

한국육아지원학회 2023년도 하계 연구모임 안내

유아교육분야에서 텍스트 마이닝 연구방법 활용: 키워드 네트워크 분석 방법의 활용 및 실제

최지은(서울여대, 아동학과)
jechoi@swu.ac.kr

Copyright © 2023 All rights reserved
저자의 동의없이 무단복사를 금합니다.
본 교재는 하계 연구모임을 위해 제작된 것이며,
다른 곳에서 강의 교재나 세미나 교재로 사용하는 것을 금합니다.

유아교육분야에서 텍스트 마이닝 연구방법 활용: 키워드 네트워크 분석 방법의 활용 및 실제

I. 서론

II. 네트워크 연구의 절차

III. 네트워크 분석

IV. 실제

1. 네트워크 연구의 개요
2. 네트워크 분석의 단계
3. 네트워크 연구의 유의점

1. 네트워크 수준 분석
2. 노드 수준 분석
3. 네트워크에 내재된 속성 분석
4. 중심성 분석
5. 하위집단 분석
6. 예고 네트워크 분석

1. 텍스트 매뉴얼
2. UCINET/NetDraw를 활용한 네트워크 작성
3. UCINET/NetDraw를 활용한 CONCOR 분석
4. UCINET/NetDraw를 활용한 예고 네트워크 분석

01 용어 정의: 텍스트 마이닝(Text mining)이란?

Mining: 채굴하다 ▶ 어떠한 숨겨진 가치를 찾아내는 과정

1) 텍스트 마이닝(Text mining)

웹문서, SNS 대화, 블로그 등의 비정형데이터에 대한 마이닝 과정

[자연히 발생하여 사람이 의사소통 등에 사용되는 '자연어' 기반의 비정형(非定型) 텍스트 데이터에서 특정한 단어(Keyword)나 문맥(Context)을 바탕으로 인사이트(Insight)를 발견하는 기법]

vs.

Data mining: 스프레드시트 데이터, CSV 데이터 등과 같은 고정된 필드에 저장된 데이터, 즉 정형데이터에 대한 마이닝 과정

2) 텍스트 마이닝 기능

- ① 문서요약(summarize)
- ② 문서분류(classification)
- ③ 문서군집(clustering)
- ④ 특성추출(feature extraction)

01 용어 정의: 텍스트 마이닝(Text mining)이란?

3) 텍스트 마이닝 절차

- ① 데이터 수집(Data Gathering) : 뉴스, SNS(Social Network System) 등에서 텍스트 데이터 수집
- ② 전처리(Preprocessing) : 비(非)정형 데이터를 정형 데이터로 변환
- ③ 특징 추출(Feature Extraction) : 토큰화(Tokenization)를 통해 단어 단위로 나누어 중요한 단어를 추출
- ④ 텍스트 범주화(Text Categorization) : 학습 집합과 실험 집합으로 나누고, 학습 집합에 학습 알고리즘을 적용해서 텍스트를 범주화
- ⑤ 자동 요약(Automatic Summarization) : 주제를 질의 중심으로 요약, 혹은 원문 내용을 요약하거나, 문장 추출 혹은 담화 구조를 이용하여 자동으로 요약

4) 텍스트 마이닝 활용

- 비정형 데이터 분석의 다양한 분야에서 활용
- 대표적으로 특정 주제를 찾거나 특정 키워드로부터 내용을 추출하는 데 활용
 - [예] 뉴스 기사나 특정 글의 논조 분석
 - [예] 제품에 대한 의견이나 감성 반응 등을 분석하는 소셜 미디어 분석(Social Media Analysis) → 고객 맞춤형 서비스 개발
 - [예] 'Word Cloud', 'Text Network': 텍스트 마이닝 분석을 시각화한 대표적인 사례

02 용어 정의: 네트워크 분석(Network analysis)이란?

텍스트 마이닝: 텍스트 데이터에 특화된 기법 사용 vs. 네트워크 분석: 그래프 형태의 데이터에 특화된 기법 사용

텍스트 마이닝에서 추출된 정보를 기반으로 네트워크 분석을 수행
▶ 텍스트 데이터와 네트워크 구조를 통합하여 더 깊은 통찰력을 얻을 수 있음

1) 네트워크 분석

- 노드(Node)와 노드 사이의 연결(Edge)을 가진 그래프를 구성하고, 이를 시각화하고 분석하여 네트워크 내의 구조와 속성, 그리고 상호작용을 이해하는데 사용
- 중심성 지표, 커뮤니티 탐지, 연결 강도 측정 등의 기법을 통해 네트워크 내의 핵심 노드나 구조를 파악 & 정보 전파, 영향력 분석 등에 활용

01 네트워크 연구의 개요

1) 개요:

- 연구대상이나 현상에 대한 네트워크 데이터 수집
- 네트워크 데이터로 적절한 네트워크 표현
- 표현된 네트워크의 기본적인 구조나 형태의 특징 도출
- 국지적/전역적, 개별 노드와 링크, 부분 집단, 예고와 타자들 등에 나타나는 관계성의 특성 파악
- 네트워크로 구성되어 배태되는(embedded) 다양한 사물이나 현상의 미시적/거시적 특성을 분석하여 설명

01 네트워크 연구의 개요

1) 개요

- ▶ 네트워크 연구는 크게 네트워크 거버넌스 접근과 네트워크 구조 접근의 두 가지 접근 방법이 있음

네트워크 거버넌스(network governance) 접근

- 네트워크의 지배구조에 관심
- 네트워크의 행위주체인 노드들의 배태성(embeddedness)에 관심
- **배태성: 행위자들이 네트워크에서 어떻게 자리매김(embedded)에 따라 다른 효과가 나타나는 것 의미**
- 네트워크 내에서 행위자가 중심에 존재, 변방에 존재, 고립, 군집 구성 등에 따라 행위자에게 미치는 영향, 효과, 보상 등에 차이 발생

네트워크 구조 접근

- 네트워크의 구조적 특성에 따른 노드의 지위와 역할에 관심
- 네트워크에서 행위자는 구조적 공백이라는 특정한 위치를 차지하려는 전략을 구사하려고 함 = 행위자의 자리매김(embedding) 전략

01 네트워크 연구의 개요

1) 개요

- **네트워크 분석 방법 사용 영역:** 사회과학(사회학, 행정학, 신문방송학, **교육학**, 사회복지학, 정치학, 문헌정보학 등), 인문과학(언어학, 문학 등), 자연과학/공학 등(간호학, 생의학, 정보과학, 산업공학 등), 스포츠 과학 등
- **다양한 영역에서 활용하고 있는 네트워크 분석 방법 특징:**
 - ① 네트워크 구성원들을 연결해주는 관계들에 기초한 구조적 특성에 의해 그 방법적인 동기를 부여 받음
 - ② 체계적인 실증적 자료에 기초, 네트워크 구조의 그래픽화한 모습의 발견에 그 초점을 둠
 - ③ 주로 수학적 혹은 통계적인 분석 모델에 의존

[예] 네트워크의 구조에 관한 지표, 네트워크 주요 행위자에 관한 중심성 자료들을 이용한 연구

01 네트워크 연구의 개요

2) 네트워크 분석의 목적과 유형

- 네트워크 분석의 중요한 의의:

‘우리가 사는 세상은 어떻게 만들어질까?’

매우 본질적인 질문에 대해 해답을 줄 가능성 있음

- 즉, 세상을 구조(structure)와 행위(action)라는 두 가지 요인의 상호역동성으로 설명하는 이론이자 방법
- ▶ 네트워크에서 행위자의 역할은 그들이 연결되는 관계에 배태(embedded)되므로, 행위자들 간의 연결 관계의 구조적 특성들을 파악하면 네트워크로 표현되는 세상은 분석되고 설명됨

01 네트워크 연구의 개요

2) 네트워크 분석의 목적과 유형

- **미시적 분석:** 노드 간의 연결 관계에 나타나는 특성을 분석
 - [예] 노드와 링크의 수준에서 형성되는 다양한 역할과 행위를 분석하는 분석 기법 포함
- **거시적 분석:** 네트워크의 전체적 수준에서 네트워크에 배태된 다양한 특성들을 분석하는 기법

- **유형:**

• ①&②목적:
네트워크 구조의 해석 & 네트워크 내
군집들의 발견
→ 네트워크 구성원들이 링크를
만들어가는 '과정' 으로서 네트워크
모습 고찰, 네트워크 구조의 모습
해석에 중점

- ① **연결(connectivity)의 연구방법**
 - 네트워크 안에 있는 구성원들 간의 근접성(closeness)에 기초하여 네트워크 내의 군집들을 발견하는 것에 목적
- ② **위치(positional)의 연구방법**
 - 네트워크 안에 있는 구성원들 간의 유사성(similarity, 위치 유사성 또는 역할 유사성)에 기초하여 네트워크의 군집들을 발견하는 것에 목적

• ③목적:
다양한 네트워크 구조가 네트워크
구성원의 사회행위에 미치는 영향 분석

- ③ **네트워크 구조가 다양한 사회현상에 미치는 영향 분석**

01 네트워크 연구의 개요

3) 네트워크 분석의 특성

- ① 사물이나 현상을 네트워크로 표현하고 분석하는 것은 어렵지 않음
- ② 분석의 결과를 해석하는 것은 상당히 어려움
- ③ 네트워크 분석 결과를 토대로 어떤 일정한 패턴을 찾아내고, 이후의 행동을 예측하는 일은 매우 어려운 영역임

02 네트워크 분석의 단계

• 사회 네트워크(SNA, Social Network Analysis)의 단계:

- 분석대상(모집단)과 샘플의 규모 결정
- 조사할 사회적 관계의 내용 결정
- 사회적 관계의 측정 방법 결정
- 분석지표 결정 및 적절한 분석 도구 선택
- 분석작업 수행 및 분석결과 해석

[표 1] 네트워크 분석의 단계

단계	내용
1) 분석 문제의 제기	• 분석 문제의 설정 - 특정한 노드의 영향 분석 - 특정한 현상의 해석 - 특정한 문제의 풀이 및 예측 • 분석 대상의 설정 - 노드 데이터의 설정 - 관계 데이터의 설정
2) 네트워크 데이터의 조사	• 노드 데이터의 조사 - 인구통계학적 방법(전체 네트워크 방법) - 눈덩이 굴리기 방법(예고중심 네트워크 방법) 등 • 관계 데이터의 조사 - 사회조사 방법: 설문조사, 면담조사, 관찰조사 - 마이닝 조사방법: 데이터 마이닝, 텍스트 마이닝 - 실험조사 방법 등
3) 네트워크의 생성	• 데이터 코딩 • 데이터 변환 • 네트워크 시각화
4) 네트워크 분석의 실행	• 분석작업의 수행 - 네트워크 수준 분석 - 노드 수준 분석 - 네트워크에 내재된 속성 분석 - 중심성 분석 - 하위집단 분석 - 예고 네트워크 분석 • 분석결과와 해석(영향 분석, 현상 해석, 문제풀이 및 예측)

