v.0.4 PFNet, NodeXL
Kwon & Bae (2016)	A study on the knowledge structure of cancer survivors based on social network analysis	Author-keyword from 1,099 articles (frequency, graph drawing algorithm)	Cosine similarity based on article	NA	COOC PFNet, NodeXL
Lee et al. (2017)	Analysis of research articles published in the journal of Korean academy of nursing administration for 3 years (2013~2015) : The application of text network analysis	Author-keyword from 171 articles (degree centrality, betweenness centrality)	Jaccard coefficient based on article	NA	NetMiner 4.0.
Park et al. (2017)	A comparison of hospice care research topics between Korea and other countries using text network analysis	Abstracts from 347 Korean and 1,926 international articles (degree, betweenness, and closeness centrality, clustering)	Window size ≤ 3 & article	Community modularity using GN algorithm	NetMiner 4.3.

COOC=Co-occurrence Matrix Generation Program; GN=Girvan & Newman; KrKwic=Korean keyword in context; NA=not applicable; PFNet=Pathfinder network.

집을 위한 예비검색을 여러 차례 수행하면서 검색된 문헌의 양과 목록의 차이를 파악하고 일관성 있는 검색결과가 반복적으로 나오는 최종 검색식을 확정한다. 이 과정에서 웹 검색 전문 도서관 사서나 문헌연구 전문가의 자문을 구하는 것이 도움이 된다.

문헌검색은 여러 개의 데이터베이스를 이용하는 것이 문헌 누락의 위험을 줄일 수 있지만 연구목적에 따라 공인된 수준 높은 문헌검색 데이터베이스만을 이용하기도 한다.<sup>15,28)</sup> 또한 각 데이터베이스에서 제공하는 “서지정보 상세 내보내기” 기능이 있는지 확인하고, 한번에 최대 몇 개의 서지정보를 내보낼 수 있는지 파악하면 문헌수집 전략수립에 참고가 된다. 국내 간호학 분야에서 논문 수집을 위해 이용되는 데이터베이스에는 한국교육학술정보원(Research Information Sharing Service, RISS), 한국학술지인용색인(Korea Citation Index, KCI), 한국학술정보(Koreanstudies Information Service System, KISS), 누리미디어의 DBpia<sup>®</sup>, 국가과학기술정보센터(National Digital Science Library, NDSL) 등이 있다. 해외는 Ovid MEDLINE (Medical Literature Analysis and Retrieval System Online), EMBASE (Excerpta Medica Database), WOS (Web of Science), CINAHL (Cumulative Index to Nursing and Allied Health Literature), PubMed, Scopus<sup>®</sup> 등이다.

이 단계에서 고려할 점은 많은 양의 문헌을 수집하여 자료를 처리하는 데 드는 시간과 노력, 기술적인 측면이다. 데이터베이스에서 검색된 문헌을 바로 엑셀에 저장해 주는 내보내기 기능을 사용할 수도 있지만, 여러 데이터베이스에서 문헌을 수집할 때는 서지정리를 위한 전문 소프트웨어를 이용하는 것이 편리하다. 가령 EndNote<sup>™</sup> 프로그램은 중복문헌을 찾아 제거하고, 연도별로 정리하기 편하다. 또한 연구자가 필요한 서지정보만 지정하여 반출하고 엑셀 형태로 변환하는 것이 가능하므로 초록, 저자키워드, 연도, 제목과 같이 분석에 사용되는 텍스트를 추출하는데 도움이 된다. 반출방법은 EndNote “Edit” 메뉴에서 “Output Style > New style”을 생성하여 연구자가 원하는 서지정보(예: 논문 제목, 연도, 저자키워드, 초록)를 분리하여 새로운 “Output Style”이름을 만든다. 원하는 논문을 모두 선택한 후 연구자가 만든 반출 양식을 지정한 후 반출한다. 목록은 “text 파일”로 저장이 되므로 엑셀 프로그램에서 저장되었던 파일을 열어 연도순으로 정렬하여 일련번호를 부여하고 엑셀파일을 저장한다. 엑셀파일로 변환하는 과정에서 비정형자료인 초록은 여러 문장으로 이루어진 하나의 단락이므로 셀 하나에 하나의 초록이 한 개의 말뭉치로 들어가

도록 한다. 그러나 저자키워드는 있는 그대로 정형자료로 이용할 수 있어 셀 하나에 저자키워드 한 개만 들어가도록 분리한다.