02 네트워크 분석의 단계

1) 분석 문제의 제기

(1) 분석 문제의 설정

① 영향의 목적:

- 네트워크 구조에 나타난 특성이 구성요소의 행위에 미치는 영향 분석

[예] 중심적인 역할, 매개와 중개의 역할, 특정한 연결 형태를 통해 구성요소들 상호 간의 제약하는 역할 등

② 해석의 목적:

- 네트워크에서 발현(창출)되는 다양한 현상을 이해하고 해석(설명)

③ 문제풀이나 예측의 목적:

- 문제풀이: 각종 문제 설정하고, 네트워크 분석하여 해당 찾기

[예] 9·11 테러범 검거, 전염병 감염경로, 연관된 전문가의 탐색 등

- 예측: 대상의 공간적·시간적 변화에 대한 '일정한 패턴' 파악하여 예측

[예] 노벨상 수상자 예측, 독감발생의 예측, 범죄행위의 용의자 예측 등

단계	내용
1) 분석 문제의 제기	- 분석 문제의 설정 - 특정한 노드의 영향 분석 - 특정한 현상의 해석 - 특정한 문제의 풀이 및 예측 - 분석 대상의 설정 - 노드 데이터의 설정 - 관계 데이터의 설정

02 네트워크 분석의 단계

1) 분석 문제의 제기

(2) 분석 대상의 설정

① 노드 데이터의 설정

- 일원모드(1-mode): 사람과 사람, 사물과 사물의 관계와 같이 관계의 양측 대상이 동일한 개체(사람, 사물)의 네트워크를 구성

- 이원모드(2-mode): 개체와 사건의 관계처럼 관계의 양측 대상이 사람과 특정한 사건을 나타냄 즉, 어떤 개체(사람, 사물)가 특정한 사건(소속, 활동)에 대한 포함 여부나 포함의 정도(빈도) 등을 나타내는 개체*사건의 네트워크를 구성

② 노드의 크기:

- 소규모 노드(100노드 이내), 중규모 노드(100-1,000 노드), 대규모 노드(1,000-10,000 노드), 초대규모 노드(10,000노드 이상)

③ 노드의 변화 가능성:

- 고정 데이터, 가변 데이터

단계	내용
1) 분석 문제의 제기	- 분석 문제의 설정 - 특정한 노드의 영향 분석 - 특정한 현상의 해석 - 특정한 문제의 풀이 및 예측 - 분석 대상의 설정 - 노드 데이터의 설정 - 관계 데이터의 설정

02 네트워크 분석의 단계

1) 분석 문제의 제기

(2) 분석 대상의 설정

② 관계 데이터의 설정

- 분석하고자 하는 목적에 따라 관계 데이터 설정
- 관계는 추상적인 개념이므로 측정이 쉽지 않음 → 관계를 명확하게 정의하고 설명을 제시하는 조작화 작업 필요

③ 관계의 유형 :

- 사회적 관계: 노드가 사람, 조직/기관, 지역, 생물 등의 개체 네트워크에서 가장 많이 다루는 관계
- 비사회적 관계: 노드는 웹사이트, 키워드, 문헌 등 비사회적 개체인 사물 사이에 나타나는 관계

④ 관계의 강도:

- 이진 데이터, 가중 데이터

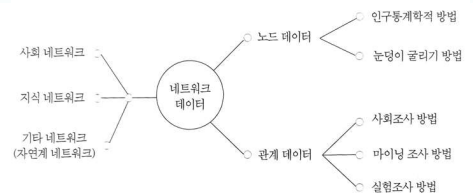
⑤ 관계의 방향:

- 무방향 관계, 송신 관계, 양방향 관계

단계	내용
1) 분석 문제의 제기	• 분석 문제의 설정 - 특정한 노드의 영향 분석 - 특정한 현상의 해석 - 특정한 문제의 풀이 및 예측 • 분석 대상의 설정 - 노드 데이터의 설정 - 관계 데이터의 설정

02 네트워크 분석의 단계

2) 네트워크 데이터의 조사



[그림1] 네트워크 데이터 조사의 고려사항

- **네트워크 데이터:** 분석하고자 하는 네트워크의 데이터 세트
노드(행위자)에 대한 데이터와 관계에 대한 데이터로 구성
- **네트워크:** 노드와 관계의 집합
- 네트워크 데이터는 주로 사회 네트워크와 지식 네트워크 영역에서 주로 논의되어 옴
 - **사회 네트워크 영역:** 사회조사 방법([예] 설문조사, 면담조사, 관찰조사, 실험조사 등)을 중심으로 관련된 이론이나 기법 논의
 - **지식 네트워크 영역:** 대규모의 데이터 대상으로 하는 마이닝 방법([예] 데이터 마이닝, 텍스트 마이닝 등)에 의해 노드를 추출하고 관계를 설정하는 이론과 기법 논의
 - **기타 네트워크 영역([예] 생태 네트워크 등):** 사회/지식 네트워크 영역에서 개발된 데이터 조사 방법 응용

02 네트워크 분석의 단계

2) 네트워크 데이터의 조사

(1) 노드 데이터: 분석의 대상이 되는 노드들의 집단에 대한 데이터

- 노드의 규모(크기), 노드의 속성에 대한 데이터 확보
- **노드 데이터 조사 = 노드의 규모 결정**: 네트워크의 경계 설정

• **네트워크 분석 기본 원칙**: 사회조사에서 사용하는 전통적인 의미를 가지는 표본의 개념 No 사용: 모집단을 대표하는 표본을 선정하고, 표본 개체에 대한 데이터들을 대상으로 분석하여 **모집단의 특성을 유추하지 않음**

→ 개체 간의 관계 데이터에 대한 일종의 계량정보학/계량사회학적 분석이기 때문

- ① 인구통계학적 방법(전체 네트워크 방법): 전체 노드의 규모를 사전에 결정 → 노드들 간의 관계 데이터 조사
 - 모집단의 경계가 명확하게 정해져 있을 경우
 - 분석(연구)의 목적이 되는 특정 지역 또는 특정 조직의 구성원들을 인위적으로 설정하는 경우
 - [예] 전수조사(census survey)
- ② 눈덩이 굴리기 방법(예고 네트워크 방법): 소규모 노드들로부터 시작 → 적정수준의 규모를 결정
 - 모집단의 경계가 명확하게 정해져 있지 않을 경우: 초기의 특정한 개체(노드 선택) → 초기 개체들에 질문하여 관계가 있는 다른 개체들 조사 → 해당 방식을 반복하여 개체 증가시켜 나감
 - [장점] 특수한 모집단의 노드들 탐색하는 데 유리, 개체(노드)의 탐색과 관계 데이터 수집이 동시에
 - [단점] 고립된 노드 발견이 어려움, 얻어지는 개체들의 전체 집단 변동 가능성↑

단계	내용
2) 네트워크 데이터의 조사	• 노드 데이터의 조사 - 인구통계학적 방법(전체 네트워크 방법) - 눈덩이 굴리기 방법(예고중심 네트워크 방법) 등 • 관계 데이터의 조사 - 사회조사 방법: 설문조사, 면담조사, 관찰조사 - 마이닝 조사방법: 데이터 마이닝, 텍스트 마이닝 - 실험조사 방법 등

02 네트워크 분석의 단계

2) 네트워크 데이터의 조사

(2) 관계 데이터: 분석대상의 노드들 간에 형성되는 관계에 대한 데이터

- 관계의 명칭, 관계의 조작적 정의, 관계의 강도, 관계의 방향 등 관계의 속성 데이터
- 노드들 간의 관계: 사회적 관계/비사회적 관계, 단일형 관계/다중형 관계

[표 2] 관계 데이터 조사방법의 유형

구분	방법	내용
관계탐색	설문조사	• 설문을 통해 관계 데이터를 조사
	면담조사	• 면담(인터뷰)를 통해 관계 데이터를 조사
	관찰조사	• 특정한 현장 또는 실험장에서 관계를 관찰
관계추출	마이닝 조사	• 각종 정보매체(기록매체)의 데이터에 관계형성 규칙을 적용한 추출 알고리즘으로 관계 데이터를 계산
관계발생	실험조사(시뮬레이션)	• 조성된 현장이나 실험 환경에서 발생하는 관계 데이터를 조사

단계	내용
2) 네트워크 데이터의 조사	• 노드 데이터의 조사 - 인구통계학적 방법(전체 네트워크 방법) - 눈덩이 굴리기 방법(예고중심 네트워크 방법) 등 • 관계 데이터의 조사 - 사회조사 방법: 설문조사, 면담조사, 관찰조사 - 마이닝 조사방법: 데이터 마이닝, 텍스트 마이닝 - 실험조사 방법 등

02 네트워크 분석의 단계

단계	내용
3)네트워크의 생성	• 데이터 코딩 • 데이터 변환(유사성 측정 포함) • 네트워크 시각화

3) 네트워크의 생성

(1) 데이터 코딩

(2) 데이터 변환

- 행렬로 표현된 네트워크 데이터는 다양한 용도에 의해 변환됨

(3) 네트워크의 시각화

- 네트워크를 시각적으로 표현하는 방법

- ① 그래프 표현
- ② 트리 표현
- ③ 다차원 척도 표현

02 네트워크 분석의 단계

단계	내용
4)네트워크 분석의 실행	• 분석작업의 수행 - 네트워크 수준 분석 - 노드 수준 분석 - 네트워크에 내재된 속성 분석 - 중심성 분석 - 하위집단 분석 - 에고 네트워크 분석 • 분석결과와 해석(영향 분석, 현상 해석, 문제풀이 및 예측)

4) 네트워크 분석의 실행

- 분석수준에 따라 6가지 유형으로 구분

- ① 네트워크 수준 분석
 - 거시(macro) 수준의 분석: 네트워크 크기, 네트워크 밀도 등의 분석지표 사용/네트워크 단위에서 분석의 결과가 나타남
- ② 노드 수준 분석
 - 미시(micro) 수준의 분석: 연결정도, 연결강도, 연결거리, 경로 등 사용/노드 단위에서 분석의 결과가 나타남
- ③ 네트워크에 내재된 속성 분석
 - 하이브리드(hybrid) 수준의 분석/특정 노드들이 밀접하게 구성되는 상호성, 이행성, 군집화 계수 등의 분석이 해당
- ④ 중심성 분석
 - 다양한 유형의 중심성 지표(연결정도 중심성, 근접 중심성, 매개 중심성 등)들 사용
- ⑤ 하위집단 분석
 - 컴포넌트 분석, 파당 분석, 구조적 등위성 분석(클러스터 분석, CONCOR 분석 등) 등
- ⑥ 에고 네트워크 분석
 - 에고 네트워크 특성 분석(크기, 밀도, 성분), 중개성 분석, 구조적 공백 분석(중복성, 제약성, 효과크기, 접촉효율성) 등 사용