웹 사이트에 공개된 Bibexcel (<https://homepage.univie.ac.at/juan.gorraiz/bibexcel>) 프로그램도 서지정보를 텍스트파일(.txt) 형태로 저장한 후 저자명, 초록, 저자키워드 추출에 사용할 수 있으나 호환 가능한 웹 데이터베이스가 제한적이다.<sup>29)</sup>

## 2) 텍스트 추출

수집된 문헌의 서지정보에서 분석에 포함할 텍스트의 경계를 결정하여 추출한다. 우선 서지정보 “제목, 저자, 학술지명, 연도, 저자키워드(주요어), 초록” 중에서 무엇을 분석대상 텍스트로 이용할 지 범위를 결정하는 것이 필요하다. 초기 텍스트네트워크분석에서는 저자키워드만을 추출하였거나, 제목과 저자키워드를 혼합하였다. 최근에는 비정형 텍스트인 초록을 추출하여 전처리 할 수 있는 프로그램들이 개발되어 초록 추출 연구가 늘고 있다. 그러나 초록은 수집된 문헌의 양이 많을수록 전처리 과정에 시간과 노력이 매우 많이 소요되지만 저자키워드는 비교적 간편하다. 동시단어분석법 초기 연구에서 저자키워드를 이용한 분석에서 색인자 효과(indexer effect)로 인한 편향의 우려가 있었으나 무시할만한 수준이라는 보고도 있어 장·단점을 판단할 필요가 있다.<sup>8)</sup>

텍스트를 추출하기 전에 보통 영어나 한글 중 한가지 언어를 선택하고, 목록을 통일하는 과정을 거친다. 국외 연구문헌을 포함하였다면 영어를 선택하는 것이 수월하다. 연구에 포함할 텍스트의 범위가 제목, 저자키워드, 초록이라면 3가지 모두 영어가 존재하는지 검토한다. 가령 한글 주요어만 서지상세정보로 제공하는 학술지 논문이 있다. 이때는 논문 원본을 찾아 영어로 제시된 주요어가 있다면 활용한다. 영어로 주요어가 제시되지 않은 논문은 영문 초록을 읽어보면서 저자가 제시한 한글 주요어를 영어로 번역하여 연구자료에 포함하는 노력을 통해 관련 논문이 최대한 분석자료에서 누락되지 않도록 한다. 시간에 따른 연구동향을 분석하려면 연도를 함께 추출한다. 추출방법은 서지정리 프로그램을 이용하여 추출할 서지정보를 엑셀로 변환하여 정리한다. 이 때 하나의 논문을 하나의 행에 두고 일련번호, 연도, 초록 또는 저자키워드를 각 열에 나열한다. 분석 소프트웨어 프로그램에 따라 텍스트파일만 가능한 경우도 있지만 연구자가 목록을 확인하고 연도별 집계를 하고 추이를 그래프로 분석하려면 엑셀이 편리하다. 또한 엑셀문서는 여러 가지 확장자로 호환된다.

### 3) 전처리 과정(pre-processing)

텍스트에서 의미 있는 단어를 선정하기 전에 연구자는 텍스트를 글자(형태소, morpheme) 또는 단어(구), 용어 단위로 해체하는 과정을 수행한다. 해체된 단어를 의미 있는 용어 형태로 통일하여 지정하고, 동일한 의미를 지니지만 다양하게 표현된 용어는 통일된 대표어를 정한다. 또한 특별한 의미를 가지지 못하는 일반적인 용어나 단어는 제외어로 결정한다. 이러한 일련의 과정을 단어정제(word clearing or refining)라고 부른다.

텍스트로 초록을 추출했다면 자연어처리과정을 거쳐 의미 있는 단어를 분리하고, 불필요하다고 판단되는 대명사, 부사, 숫자 등의 불용어(stop-word)를 제외하는 과정이 필요하다. 컴퓨터 프로그램을 이용하여 연구목적에 따라 명사, 형용사, 고유명사 등을 지정하여 분리할 수 있다. 초록은 여러 문장으로 이루어진 한 단락의 텍스트이므로 단어정제의 어려움이 있다. 동일한 단어가 문장에서의 위치에 따라 동사, 형용사, 부사의 기능을 하기도 하고, 영어는 품사에 따라 단어 형태-대소문자 구별, 부사형, 단·복수형, 동사변화 등·가 다르므로 통일 작업이 필요하다. 한국어는 동음이의어, 띄어쓰기 등의 문제를 고려하면서 정제작업을 수행한다. 가령 NetMiner 프로그램 “사전(dictionary)” 메뉴에 제외어, 지정어, 대표어(유의어)를 연구자가 등록하여 단어정제에 도움을 받을 수 있다. 연구목적과 무관하거나 의미 혼동을 부르는 단어는 제외어로, 명확한 의미를 전달하는 단어구(word phrase)는 지정어로, 여러 형태로 표현된 단어(구)는 대표어로 정리한 엑셀파일을 만들어 등록한다. Park 등<sup>3)</sup>은 NetMiner 프로그램을 이용하여 영문초록을 텍스트 자료로 분석하였는데 국내 347편에서 3,046개, 해외 1,926편에서 8,705개의 키워드를 정제하였고 10회 이상 출현한 국내 198개, 해외 594개 키워드를 선정하였다.

텍스트로 저자키워드를 추출했다면 자연어처리과정 없이 단어정제가 가능하다. 저자키워드에서도 동일한 의미로 사용된 여러 형태의 단어(구)와 약어를 통일하는 작업이 중요하다. 간호학에서는 MeSH 표목을 확인하면서 대표어를 지정하거나 빈도수가 많은 형태로 대표어를 지정할 것을 권한다. 예를 들어 Kwon과 Park<sup>14)</sup>은 1991~2014년 동안 한국직업건강간호학회지에 발표되었던 400편 논문에서 1,379개의 영문저자키워드를 추출하여 652개로 정제하였다. Lee 등<sup>26)</sup>은 2013~2015년 동안의 행정간호학회지 논문 171편을 수집하고 주요어 679개를 정제하였고 출현빈도 4회 이상인 31개 주요어를 선정하였다.

전처리 과정에서 논문의 저자가 선택한 단어에 대한 언어적 모호성, 관용적 표현의 차이를 완전하게 극복할 수 있는 컴퓨

터 프로그램은 현재 없으므로 연구자는 주관성을 배제한 단어 정제과정을 수행하기 위해 노력하고 그 과정을 논문에 자세히 기술하는 것이 바람직하다.

### 4) 동시출현 행렬 개발

이 단계에서는 정제된 단어 중에서 행렬 개발에 이용할 단어 선정과 동시출현 기준을 정하는 것이 필요하다. 동시출현 행렬 개발과정에서 단어의 수가 너무 많으면 기술적 어려움과 함께 주요한 현상을 왜곡할 수 있다.<sup>23)</sup> 텍스트네트워크분석에서 단어선정기준에 대한 공식적인 합의는 없지만 선행연구에서는 대체로 출현빈도가 높은 순서에 따라 상위 10~200여개의 단어를 선정하고 있는 것으로 보인다.

동시출현은 문장(sentence), 문단(paragraph), 문서(document)를 기준으로 한다. 각 기준에서 두 개의 단어가 자주 함께 나타나면 유사한 연관성이 있고 중요한 맥락적 관계를 가지는 것으로 간주된다.<sup>8)</sup> 함께 나타난 두 단어는 ‘공 출현(동시출현, co-occurrence)’ 관계이며, 분석결과를 해석할 때는 공 출현 기준을 고려한다. 동시출현 행렬 개발은 앞에서 소개한 컴퓨터 프로그램을 이용한다. 저자키워드를 추출한 경우에는 단어\*단어 형태의 1원 모드인 단어 간 동시출현 행렬 개발을 곧바로 할 수 있다. 초록의 경우 논문\*단어 형태의 2원 모드 행렬을 먼저 개발한 후 이를 이용하여 1원 모드인 단어간 동시출현 행렬(Fig. 2)을 개발하게 된다. 2원 모드에서 1원 모드로 변환할 때 유사성 지표로 코사인 유사도(cosine similarity), 자카드 계수(Jaccard coefficient)를 이용한다. 코사인 계수는 단어의 사용 빈도수로 인한 변이를 보정한 유사도 계수이며, 자카드 계수는 집단의 크기로 인한 변이를 보정한 유사도 계수인데, 초기 텍스트분석에서는 코사인 계수가 더 많이 사용되었으나 자카드 계수 사용이 늘고 있다. 두 지표 모두 ‘0~1’의 범위를 가지며 ‘1’에 가까울수록 동시출현 빈도가 높다. 두 계수는 문서 단위를 기준으로 동시 출현 관계를 계산한 값이다. Kwon과 Park<sup>14)</sup>은 코사인 유사도 계수를 기준으로 단어 간 동시출현 행렬을 생성하였고 중요한 관계만 남기는 Pathfinder network (PFnet)방식으로 노드와 관계를 축약한 지식지도를 시각화하였다.