02 네트워크 분석의 단계

4) 네트워크 분석의 실행

• 분석 도구

① 범용 SNA 도구

- NetMiner: 한국에서 개발된 분석 도구 및 시각화 도구
- UCINET: 종합적인 네트워크 분석 도구

② 시각화 도구

- NetDraw: 가장 유명하며, UCINET에서 연동하여 사용

단계	내용
4)네트워크 분석의 실행	• 분석작업의 수행 - 네트워크 수준 분석 - 노드 수준 분석 - 네트워크에 내재된 특성 분석 - 중심성 분석 - 하위집단 분석 - 예고 네트워크 분석 • 분석결과와 해석(영향 분석, 현상 해석, 문제풀이 및 예측)

03 네트워크 연구의 유의점

- 인간 관계 네트워크를 다루는 연구([예] 특정 집단 내 구성원들 간의 성접촉 관계의 네트워크, 이메일 송수신 네트워크, 휴대전화 연결 관계의 네트워크 등)에서는 윤리적 문제에 신경해야 함
 - 네트워크 데이터에 나타난 조사 대상자의 개인정보가 노출될 위험이 있으며, 분석 결과에 따른 개인들 간의 이해관계가 드러날 수 있기 때문

- ➔ 조사과정에서 분석과정까지 개인들 간의 관계를 다루게 될 경우, 개인의 속성을 조사하는 통계분석(사회조사)의 연구윤리보다 더 세심한 신경을 써야 함
- ➔ 프라이버시 보호를 위해 개인명, 조직명, 전화번호, 메일주소 등의 데이터를 코드화하거나 익명처리 해야 함
- ➔ 연구윤리심의위원회의 심의를 통해 연구에 사용되는 각종 개인정보가 연구윤리 문제에 저촉이 없도록 유의

01 네트워크의 기본 속성 분석1: 네트워크 수준의 속성 분석

1) 네트워크 크기(network size)

- 네트워크를 구성하는 노드의 수, 네트워크의 규모 (링크의 수 포함)
 - 네트워크 크기가 작은 경우: 관계 데이터는 비교적 쉽게 조사 가능
 - 네트워크 크기가 매우 큰 경우: 노드들 간의 연결 관계를 파악하는 일이 쉽지 않음
- 방향이 있을 경우: AB와 BA는 서로 다른 노드 쌍. 노드수 n , $n(n-1)$ 의 연결관계를 가지는 노드 쌍 존재
- 방향이 없을 경우: AB와 BA의 연결관계 동일. $n(n-1)/2$ 의 연결관계 존재

01 네트워크의 기본 속성 분석1: 네트워크 수준의 속성 분석

2) 밀도(density)

- 네트워크 노드들 사이에 연결된 정도, 전체 노드들이 연결된 개수로 표현
 - 밀도가 높은 네트워크:
 - 노드의 연결 관계가 많음, 상호 간 도움이나 교류가 많음, 구조적으로 서로 긴밀하게 연결되어 복잡한 구조
 - 밀도가 낮은 네트워크:
 - 노드들 간의 연결 관계 적음, 상호 긴밀하게 연결되어 있지 못함, 복잡하지 않은 구조
- ▶ 밀도 값을 통해 네트워크가 형성하는 연결 관계의 응집성, 결속도, 복잡성 등을 설명
- 밀도: 0에서 1의 범위
 - 밀도=0: 연결선이 하나도 없는 네트워크
 - 밀도=1: 모든 노드들이 연결되어 있음 (1에 가까울수록 네트워크 연결 관계의 응집성, 결속도, 복잡성이 높음)



02 네트워크의 기본 속성 분석2: 노드 수준의 속성 분석

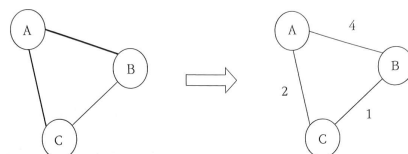
1) 연결정도(degree)

- 네트워크에서 특정한 한 노드의 자체적 속성, 해당 노드에 직접 연결되어 있는 노드들의 개수(또는 연결선의 개수)
- 연결정도의 계산 방식:
 - 노드에 연결된 링크의 수(절대적 연결정도)
 - 노드에 연결된 링크의 수에서 정규화하여 계산(정규화된 연결정도, 상대적 연결정도)
- 연결정도 값: 네트워크를 형성하는 각 노드들 간의 상대적인 순위를 의미
 - 노드와 연결된 링크를 기초로 계산되므로, 어떠한 계산 방법을 적용하더라도 노드의 연결정도 값의 순위는 변하지 않음

02 네트워크의 기본 속성 분석2: 노드 수준의 속성 분석

2) 연결강도(strength)

- 두 노드가 링크로 연결되는 강도(크기), 노드 간 연결의 중요도를 표현하는 지표
 - 관계의 강도는 두 노드 간에 형성되는 신뢰, 친밀도, 지속성, 접촉 빈도 등을 의미
 - 연결강도는 노드 간의 연결에 대한 관계의 강도를 나타내는 가중치로 표현, 그렇게 표현된 네트워크는 **가중 네트워크**
 - 가중 네트워크: 상대적 비교가 되는 가중치로 표현(vs. 링크의 유무)
 - 링크의 가중치=연결강도(vs. 모든 링크에 동일하게 1이 부여됨)
- 연결강도의 값이 크면: 강한 연결(strong ties)
- 연결강도의 값이 작으면: 약한 연결(weak ties)
- 연결강도의 '강함'과 '약함'은 전체적인 관계 데이터의 크기, 분석의 목적에 따라 상대적으로 판단됨
- 표시 방법: 가중치가 높은 링크에 선을 굵게 하거나, 가중치를 직접 표시하는 방법

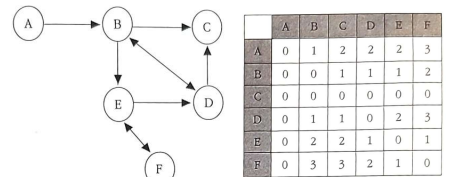


[그림2] 가중치가 있는 링크의 표시 방법

02 네트워크의 기본 속성 분석2: 노드 수준의 속성 분석

3) 연결거리(distance)

- 특정 노드 간에 연결된 거리, 즉 단계의 의미를 지님
 - 길이(length)의 개념을 강조해야 하는 경우, 연결거리는 경로길이(path length)라고 함
 - 연결거리, 경로길이 모두 많이 사용하는 용어
 - [예] 노드 A와 B 사이의 연결거리가 'd'인 경우, 그들은 d단계 떨어진 거리에 의해 연결되어 있다는 의미임
- 노드 A와 B 사이에는 다양한 경로가 존재할 수 있음
 - 최단연결거리(geodesic distance, 최단거리): 연결경로 중 가장 짧은 거리(최단경로거리, 최단경로길이)
 - ▶ 연결거리가 짧을수록 두 노드 간에는 가깝게 연결되어 있으며, 연결성이 높음
- 정보, 소문 등이 얼마나 신속하게 전달되는 지를 판단하는 지표
 - 중심성 분석, 하위집단 분석 등과 같은 네트워크 분석에서 활용 많음
 - [예] 가장 먼 연결거리가 3이면, 모든 노드들은 연결거리 3 이내에서 결속되어 있음을 의미
 - ▶ 최단거리는 네트워크 내 노드들의 결속성을 분석하는 척도로 활용



[그림3] 연결거리 사례

02 네트워크의 기본 속성 분석2: 노드 수준의 속성 분석

4) 직경(diameter)

- 네트워크 내의 연결거리 중 가장 긴 연결거리, 즉 네트워크에서 가장 멀리 떨어져있는 두 노드 간의 거리
- 네트워크의 지름

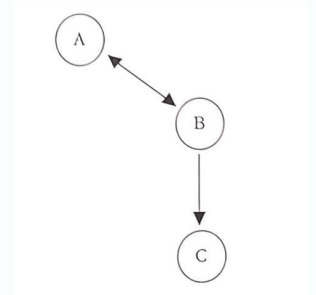
5) 평균연결거리(average distance)

- 네트워크 내 모든 노드 쌍의 연결거리 평균
- 평균거리, 평균경로길이, 특성경로길이(characteristic path length), 특성연결거리 등으로 불림

03 네트워크에 내재된 속성 분석

1) 상호성(reciprocity)

- 노드 쌍에 나타난 연결 관계의 유형 중 양자 간 연결 관계일 경우, 두 노드는 상호적임
 - 상호성은 밀도와 유사한 개념(지표임)
 - 밀도: 전체 네트워크에서 노드의 연결성 정도만 파악
vs. 양자 관계가 상호적인지, 일방적인지, 관계가 없는지 등 파악 가능
 - 상호 연결되어 있으면, 평형상태(equilibrium) 또는 안정된(stable) 상태
 - 일방적일 경우, 불안정된(unstable) 상태



[그림4] 노드 쌍에서 상호성의 사례

03 네트워크에 내재된 속성 분석

2) 이행성(transitivity)

- 세 노드에서 발생하는 관계, 노드의 삼자 간 관계와 관련된 지표
 - 현실세계에서 삼자 간 관계가 많이 나타나며, 의미 있는 분석 가능

3) 군집화 계수(clustering coefficient)

- 군집화(clustering): 특정 지역의 이웃 노드들이 밀접하게 구성되는 현상

04 중심성 분석

중심성(centrality)

- 한 행위자가 전체 네트워크에서 중심에 위치하는 정도를 표현하는 지표

- 프리만이 제안한
 - 연결정도 중심성
 - 근접 중심성
 - 매개 중심성

[표 3] 중심성 지표의 요약

기본 중심성	주요 근거	중심성 지표	비고
연결정도 중심성	연결정도	프리만 연결정도 중심성	위세 중심성
		보나시치 파워 중심성	
근접 중심성	연결거리	프리만 근접 중심성	위세 중심성
		보나시치 아이겐벡터 중심성	
		페이지 랭크	
매개 중심성	연결거리	프리만 노드 매개 중심성	
		프리만 링크 매개 중심성	
		흐름 매개 중심성	

04 중심성 분석

1) 연결정도 중심성(degree centrality)