NetMiner 프로그램은 “window size”를 기준으로 동시출현 기준을 정하는 기능을 제공하고 있는데, 문장 안에서 두 단어가 나란히 붙어 사용되고 있는 관계까지 반영할 수 있어 맥락구조 해석에 도움이 된다. 또한 추출된 단어의 TF-IDF (Term Frequency-Inverse Document Frequency, 문서 내 단어 중요도 평가 지수)를 계산해 주므로 문서와 문장에서의

		1	2	3	4	5	6	7	8
		NRS	PCA	PONV	QOL	RNs	RT	SRS	abdominal
1	NRS	2.0							
2	PCA		6.0						
3	PONV								
4	QOL				2.0				
5	RNs					2.0			
6	RT							3.0	
7	SRS						3.0		
8	abdominal								
9	ability					2.0			
10	activity								
11	adult								
12	age								
13	analgesia		5.0						
14	anxiety	3.0							
15	aromatherapy								
16	assessment								
17	attitude					8.0			
18	back pain								
19	barrier								
20	behavior				2.0				
21	blood	2.0							
22	burden					4.0			
23	cancer								
24	cancer pain								
25	care								

Fig. 2. An example of co-occurrence matrix.

단어 출현빈도를 모두 반영한 값을 얻을 수 있다. Park 등<sup>3)</sup>은 NetMiner를 이용하여 ‘window size 3 이하’를 기준으로 동시출현 행렬을 개발하였다. 이는 한 문장에서 3개 이하의 단어 간 거리에서 동시 출현하는 관계를 의미하며, 네트워크 분석 결과를 나타낸 소시오그램에서 키워드간의 맥락을 파악하는데 유용한 것으로 보인다.

시간에 따른 연구동향을 분석하려면 연도별로 논문을 추출하고 시간구간별로 동시출현 행렬을 개발한다. 가령 시간구간이 5개라면 5개의 동시출현 행렬을 개발하게 된다. 시간에 따른 연구동향을 분석한 선행연구를 살펴보면 Kim 등<sup>13)</sup>은 사회적 이슈가 있었던 특정 시점의 전과 후로 나누어 2개의 동시출현 행렬을 개발하여 연결중심성 분석을 수행하였고, Park 등<sup>3)</sup>은 총 19년동안의 논문을 약 5년의 구간으로 나누어 국내와 국외 각 3개씩 동시출현 행렬을 개발하여 연결중심성 분석을 수행하였다. 그러나 초기 연구에 해당되는 Lee 등<sup>12)</sup>은 총 15년동안의 논문을 3년 단위로 5개 구간으로 나누어 각 구간에서 출현빈도 상위 2%의 용어를 순위에 따라 제시하였는데 각 구간별 동시출현 행렬 개발은 하지 않은 것으로 사료된다.

##### 5) 사회네트워크분석

개발된 동시출현 행렬에 대하여 사회네트워크분석을 수행하는 단계이다. 사회네트워크분석 지표 중 텍스트네트워

크분석에서 주로 이용되는 노드 수준의 지표는 연결중심성(degree centrality), 매개중심성(betweenness centrality), 근접중심성(closeness centrality)이며, 아이겐 벡터 중심성(eigenvector centrality)을 이용하기도 한다. 노드들의 응집성(cohesion)을 보여주는 사회네트워크분석 지표는 컴포넌트(component), 커뮤니티(community)가 주로 이용된다. 네트워크 수준의 특성을 나타내는 네트워크 크기(size), 밀도(density), 포괄성(inclusiveness), 집중도(centralization)나 노드 수준의 특성을 나타내는 연결정도(degree), 연결강도(strength), 도달가능성(reachability) 등을 제시하는 경우도 있다. 텍스트네트워크분석에서 연결정도와 연결강도는 매우 중요한 개념이므로 이를 중심으로 살펴보고자 한다. 연결정도는 한 단어가 얼마나 많은 단어와 동시출현하고 있는가를 의미한다. 연결강도는 두 단어가 얼마나 많은 문장, 문단, 문서에서 동시출현 하였는가를 의미한다. 따라서 연결정도나 연결강도가 큰 단어는 중심성 값이 커진다. 연결중심성이 높을수록 많은 단어들과 연결됨을 의미하고 네트워크 안에서 중심에 위치하게 되므로 중요한 핵심주제로 간주된다. 근접중심성이 높을수록 네트워크 안에서 다른 단어와 가깝게 위치하므로 주변 단어와 맥락적 유사성이 높은 것으로 간주된다. 매개중심성이 높을수록 단어들 사이에서 유일하게 매개하는 역할을 하므로 주제의 확장을 설명하는 핵심주제로 간주된다.<sup>30)</sup> 사회네트워크분석



결과를 연결중심성, 근접중심성, 매개중심성 순위에 따라 표로 정리하여 제시할 수 있다(Table 2).

다음으로 사회네트워크분석 결과 발견된 지식구조를 소시오그램으로 시각화한다. 이때 노드와 링크 수가 너무 많으면 가시성이 떨어지므로 소시오그램에 포함할 노드와 연결강도를

결정한다. 수집된 문헌 수가 많을수록 연결강도 값의 분포가 커서 연결선이 매우 복잡하다. 본 연구자의 견해로는 네트워크 크기나 밀도에 따라 다르지만 보통 상위 10~30% 값을 포함할 때 가시성이 좋았고, 응집성 분석에서도 모듈성이 우수한 하위주제그룹 탐색이 가능하였다. 주요 의미 구조를 직관

**Table 2.** An Example of Keyword Ranks based on Freeman's Centrality

Rank	Keyword	Degree centrality	Keyword	Closeness centrality	Keyword	Betweenness centrality
1	Pain	0.589	Pain	0.688	Pain	0.653
2	Cancer	0.221	Patient	0.500	Cancer	0.152
3	Knowledge	0.189	Cancer	0.497	Hospice palliative care	0.095
4	Patient	0.179	Hospice palliative care	0.492	Nursing care	0.085
5	Hospice palliative care	0.168	RNs	0.490	Therapy	0.079
6	Pain management	0.168	Pain management	0.482	Patient	0.050
7	RNs	0.147	Knowledge	0.475	Satisfaction	0.043
8	Nursing care	0.137	Care	0.466	Knowledge	0.042
9	PCA	0.116	Therapy	0.463	Pain management	0.040
10	Symptom	0.105	South Korea	0.463	Symptom	0.038
11	South Korea	0.095	Symptom	0.461	Aromatherapy	0.035
12	Care	0.095	QOL	0.461	Management	0.032
13	Satisfaction	0.095	Nursing care	0.459	Home	0.029
14	QOL	0.084	Home	0.446	Care	0.028
15	Attitude	0.084	Analgesia	0.444	Breast	0.027
16	Education	0.084	PCA	0.442	RNs	0.024
17	Postoperative	0.084	Satisfaction	0.440	Surgery	0.022
18	Analgesia	0.074	Education	0.434	Massage	0.022
19	Home	0.074	Children	0.434	Older adults	0.021
20	Therapy	0.074	Attitude	0.434	Cancer pain	0.021
21	Cancer Pain	0.063	Depression	0.432	Self	0.021
22	Children	0.063	Assessment	0.432	Music therapy	0.021
23	Health	0.063	Postoperative	0.430	Intervention	0.021
24	Arthritis	0.053	Meta-analysis	0.430	Anxiety	0.021
25	Assessment	0.053	Cancer pain	0.430	South Korea	0.018
26	Management	0.053	Health	0.428	QOL	0.013
27	Nausea	0.053	Dementia	0.426	Arthritis	0.010
28	Older adults	0.053	Management	0.424	Attitude	0.008
29	Surgery	0.053	Female	0.424	PCA	0.007
30	Vomiting	0.053	Anxiety	0.424	Family	0.003

PCA=Patient controlled analgesia; QOL=Quality of life; RNs=Registered nurses.

적으로 파악하기에 적절하면서 하위주제그룹 탐색에도 적절한 연결강도 값을 결정하는 과정은 여러 차례의 모의분석을 통한 세심한 관찰과 분석이 요구된다. 사회네트워크분석 전문가나 프로그램 개발 전문가의 자문을 받아 과학적 엄격성을 확보하는 것이 도움이 된다.