- 네트워크의 노드들이 얼마나 많은 연결을 가지고 있는지를 측정
 - 연결된 노드의 수가 많을수록 연결정도 중심성은 높아짐
 - 계산: 프리만의 연결정도 중심성 척도가 가장 많이 알려짐

2) 근접 중심성(closeness centrality)

- 근접도(closeness): 네트워크의 다른 모든 노드들과 얼마나 근접하게 연결되어 있는가, 연결거리와 연관
 - 한 노드의 근접 중심성은 해당 노드가 얼마나 네트워크의 **중앙**에 있는지를 측정
 - ▶ 네트워크 전역에서 가장 일반적인 영향력을 가지는 노드가 무엇인지 알 수 있음
 - 근접 중심성이 높은 노드는 네트워크 내 다른 모든 노드들과 가장 짧은 연결거리를 가지고 있음을 의미

04 중심성 분석

3) 매개 중심성(betweenness centrality)

- 한 노드가 다른 노드와 네트워크를 구축하는 데 있어 중개자 혹은 다리 역할을 얼마나 수행하느냐를 측정
- 매개 중심성이 높으면, 중개자, 문지기의 역할을 한다고 볼 수 있음

4) 위세 중심성(prestige centrality)

- 개별 행위자의 중심성과 해당 행위자를 에고노드(ego)로 하고, 이 에고노드와 연결된 이웃노드들(alter)의 중심성 지표를 함께 고려한 것
- 한 노드의 영향력 또는 중요도를 측정하고 평가하는 데 사용하는 척도
- 연결된 노드들의 중심성에 가중치를 부여하여 노드의 중심성을 측정
- [계산] 보나시치가 개발한 파워중심성(power centrality, 연결정도 중심성 보완), 아이겐벡터 중심성(eigenvector centrality, 근접중심성 보완)

05 하위집단 분석

하위집단 분석(sub-group analysis)

- 네트워크의 구조분석
- 전체 네트워크의 노드들을 유사한 속성을 가지는 몇 개의 노드집단으로 구분하고, 이들로 구성되는 하위 네트워크 또는 하위집단들을 구분해내는 작업

[표 4] 하위집단 분석의 유형 구분

유형	구체적인 방법		비고
컴포넌트 분석 (구성집단분석)	단순 컴포넌트 방법	시작점에서 직간접으로 연결된 노드들을 추적	응집도 (연결된 형태)
	사이클(cycle) 방법	K-사이클: 사이클 경로 길이가 k 이내	
	노드(knot) 방법	분절점(cut-point) 제거방식	
파당 분석 (결속집단 분석)	파당(clique) 방법	파당: 모든 노드들이 직접 연결된 하위 네트워크	응집도 (파당의 성격)
	n-파당(clique) 방법	노드 간 연결거리가 n 이내	
	k-플렉스(plex) 방법	직접 연결되는 노드 수가 n-k개 이상	
	k-코어(core) 방법	각 노드의 연결정도가 k 이상	
	n-클랜(clan) 방법	n-파당 집단들에서 직경이 n보다 큰 집단을 제외한 집단들	
구조적 등위성 분석	n-클럽(club) 방법	직경이 n인 노드들 집단 중에서 가장 규모가 큰 것	지위와 역할 의 등위성
	클러스터 분석	일반적인 클러스터링, 프로파일 클러스터링	
	CONCOR 분석	블록모델링 방법	
	다차원척도 분석	계량적 다차원척도 분석, 비계량적 다차원척도 분석	

05 하위집단 분석

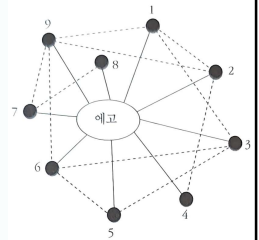
1) CONCOR 분석(CONvergence of iterated CORrelations)

- 상관관계 분석을 반복적으로 수행하여 적절한 수준의 유사성 집단을 찾아내는 방법
- 구분된 구조적 등위성 집단을 블록(block)이라고 하며, 블록을 대상으로 추가적인 분석 수행
= 블록 모델링(block modeling)

06 에고 네트워크 분석

1) 에고 네트워크

- 네트워크 분석에서 전체 네트워크를 대상으로 하는 분석이 아니라, 특정한 노드를 중심으로 하는 에고 네트워크 분석(ego network analysis)이 필요한 경우가 있음
>> 특정 지역과 관련된 국지적 사회구조에 내재된 구성원들 간의 특성을 기술하고 파악하는 분석에 보다 유용
 - 에고 네트워크 = 개인 네트워크(personal network)
 - 에고: 개인, 집단, 조직, 커뮤니티 등 의미
- 에고 = 에고노드, 초점노드(focal node): 타자노드(alter)들과 다양한 링크(연결 관계)를 형성하고 있음
- 타자노드: 에고와 어떤 기준에 부합하는 연결거리 내에 존재하는 노드들
- 에고 네트워크의 구성방법
 - ① 눈덩이 굴리기와 같이 에고 중심 네트워크 방법으로 에고 네트워크 데이터를 직접 수집하여 구성: 응답자(에고), 응답자가 지명한 노드(타자)
 - ② 기존의 전체 네트워크로부터 특정 노드를 대상으로 하는 에고 네트워크를 추출([그림5]): 에고와 연결된 타자노드들은 그대로 연결 관계를 지님, 이들의 연결 관계는 에고에 의미있는 영향을 미침



[그림5] 에고 네트워크 사례

06 에고 네트워크 분석

2) 에고 네트워크의 주요 분석 지표

- **네트워크 크기(size)**: 1단계 이웃 관계에 있는 모든 노드 수, 에고노드를 제외한 타자노드 모두 의미
- **네트워크 밀도(ego-density)**: 에고와 연결되어 있는 타자들이 상호 간 얼마나 잘 연결되어 있는지를 나타내는 지표, 즉 타자노드 쌍 사이에 직접 연결된 링크수의 비율
 - 에고와 연결된 타자들이 많이 연결되어 있을수록(에고밀도가 높을수록) 에고는 집단 내에 강하게 배태되어(embeddedness) 있음
- **연결강도(strength)**: 에고가 타자노드들과 얼마나 잘 연결되어 있는지 나타남
 - 기준에 따라 강한 연결 강도, 약한 연결 강도로 구분
- **중개자(broker)**: 직접 연결되지 않은 노드 쌍의 수
 - 중개자는 전체 가능한 링크수에서 실제 링크수를 제외한 것
 - 타자들이 서로 연결되어 있지 않으면, 에고는 중개자 역할을 수행하게 됨
 - 즉 다른 노드 쌍 사이에 있는(go-between) 형태가 됨
- **에고 매개성(ego betweenness)**: 에고 네트워크에서 중심성의 개념, 에고노드에 대한 매개 중심성
 - 매개 중심성 지표처럼, 에고 매개성은 모든 최단거리와 에고를 지나는 최단거리의 비율로 계산됨

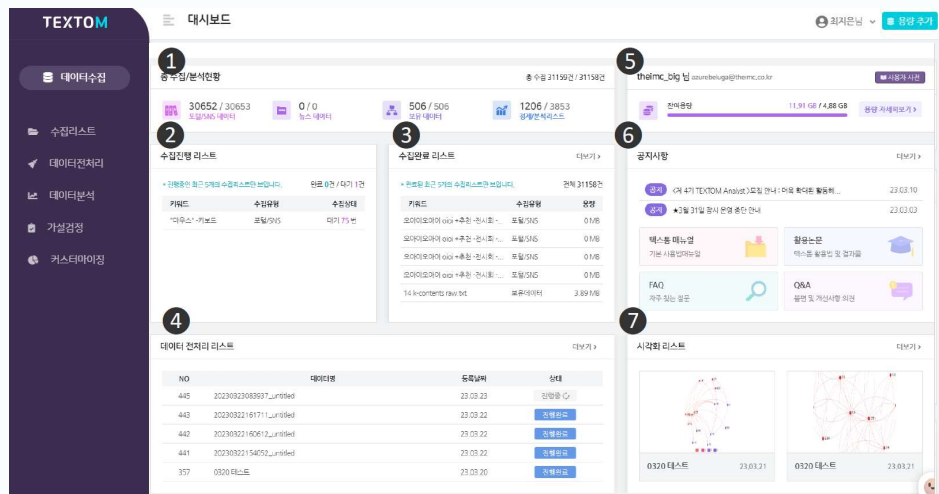
네트워크 크기	전체 링크수	연결가능 최대 링크수	에고밀도	중개자
9	11 (22/2)	36 (72/2)	0.31 (11/36)	25 (36-11)

	크기	링크수	노드쌍	에고밀도	중개자	에고 매개 중심성(정규화)
에고	9	22	72	0.31	25	51.85
1	4	8	12	0.67	2	16.67
2	4	8	12	0.67	2	16.67
3	4	8	12	0.67	2	16.67
4	2	2	2	1	0	0
5	3	6	6	1	0	0