네트워크는 노드들이 밀집되어 뭉쳐 보이는 부분, 연결선의 밀도가 낮은 부분, 연결선 없이 고립된 노드 등 여러 형태의 하위구조를 형성한다. 이러한 네트워크의 자연적 특성을 사회네트워크분석방법론에서는 다양한 “응집성” 지표로 개발하여 분석에 활용하고 있다. 응집성 지표 중에서 커뮤니티는 네트워크 안의 연결선은 촘촘하나, 밖의 연결선은 적은 상태의 하위 네트워크를 발견하는 알고리즘이다. 안과 밖에서 연결선의 밀도 차이가 명확할 때 모듈성(modularity)이 좋은데, 커뮤니티 알고리즘 방법에 따라 군집상태가 좋은 모듈성 값에 대한 기준이 다르다. 컴포넌트는 고립된 노드가 없어 하나로 연결된 상태로 커뮤니티보다 느슨한 응집성을 보인다. 따라서 텍스트네트워크분석에서는 모듈성이 우수한 하위주제그룹을 발견하기 위해 컴포넌트 분석을 먼저 수행하고 발견된 가장 큰 컴포넌트를 분리하여 커뮤니티분석을 추가로 수행하는 방법이 권장된다.

마지막으로 텍스트네트워크분석 결과 발견된 핵심주제를 연결중심성 순위에 근거하여 단어구름(word cloud)으로 표현하거나(Fig. 3A), 소시오그램 지도를 그림으로 제시한다. Kamada와 Kawai 스프링 지도(Fig. 3B)는 네트워크 구조에서 두 노드 간의 거리를 최적화하여 시각적으로 잘 볼 수 있도록 구현한 지도로 알려져 있다. 다차원축적도(MultiDimensional Scaling, MDS)는 실제의 거리와 방향을 반영하여 지도로 나타내는 방법인데 초기 네트워크 소시오그램에서 주로 이용되었다.

지금까지 제시한 동시출현 행렬(Fig. 2), 단어구름(Fig. 3A), 스프링지도 소시오그램(Fig. 3B)과 Table 2는 예시를 위해 본 연구자가 수집한 국내 통증관리 간호논문 321편의 서지정보에서 저자키워드를 추출하여 분석한 결과이다. Fig. 2의 행렬은 동일한 단어가 행과 열에 나타나므로 대각선은 자기상관(self-loop)을 형성하고 있다. 자기상관 값은 네트워크분석을 수행하기 전에 모두 “0”으로 치환하여 중심성분석 등에 영향을 주지 않도록 한다. 각 셀에 나타난 숫자는 문서 기준의 동시출현빈도이다. Fig. 3A의 단어구름에서는 단어 크기와 위치의 차이를 비교하면서 직관적으로 주제어를 파악할 수 있다. “통증(pain)”이 가장 크고 주변에 “호스피스 완화의료(hospice palliative care)”, “간호(nursing care)”, “증상(symptom)”,

“지식(knowledge)”이 가까이 있고, “암(cancer)”, “환자(patient)”, “통증관리(pain management)”, “자가통증조절기(PCA)”는 떨어져 있다. Fig. 3B의 소시오그램에서는 노드간의 맥락적 구조를 볼 수 있는데, 크기가 큰 노드는 높은 연결중심성을, 굵은 링크는 높은 연결강도를 의미한다. “암(cancer)”은 “유방(breast), 소화기관(gastrointestinal), 폐(lung), 생존(survival)”과 연결되어 있고 이 중에서 “유방(breast)”과 연결된 선이 가장 굵어 “유방암”에 대한 연구가 많았음을 암시한다. Table 2에서는 각 중심성 값의 순위에 따른 키워드를 살펴볼 수 있는데, 매개중심성에 비해서 연결중심성과 근접중심성 순위에서 동일한 키워드가 많으며 이들 키워드들은 단어구름에서 크게 나타난 키워드들과 일치하며, 핵심주제라고 판단된다. 매개중심성에서만 순위가 높은 “향기요법(romatherapy)”, “유방(breast)”, “마사지(massage)”, “음악요법(music therapy)”, “중재(intervention)”는 통증관리 간호연구에서 확장된 주제영역을 매개하는 주제어로 간주된다.

## 6) 결과 해석 및 논의

의미 있는 텍스트네트워크분석 결과를 정리하여 표와 그림으로 제시하고 종합적으로 발견된 현상이 무엇인지 기술하고 현상의 의미를 논의하면서 연구의 의의를 명확히 하는 것은 논문 작성에서 매우 중요하다. 텍스트네트워크분석 결과를 통해 발견된 연구동향을 총체적으로 논의하려면 연구자는 핵심주제, 주요 의미구조와 관련된 맥락을 중심으로 수집된 논문의 초록이나 전문을 찾아보고 현상의 의미를 탐색하는 과정이 필요하다. 또한 선행연구를 자신의 연구결과와 비교하면서 일치유무, 보완 및 강조가 필요한 부분을 정리하고, 관련주제에 대하여 전문가나 임상 실무자의 자문을 구하면서 연구결과 해석에 대한 통찰을 얻을 수 있다. 본 연구에서 고찰한 국내 간호학 논문 6편의 논의를 방법론적 관점을 중심으로 종합하여 살펴보면 다음과 같다. 텍스트네트워크분석 연구를 통해 발견된 지식구조는 현재까지의 연구동향을 체계적으로 조망하는데 도움이 되며, 주제 간의 맥락적 관계나 시간흐름에 따른 연구동향을 파악할 수 있다는 측면에서 텍스트네트워크분석 연구의 의의가 있다. 연구결과를 바탕으로 연구동향을 되돌아보고 미래 연구방향을 제시할 수 있다. 텍스트네트워크분석 연구의 한계점으로는 전통적 내용분석처럼 연구설계, 대상, 측정도구 등에 대한 통찰을 할 수 없어 필요하다면 추가로 관련 내용에 대한 연구분석을 수행해야 한다는 점과 단어의 전처리 과정에서 많은 시간과 노력이 필요하고, 연구결과 해석과정에서 주관적 편향의 위험을 배제할 수 없다는 점이었다.