[그림6] (위) 에고노드에 대한 주요 분석의 결과
(아래) 전체 노드를 에고로 설정하고 분석한 결과

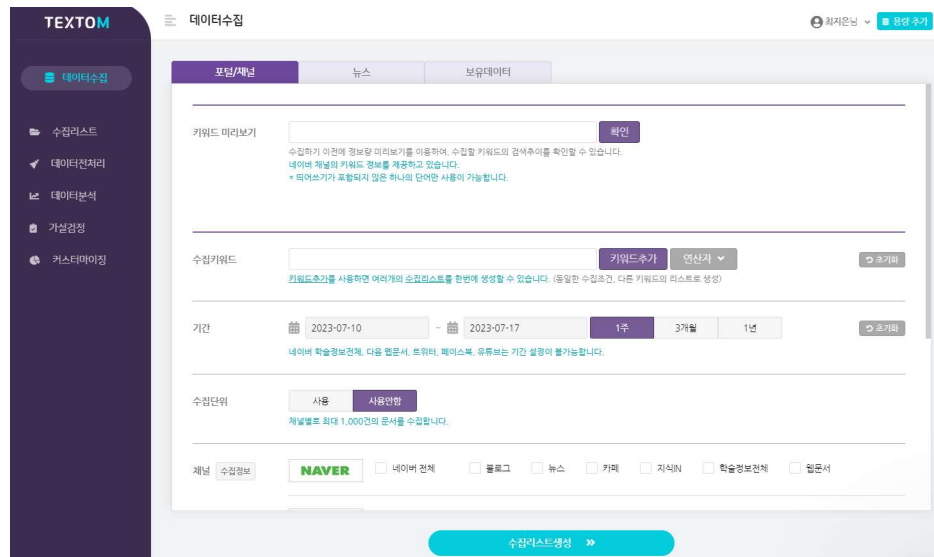
01 TEXTOM(텍스툼) 매뉴얼

◎ 대시보드



01 TEXTOM(텍스툼) 매뉴얼

① 데이터 수집

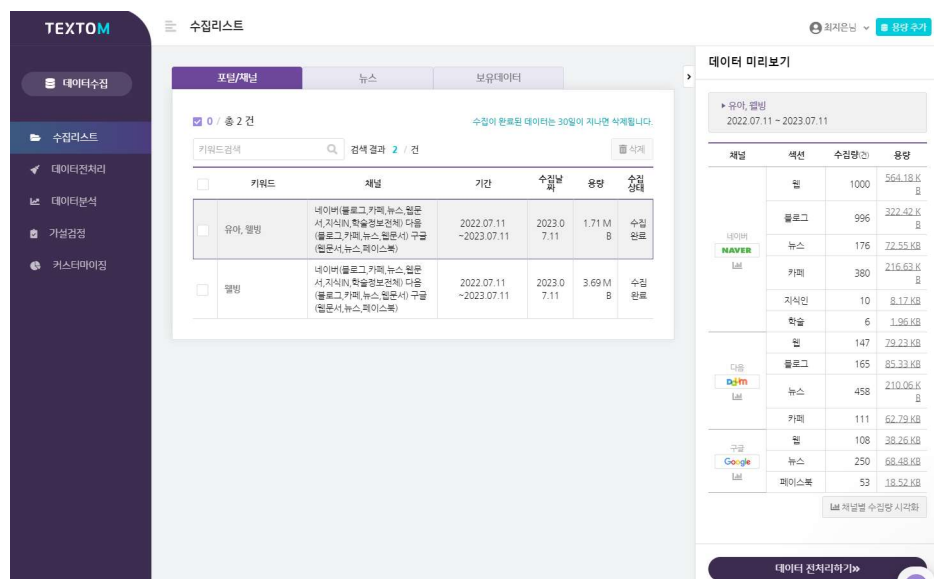


유아교육분야에서 텍스트 마이닝 연구 방법 활용: 키워드 네트워크 분석 방법의 활용 및 실제

039 |

01 TEXTOM(텍스툼) 매뉴얼

② 수집리스트



유아교육분야에서 텍스트 마이닝 연구 방법 활용 키워드 네트워크 분석 방법의 활용 및 실제

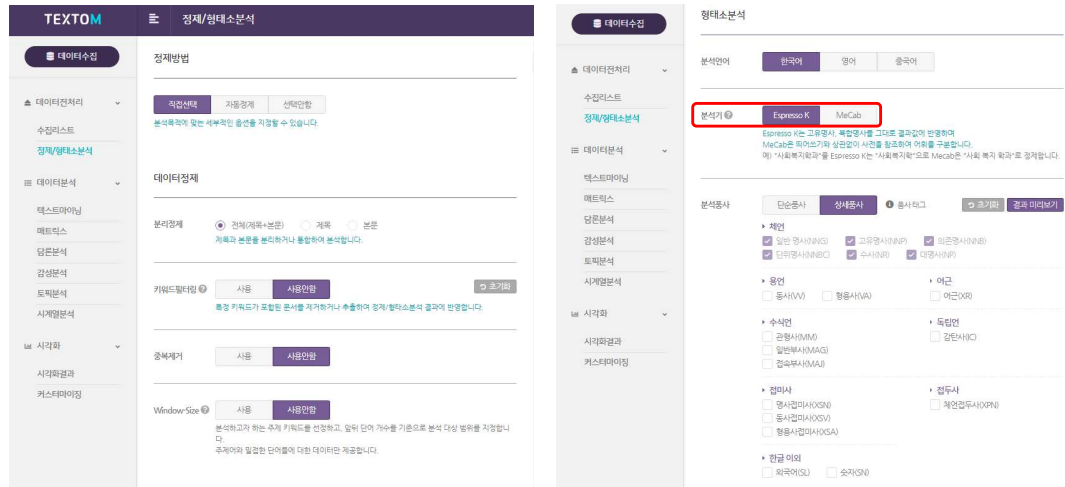
040 |

01 TEXTOM(텍스툼) 매뉴얼

③ 데이터 전처리: 정제/형태소 분석

2023.08.01 이전 구버전 캡처
(참고. 텍스트SV 2023.08.01부터 실시)

※ Espresso K는 고유명사, 복합명사를 그대로 결과값에 반영
 ※ MeCab은 띄어쓰기와 상관없이 사전을 참조하여 어휘를 구분
 [예] "사회복지학과"를 Espresso K는 "사회복지학"으로
 Mecab은 "사회복지학과"로 정제



01 TEXTOM(텍스툼) 매뉴얼

④ 키워드 정제



01 TEXTOM(텍스툼) 매뉴얼

⑤ 데이터분석

2023.08.01 이전 구버전 캡처
(참고, 텍스툼SV 2023.08.01부터 실시)

TEXTOM

데이터수입

데이터전처리

수집리스트

정제/정제소분석

데이터분석

텍스트마이닝

매트릭스

답문분석

감성분석

토픽분석

시계열분석

시각화

시각화결과

커스터마이징

매트릭스 데이터 Matrix → 빈도매트릭스

다운로드

	연속	5월	여형	준비	해피	추천	계주	예악	4월	여형	국내	공유	확정	수업
연속	0	6799	4162	2055	983	748	501	502	456	513	455	270	684	682
5월	6799	0	2589	1152	639	569	331	309	378	395	292	255	227	228
여형	4162	2589	0	967	147	277	303	263	162	179	72	25	227	227
준비	2055	1152	967	0	148	58	114	57	85	45	40	9	226	226
해피	983	639	147	148	0	111	7	63	69	34	45	3	1	0
추천	748	569	277	58	111	0	40	17	21	211	83	6	0	0
계주	501	331	303	114	7	40	0	34	26	33	22	10	0	0
예악	502	309	263	57	63	17	34	0	39	18	15	0	0	0
4월	456	378	162	85	69	21	26	39	0	18	22	2	2	1
여형	513	395	179	45	34	211	33	18	18	0	45	2	0	0
국내	455	292	72	40	45	83	22	15	22	45	0	3	0	0
공유	270	255	25	9	3	6	10	0	2	2	3	0	0	0
확정	684	227	227	226	1	0	0	0	0	0	0	0	0	0

유아교육분야에서 텍스트 마이닝 연구방법 활용: 키워드 네트워크 분석 방법의 활용 및 실제

043

01 TEXTOM(텍스툼) 매뉴얼

⑥ 단어분석

단어빈도

N-gram

TF-IDF

개체명 인식

시계열

단어빈도(Term Frequency)

1 분석 알고리즘

단어 빈도는 단어가 문서 내에서 얼마나 자주 발생하는지를 나타내는 척도입니다. 빈도 분석은 특정 문서들 내에서 자주 사용되는 단어를 추출하여 언급되는 빈도수에 따라 중요도를 분석하는 방법으로 발생빈도를 파악하여 문서의 주제 또는 문서에 대한 태도나 감성을 추론할 수 있습니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

2 텍스트 다운로드

3 엑셀 다운로드

단어	빈도(건)	백분율(%)	누적백분율(%)
외식	335	3.244%	3.244%
포장	301	2.915%	6.16%
맛집	264	2.557%	8.717%
역	245	2.373%	11.09%
출	150	1.452%	12.543%
하	143	1.385%	13.928%
있	135	1.307%	15.236%
집	131	1.268%	16.505%
가족	107	1.036%	17.541%
것	81	0.784%	18.326%

유아교육분야에서 텍스트 마이닝 연구방법 활용: 키워드 네트워크 분석 방법의 활용 및 실제

044

01 TEXTOM(텍스툼) 매뉴얼

⑥ 단어분석

단어빈도

N-gram

TF-IDF

개체명 인식

사계열

N-gram(N-gram)

1 분석 알고리즘

N-gram은 문자열에서 N개의 연속된 요소를 추출하는 통계 기반 언어 분석모델로, 문장 내 연속하여 동시 출현하는 단어와 그 빈도를 계산합니다. N-gram은 문장을 몇 개의 단어 개수에 따라 나눌지에 따라 종류가 결정되는데, 텍스트툼은 N=2인 Bigram 모델을 제공합니다.

상위 200개까지 단어를 따라 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

2 텍스트 다운로드 3 엑셀 다운로드

단어1	단어2	빈도(회)
가족	외식	85
수	있	44
것	같	43
포장	가능	38
외식	하	30
매뉴	포장	20
외식	잘	19
외식	장소	19
를	것	17
먹	수	17

N-gram

- N-gram은 문자열에서 N개의 연속된 요소를 추출하는 통계 기반 언어 분석모델로, 문장 내 연속하여 동시 출현하는 단어와 그 빈도를 계산
- N-gram은 문장을 몇 개의 단어 개수에 따라 나눌지에 따라 종류가 결정되는데, 텍스트툼은 N=2인 Bigram 모델을 제공

01 TEXTOM(텍스툼) 매뉴얼

⑥ 단어분석

단어빈도

N-gram

TF-IDF

개체명 인식

사계열

TF-IDF(Term Frequency-Inverse Document Frequency)

1 분석 알고리즘

TF-IDF는 단어의 빈도와 역문서 빈도를 사용하여 문서 내의 각 단어에 가중치를 부여하는 것으로, 문서 내 단어가 얼마나 중요한지를 평가합니다. TF-IDF값이 높은 단어는 문서 내에서 핵심적인 메시지를 담고 있을 확률이 높고, TF-IDF값이 낮은 단어는 문서 내에서 흔히 사용되고 있을 수 있습니다.

상위 200개까지 단어를 따라 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

2 텍스트 다운로드 3 엑셀 다운로드

단어	TF-IDF	DF	IDF
먹	218.36591025884076	162	0.8312894295667378
맛있	189.07563205300642	193	0.7161955750962364
잘	178.70910331785135	120	1.1913340221190757
집	173.5652293314114	105	1.3249254147435985
하	169.18261636052645	121	1.1830952193043809
있	167.76278772839456	114	1.2426873163066263
가족	150.24670214455807	97	1.404174786397739
매뉴	139.38008378879317	57	1.9358344970665717
것	137.8797913166904	72	1.7022196458850665
가능	128.89282115021894	61	1.8680119007278106

TF-IDF(Term Frequency-Inverse Document Frequency)

- TF-IDF는 단어의 빈도와 역문서 빈도를 사용하여 문서 내의 각 단어에 가중치를 부여하는 것
- 문서 내 단어가 중요한지 평가
- TF-IDF값이 높은 단어는 문서 내에서 핵심적인 메시지를 담고 있을 확률이 높고, TF-IDF값이 낮은 단어는 문서 내에서 흔히 사용되고 있을 수 있음

01 TEXTOM(텍스툼) 매뉴얼

⑦ 네트워크 분석

네트워크 작성	중심성	에고 네트워크	최단경로
CONCOR	클러스터링		

네트워크 속성

네트워크 전체에 대한 구조적 기술 통계량을 분석합니다. 네트워크의 노드들이 상호 간에 얼마나 연결이 잘되어 있는가를 관련 지수로 측정합니다.