## REFERENCES

1. Roberts CW, Popping R. Themes, syntax and other necessary steps in the network analysis of texts: a research paper. *Soc Sci Info*. 1996 Dec 1;35(4):657-65.  
<https://doi.org/10.1177/053901896035004005>
2. Zhang L, Hall M, Bastola D. Utilizing twitter data for analysis of chemotherapy. *Int J Med Inform*. 2018 Dec;120:92-100.  
<https://doi.org/10.1016/j.ijmedinf.2018.10.002>
3. Park EJ, Kim YJ, Park CS. A comparison of hospice care research topics between Korea and other countries using text network analysis. *J Korean Acad Nurs*. 2017 Oct;47(5):600-12.  
<https://doi.org/10.4040/jkan.2017.47.5.600>
4. Zavarone E, Grassia MG. Text analysis: an overview. *Wiley StatsRef*. 2018.  
<https://doi.org/10.1002/9781118445112.stat08089>
5. Paranyushkin D. [Internet]. Identifying the pathways for meaning circulation using text network analysis. Berlin: Nodus Labs; c2011 [up-dated 2011 Dec 25; cited 2019 Jan 20]. Available from:  
<http://noduslabs.com/research/pathways-meaning-circulation-text-network-analysis>
6. Park CS, Jung JW. Text network analysis: detecting shared meaning through socio-cognitive networks of policy stakeholders. *J Gov Studies*. 2013 Aug 23;19(2):73-108.
7. Park EJ, Cho SZ, editors. KoNLPy: Korean natural language processing in python. Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology; 2014 Oct 10-11; Kangwon National University, Chuncheon. 2014 Oct; 10:133-6.
8. He Q. Knowledge discovery through co-word analysis. *Libr Trends*. 1999 Summer;48(1):133-59.
9. Wasserman S, Faust K. Social network analysis: methods and applications. New York: Cambridge University Press; 1994. 825 p.
10. Baldwin C, Hughes J, Hope T, Jacoby R, Ziebland S. Ethics and dementia: mapping the literature by bibliometric analysis. *Int J Geriatr Psychiatry*. 2003 Oct 9;18(1):41-54.  
<https://doi.org/10.1002/gps.770>
11. Scott SD, Profetto-McGrath J, Estabrooks CA, Winther C, Wallin L, Lavis JN. Mapping the knowledge utilization field in nursing from 1945 to 2004: a bibliometric analysis. *Worldviews Evid Based Nurs*. 2010 Jun 14;7(4):226-37.  
<https://doi.org/10.1111/j.1741-6787.2010.00197.x>
12. Lee SK, Jeong S, Kim HG, Yom YH. A social network analysis of research topics in Korean nursing science. *J Korean Acad Nurs*. 2011 Oct 24;41(5):623-32.  
<https://doi.org/10.4040/jkan.2011.41.5.623>
13. Kim MJ, Noh SM, Ryu EJ, Shin SM. Research trend analysis of do-not-resuscitate decision: based on text network analysis. *Asian Onco Nurs*. 2014 Dec 24;14(4):254-64.  
<https://doi.org/10.5388/aon.2014.14.4.254>
14. Kwon SY, Park EJ. Knowledge structure of the Korean journal of occupational health nursing through network analysis. *Korean J Occup. Health Nurs*. 2015 May;24(2):76-85.  
<https://doi.org/10.5807/kjohn.2015.24.2.76>
15. Kwon SY, Bae KR. A study on the knowledge structure of cancer survivors based on social network analysis. *J Korean Acad Nurs*. 2016 Feb;46(1):50-8.  
<https://doi.org/10.4040/jkan.2016.46.1.50>
16. Kim YJ, Jang SN. Mapping the knowledge structure of frailty in journal articles by text network analysis. *PloS one*. 2018 Apr 19;13(4):e0196104.  
<https://doi.org/10.1371/journal.pone.0196104>
17. Freeman LC. Centrality in social networks conceptual clarification. *Soc Netw*. 1978;1:215-39.
18. Park YS. Formation of nursing knowledge: inductive reasoning. *Perspect Nurs Sci*. 2004 Dec;1(1):36-49.
19. Lee SW. Prospects for nursing theory analysis, evaluation and development. *Perspect Nurs Sci*. 2004 Dec;1(1):1-21.
20. Lee SS. A content analysis of journal articles using the language network analysis methods. *J Korean Soc Info Manag*. 2014 Dec ;31(4):49-68. <https://doi.org/10.3743/KOSIM.2014.31.4.049>
21. Zhang Y, Chen H, Lu J, Zhang G. Detecting and predicting the topic change of Knowledge-based systems: a topic-based bibliometric analysis from 1991 to 2016. *Knowl Based Syst*. 2017 Jul 10;133:255-68. <https://doi.org/10.1016/j.knosys.2017.07.011>
22. Montemurro MA, Zanette DH. Keywords and co-occurrence patterns in the Voynich manuscript: an information-theoretic analysis. *PLoS One*. 2013 Jun 21;8(6):e66344.  
<https://doi.org/10.1371/journal.pone.0066344>
23. Lee SS. Network analysis methods. Seoul: Nonhyeong; 2012. 370 p.
24. Borgatti SP, Halgin DS. On network theory. *Organ Sci*. 2011 Sep;22(5):1168-81. <https://doi.org/10.1287/orsc.1100.0641>
25. Park HW, Leydesdorff L. Understanding the KrKwic: a computer program for the analysis of Korean text. *J Korean Data Anal Soc*. 2004 Oct;6(5):1377-87.
26. Lee TW, Park KO, Seomun GA, Kim MY, Hwang JI, Yu SY, et al. Analysis of research articles published in the Journal of Korean Academy of Nursing Administration for 3 years (2013~2015): the application of text network analysis. *J Korean Acad Nurs Adm*. 2017 Jan;23(1):101-10.  
<https://doi.org/10.1111/jkana.2017.23.1.101>
27. Kim SY, Park JE, Seo HJ, Lee YJ, Jang BH, Son HJ, et al. NECA's guidance for undertaking systematic reviews and meta-analyses for intervention. Seoul: National Evidence-based Healthcare Collaborating Agency. 2011. 288 p.
28. Zheng P, Liang X, Huang G, Liu X. Mapping the field of communication technology research in Asia: content analysis

- and text mining of SSCI journal articles 1995-2014. *Asian J Commun.* 2016 Oct 20;26(6):511-31.  
<https://doi.org/10.1080/01292986.2016.1231210>
29. Persson O, Danell R, Schneider JW. How to use Bibexcel for various types of bibliometric analysis. Åström F, Danell R, Larsen B, Schneider JW, editors. *ISSI Newsl* [Internet]. 2009 Jun [cited 2019 Jan 20];05-S:9-24. Available from: <http://issi-society.org/media/1053/ollepersson60.pdf>
30. Park EJ, Ahn DW, Park CS. Text network analysis of newspaper articles on life-sustaining treatments. *J Korean Acad Community Health Nurs.* 2018 Jun 1;29(2):244-56.  
<https://doi.org/10.12799/jkachn.2018.29.2.244>