상위 500개까지 단어를 따라 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

Network Measures	Value
Nodes	40
Total Edges	336
Diameter	2
Degree Centralization	0.59919
Closeness Centralization	0.71444
Betweenness Centralization	0.00011
Connected Components	1
Overall Reciprocity	0

1-Mode 매트릭스에 대한 네트워크 속성 분석, 에고 네트워크, CONCOR 제공

- 네트워크 속성 분석**
- Nodes: 노드의 총 개수
 - Total Edges: 엣지의 총 개수
 - Diameter: 직경
 - Degree Centralization: 연결정도 중심성
 - Closeness Centralization: 근접 중심성
 - Betweenness Centralization: 매개 중심성
 - Connected Components: 연결된 컴포넌트 수
 - Overall Reciprocity: 상호성

01 TEXTOM(텍스툼) 매뉴얼

⑦ 네트워크 분석 [참고] UCINET

Whole network measures

월빙-매트릭스-Worksheet		
1	# of nodes	50
2	# of ties	1774
3	Avg Degree	35.480
4	Indeg H-Index	33
5	K-core index	26
6	Deg Centralization	0.266
7	Out-Centralization	0.261
8	In-Centralization	0.261
9	Indeg Corr	-0.088
10	Outdeg Corr	-0.088
11	Density	0.724
12	Components	1
13	Component Ratio	0
14	Connectedness	1
15	Fragmentation	0
16	Closure	0.786
17	Avg Distance	1.276
18	Prop within 3	1
19	# w/in 3	2450
20	SD Distance	0.447
21	Diameter	2
22	Wiener Index	3126
23	Dependency Sum	676
24	Breadth	0.138
25	Compactness	0.862
26	Small Worldness	1.221
27	Mutuals	0.724
28	Asymmetries	0
29	Nulls	0.276
30	Arc Reciprocity	1
31	Dyad Reciprocity	1

31 rows, 1 columns, 1 levels.

〈표 4〉 네트워크 분석을 위한 UCINET 6.710 프로그램 실행 경로

분석	UCINET 6.710 프로그램 실행 경로
응집성	Network → Cohesion → Multiple cohesion measures
네트워크 집중도	Network → Centrality → Degree
통계적 유의성	Network → Compare densities → Against theoretical parameter
연결정도 중심성	Network → Centrality → Degree
근접 중심성	Network → Centrality → Closeness measures
매개 중심성	Network → Centrality → Freeman Betweenness → Node Betweenness

〈표 5〉 네트워크의 시각화를 위한 NetDraw 2.172 프로그램 실행 경로

분석	NetDraw 2.172 프로그램 실행 경로
전체 네트워크	File → Open → Ucinet Dataset
에고 네트워크	Layout → Ego Networks → Ego Network Viewer 키워드 선택

프로그램 실행 경로(최지은(2020), p. 11)

2) 유아 월빙에 대한 네트워크 분석

‘유아 월빙’ 관련 전체 네트워크 구조를 시각화한 결과는 그림 1과 같다. 유아 월빙에 대한 네트워크의 구조적 속성을 분석한 결과, 노드(node)는 200, 밀도(density)는 .407, 평균연결강도(average degree)는 80.930, 평균연결거리(average distance)는 1.592, 컴포넌트(components)는 2, 지름(diameter)은 3, 네트워크 집중도(network centralization)는 5.790%로 나타났다. 그리고 유아 월빙에 대한 네트워크의 통계적 유의성을 검정한 결과, 네트워크 데이터의 표집분포 평균은 3.6304, 표준오차는 0.3545로 나타났다. Z-score를 산출한 결과 Z = 10.1926으로 검정통계량이 Z-score의 절대값보다 높게 관찰됨 확률이 0.0002로 네트워크 내 데이터들 간의 관계는 통계적으로 유의한 것($p < .05$)으로 나타났다.

논문 예시 (최지은(2020), pp. 12-13)

01 TEXTOM(텍스툼) 매뉴얼

⑦ 네트워크 분석

네트워크 속성

중심성

에고 네트워크

최단경로

CONCOR

클러스터링

중심성

1 분석 알고리즘

네트워크에서 특정 노드(Node)가 갖고 있는 상대적 중요성을 수치화된 값을 계산합니다. 중심성 분석을 통해 구조적 위치의 관점에서 보다 더 영향력 있는(의미있는) 노드를 파악할 수 있습니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

2 텍스트 다운로드 엑셀 다운로드

단어	연결정도 중심성	근접 중심성	매개 중심성	아이겐벡터 중심성	페이지 랭크	클러스터링 계수
외식	1.0000000	1.0000000	0.0987763	0.2744513	0.1443191	0.4008097
포장	1.0000000	1.0000000	0.0987763	0.2744513	0.1325622	0.4008097
맛집	1.0000000	1.0000000	0.0987763	0.2744513	0.1403977	0.4008097
가족	0.9487179	0.9512195	0.0755708	0.2665096	0.0544874	0.4294294
매뉴	0.7455897	0.7959184	0.0349259	0.2267818	0.0401853	0.5098522
맛있	0.7692308	0.8125000	0.0440022	0.2274611	0.0403728	0.4804588
비달	0.6410256	0.7358491	0.0219752	0.2039866	0.0323150	0.5766667
죽밥	0.3846154	0.6190476	0.0051365	0.1416613	0.0287037	0.7333333
주운	0.6153846	0.7222222	0.0216081	0.1946801	0.0184932	0.568406
갈비	0.4358974	0.6393443	0.0047601	0.1591484	0.0193032	0.7573529

총 중심성 척도 4종 산출

- Degree Centralization: 연결정도 중심성
- Closeness Centralization: 근접 중심성
- Between Centralization: 매개 중심성
- Eigen Vector Centralization: 아이겐벡터 중심성

01 TEXTOM(텍스툼) 매뉴얼

⑦ 네트워크 분석

네트워크 속성

중심성

에고 네트워크

최단경로

에고 네트워크

1 분석 알고리즘

에고 네트워크는 한 노드를 에고(Ego)에 위치시키고, 에고와 다른 노드(타자, Alter)와의 연결을 표현한 네트워크를 의미합니다. 텍스트를 전체 네트워크에서 특정노드(에고)를 대상으로 하는 에고 네트워크를 추출하며, 에고와 연결된 타자노드들 간의 연결 관계를 분석합니다.

상위 200개까지 단어를 미리 볼 수 있습니다. 전체 단어는 다운로드하여 확인할 수 있습니다.

2 텍스트 다운로드 엑셀 다운로드

단어	Size	Ties	Pairs	Ego-Density	Broker	Ego Between
갈비	5	15	20	0.75	15	0
창업	6	20	30	0.66666	20	0
다이어트	7	27	42	0.64285	27	0
여행	7	25	42	0.59523	25	0
오리	8	34	56	0.60714	34	0
출발	9	41	72	0.56944	41	0
키메	9	40	72	0.55555	40	0
하루	10	50	90	0.55555	50	0
장가탕	10	50	90	0.55555	50	0
코리	10	50	90	0.55555	50	0

에고네트워크 :

한 노드를 에고(Ego)에 위치시키고, 에고와 다른 노드(타자, Alter)와의 연결을 표현하는 네트워크

에고네트워크 지표 6종 산출

- Size: 에고 네트워크 크기
- Ties: 직접 연결된 엣지 수
- Pairs: 전체 노드 쌍의 수
- Ego-Density: 에고 네트워크 밀도
- Broker: 중개자
- Ego Betweenness: 에고 매개성

01 TEXTOM(텍스툼) 매뉴얼

⑦ 네트워크 분석

The screenshot displays the CONCOR web application interface, which is divided into several sections:

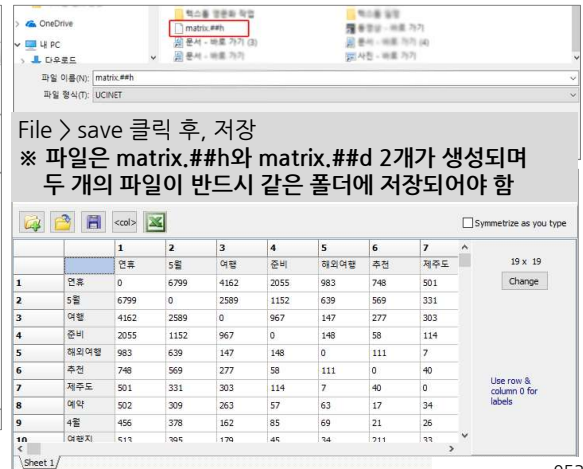
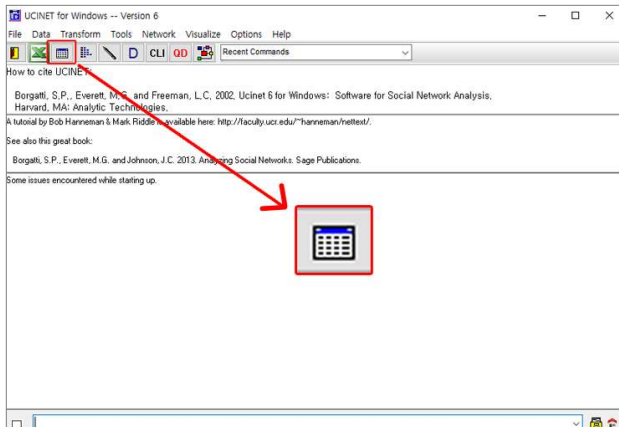
- Top Navigation Bar:** Contains tabs for "네트워크 속성" (Network Properties), "중심성" (Centrality), "애고 네트워크" (Ego Network), "최단경로" (Shortest Path), "CONCOR" (highlighted), and "클러스터링" (Clustering).
- Main Content Area:**
 - CONCOR Section:** Includes a title "CONCOR", a brief description of the analysis, and a dropdown menu for "군집 수" (Number of Clusters) set to 4.
 - 속성값 설정 (Property Value Setting) Panel:** A purple sidebar panel where users can select properties for analysis. It includes checkboxes for "단어빈도" (Word Frequency), "연결중심성" (Link Centrality), "근접중심성" (Proximity Centrality), and "매개중심성" (Mediation Centrality). A note states: "속성값은 텍스트마인딩, 중심성 분석이 완료된 경우 적용가능합니다." (Property values can be applied after text mining and centrality analysis are completed). There is a "선택 취소" (Cancel Selection) button at the bottom.
- 시각화 결과 (Visualization Results) Section:**
 - 네트워크 차트 (Network Chart):** Displays a complex network graph with nodes and edges, color-coded by cluster (1, 2, 3, 4).
 - 시각화 설정 (Visualization Settings):** A sidebar panel for customizing the visualization, including options for "글자 설정" (Text Settings), "노드 설정" (Node Settings), and "엣지 설정" (Edge Settings).

02 UCINET/NetDraw를 활용한 네트워크 작성

- ① 구글에서 'UCINET'을 검색 OR 텍스트 톨로그인 페이지 하단에 표시된 '텍스트 자료 활용 프로그램' 배너 중 'UCINET'을 클릭 → 새창으로 UCINET 접속

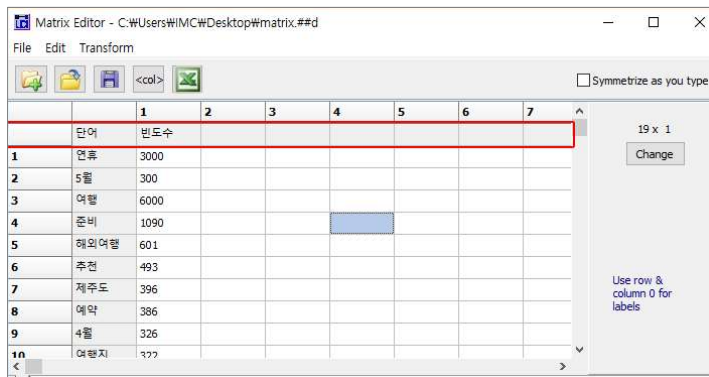
02 UCINET/NetDraw를 활용한 네트워크 작성

- ② 텍스트 'Matrix' 대메뉴에서 생성된 빈도매트릭스를 다운로드 → UCINET 상단 매트릭스 아이콘 클릭 → 빈 매트릭스 창에 excel파일에서 복사한 빈도매트릭스 붙여넣기



02 UCINET/NetDraw를 활용한 네트워크 작성

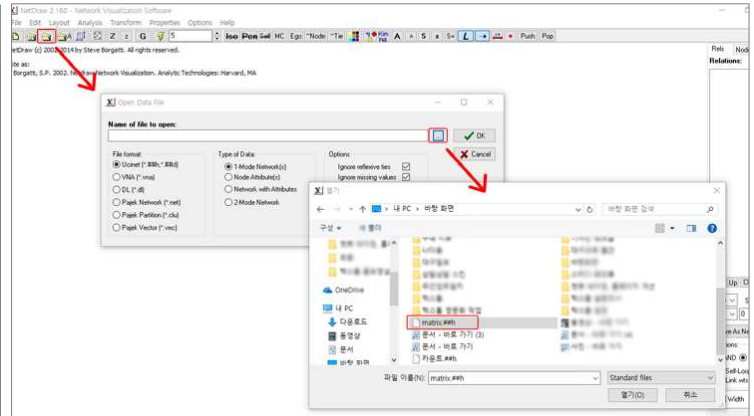
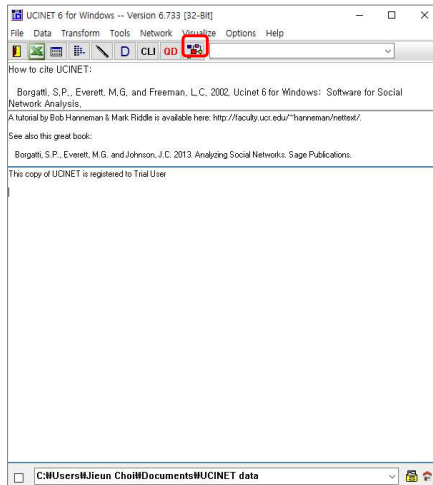
- ③ 텍스트에서 '추출단어 빈도수' 다운로드 → UCINET 입력창에 붙여넣기 (반드시 빈도매트릭스와 동일한 키워드 삽입) → 첫번째 행에 '단어'와 '빈도수'를 작성하여 라벨링



File > save 클릭 후, 저장
사용자 본인이 알아보기 쉬운 '빈도수' 또는
'카운트' 등의 이름으로 저장

02 UCINET/NetDraw를 활용한 네트워크 작성

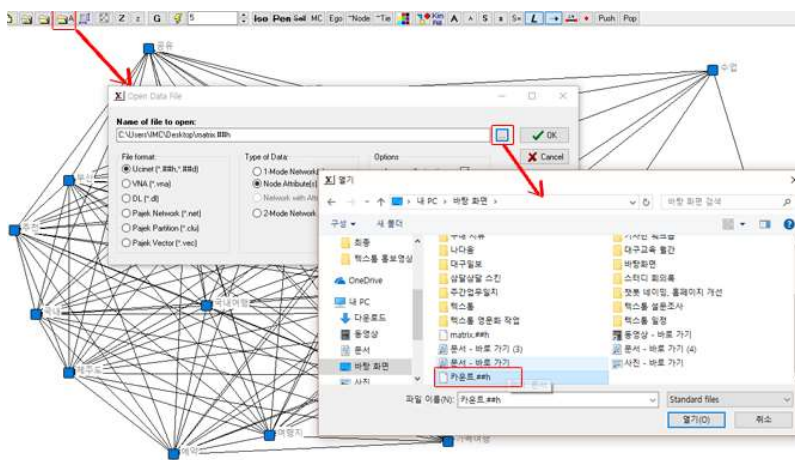
④ UCINET으로 돌아와 상단 외쪽 'NetDraw' 아이콘 클릭



Netdraw 창에서 상단 왼쪽에서 세번째 파일 열기 아이콘 클릭 → 첫 번째 지정한 빈도매트릭스 파일 오픈

02 UCINET/NetDraw를 활용한 네트워크 작성

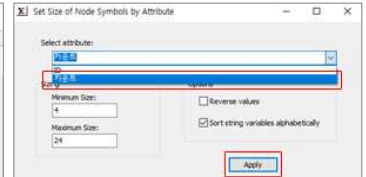
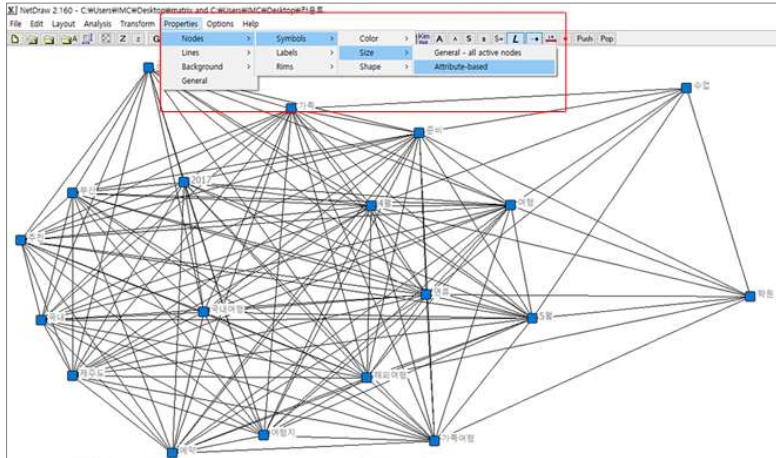
⑤ 빈도매트릭스 오픈한 상태에서 'A'가 표시된 파일 오픈 아이콘 클릭 → 두 번째 저장한 빈도 파일 오픈



빈도매트릭스 파일을 연
상태에서 눈에 보이는
변화는 나타나지 않음
노드의 크기차이를 원하지
않을 시, 이 과정
뛰어넘어도 됨

02 UCINET/NetDraw를 활용한 네트워크 작성

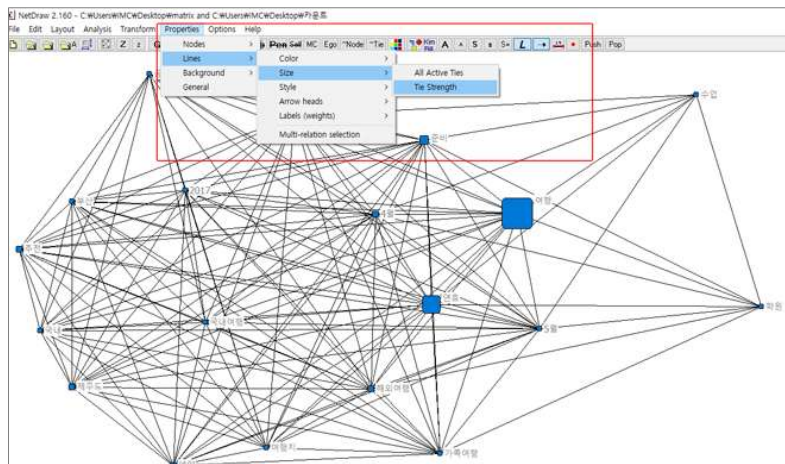
- ⑥ 빈도수에 따른 노드의 크기 조정: 상단 메뉴에 'Properties'- 'Nodes'- 'Symbols'- 'Size'- 'Attribute-based'를 차례로 클릭



'select attribute'에서 '빈도' 또는 '카운트'로 저장한 파일을 선택 → 'Apply' 클릭 (최소 및 최대값의 선정 가능)

02 UCINET/NetDraw를 활용한 네트워크 작성

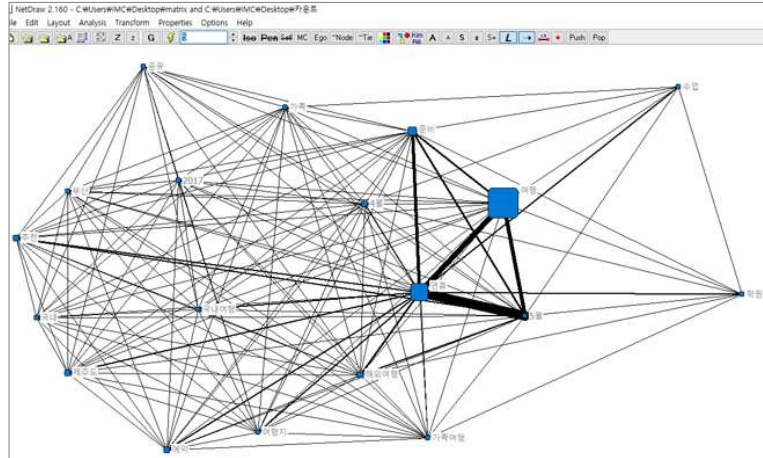
- ⑦ 빈도수에 따른 선 굵기 조정: 상단 메뉴에 'Properties'- 'Lines'- 'Size'- 'Tie Strength' 클릭



'Select relation'에서 Matrix파일을 선택 → 굵기의 최소값과 최대값을 설정 → Apply

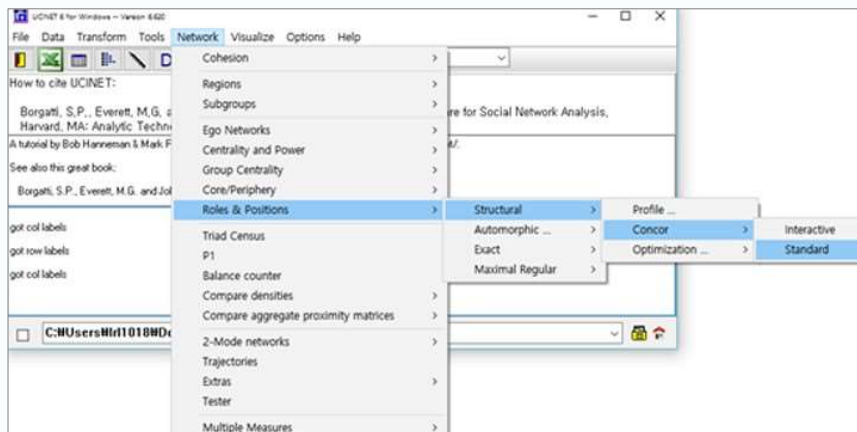
02 UCINET/NetDraw를 활용한 네트워크 작성

⑧ 네트워크 완성: UCINET NetDraw 창에서 노드나 선의 색상 조절 가능



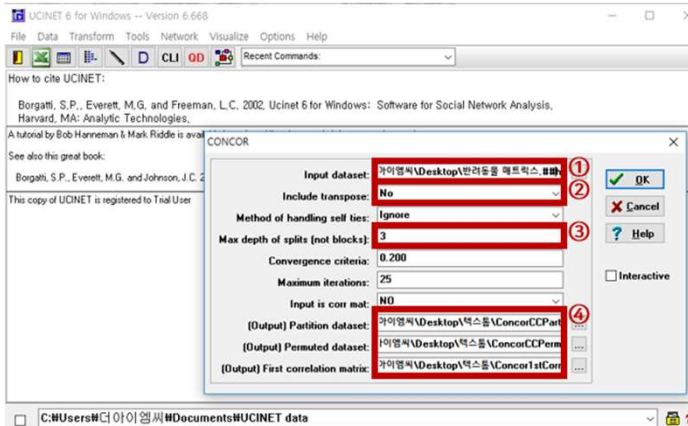
03 UCINET/NetDraw를 활용한 CONCOR 분석

① 상단 메뉴에서 'Network' - 'Roles & Positions' - 'Structural' - 'Concor' - 'Standard' 차례로 클릭



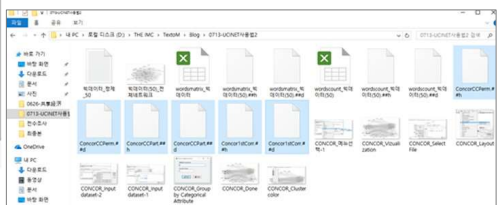
03 UCINET/NetDraw를 활용한 CONCOR 분석

② 아래 내용에 따라 설정 완료



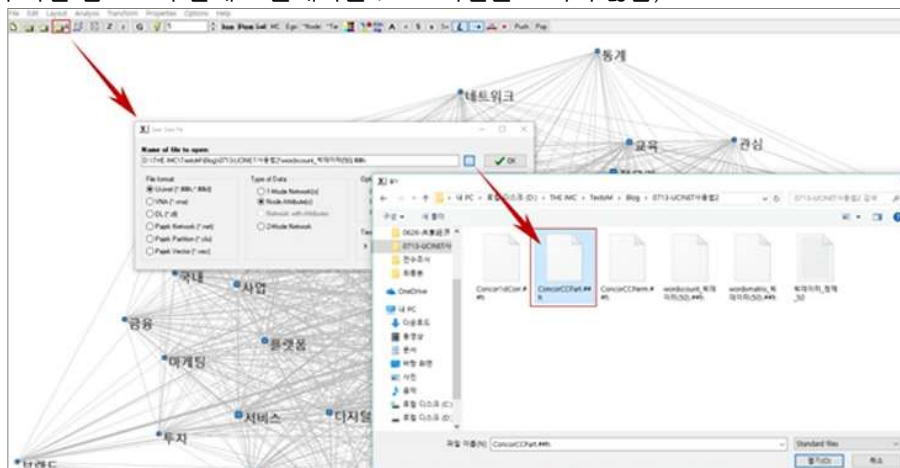
- ① : 저장된 '빅데이터_매트릭스.##h'파일 업로드
- ② : 화살표 선택안함
- ③ : Cluster Diagram의 Depth 설정
Depth를 높게 설정할 수록 클러스터(그룹)
갯수를 많이 만들 수 있음
- ④ : CONCOR분석 데이터셋(dataset) 저장위치
지정

▶ CONCOR분석에 필요한 6개의 dataset 생성



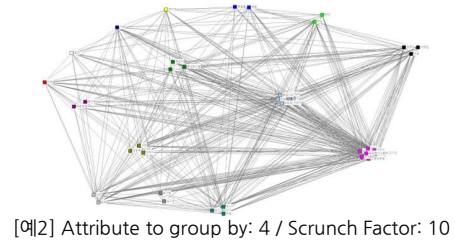
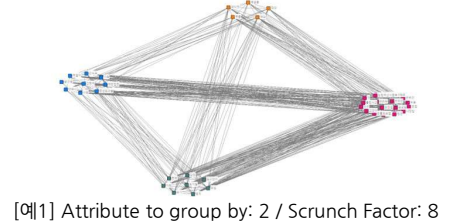
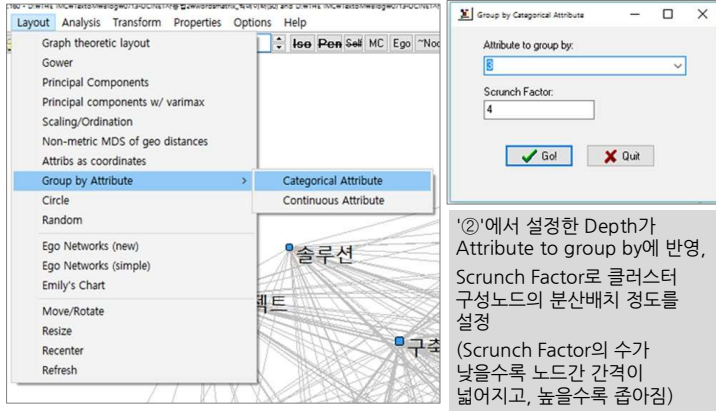
03 UCINET/NetDraw를 활용한 CONCOR 분석

③ NetDraw 오픈 → 네트워크 그래프 먼저 제작 후 보조 파일 업로드
(여기서 파일 업로드시 실제로 존재하는 .##d 파일은 보이지 않음)



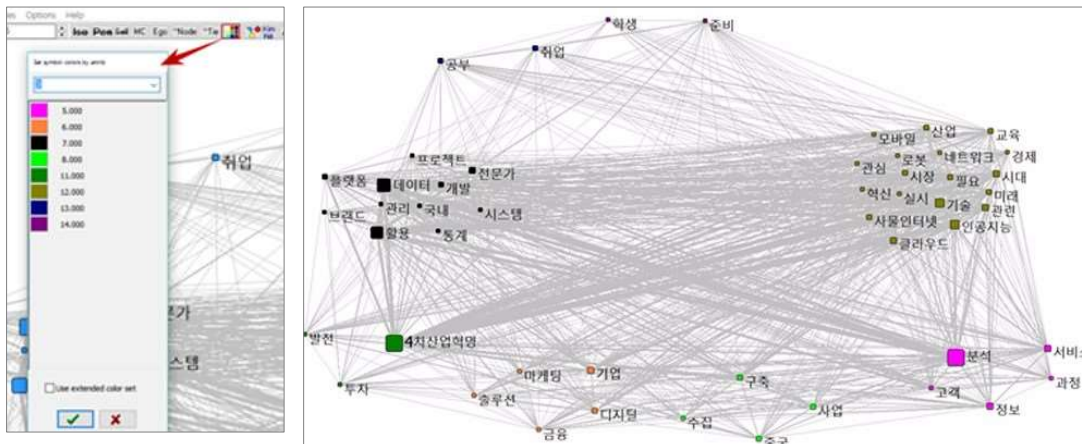
03 UCINET/NetDraw를 활용한 CONCOR 분석

④ 'Layout' - 'Group by Attribute' - 'Categorical Attribute' 차례로 클릭



03 UCINET/NetDraw를 활용한 CONCOR 분석

⑤ 클러스터별 색상 입력 ※Attribute to group by에서 선택한 숫자와 동일하게 선택해야 그룹의 갯수에 맞춰 노드 색깔 추가 'Properties' - 'Nodes' - 'Symbols' - 'Color' - 'Attribute-based'



04 UCINET/NetDraw를 활용한 에고네트워크 분석

- ① UCINET: 상단 메뉴에서 'Network' - 'Ego networks' - 'Density' 차례로 클릭

- 이수상(2019). 네트워크 분석 방법론. 논형.
- 장정윤, 엄정애(2017). 만 5세 하늬반 유아들의 또래관계 탐색. *유아교육연구*, 37(4), 877-913.
- 최지은(2019). 키워드 네트워크 분석을 활용한 영유아 놀이 관련 연구동향 분석. *학습자중심교과교육연구*, 19(14), 605-626.
- 최지은(2020). 빅데이터를 활용한 유아 웰빙(well-being)에 대한 키워드 네트워크 분석. *육아지원연구*, 15(4), 5-26.
- Hanneman, R. A., & Riddle, M. (2020). UCINET의 활용과 이해: 소셜네트워크 분석(오명열 역). 퍼플. (원서출판 2005년)
- Wang, Y., Palonen, T., Hurme, T., & Kinos, J. (2019). Do you want to play with me today? Friendship stability among preschool children. *European Early Childhood Education Research Journal*, 27(2), 170-184.
- 웹사이트:
 - 텍스트SV 매뉴얼 <https://www.textom.co.kr/home/sub/view.php?id=report&no=22>
 - 텍스트 스페셜리스트 <https://blog.naver.com/PostList.naver?blogId=textom&categoryNo=38&from=thumbnailList&parentCategoryNo=0>

Q&A



감사합니다



최지은
jechoi@swu.ac.kr