

고유과제 콜로키움 자료집

청소년 이슈 관련 네트워크 분석

일 시 2020년 8월 25일(화) 10:00 ~ 12:00
장 소 한국청소년정책연구원 7층 대회의실
주 최 한국청소년정책연구원



청소년 이슈 관련 네트워크 분석

Aug. 25, 2020
@한국청소년정책연구원

정유경
연세대학교 근대한국학연구소 HK연구교수

Outlines

- 소셜 미디어와 소셜 네트워크
- 네트워크의 시각화
 - 네트워크 분석과 관련된 이론들
 - 중요 측정치
 - 네트워크 클러스터 분석
- 네트워크 분석의 활용
 - 공기어 분석, 지리정보 결합, 시계열 자료 결합
- 청소년 관련 이슈 분석, 네트워크분석 연구 사례
- (+) 청소년 정책관련 해외 연구동향 네트워크 분석

소셜 미디어와 소셜 네트워크



이미지 출처: <https://phys.org/news/2019-06-exploring-gossip-social-networks.html>
<https://www.whatmobile.net/features/article/5-little-known-social-media-marketing-tips-that-will-take-your-marketing-to-new-heights>

3

네트워크 분석의 시작

- 사회학의 사회 연결망 (Social Network), 사람들 사이의 관계망
 - 사람들이 하는 행위의 결과물
 - 사람들이 하는 차후 행위의 선택에 영향을 주는 역할을 함
- 관련 이론 및 개념들
 - **자리매김 이론 (Theory of Embeddedness)**: 특정한 형태의 연결망 안에 '자리 매겨져 있는' 사람 (또는 객체)의 위치가 그들의 의식이나, 효용, 행위에 대한 보상(pay off)에 까지도 영향을 미친다고 보는 이론
 - **작은 세상 (Small world)**: 케빈 베이컨 게임 (6단계 게임), 여섯 다리만 건너면 지구 위에 사는 사람들은 모두 아는 사이 (Six Degrees of Separation)*
 - **유유상종 (Homophily)**: 대부분의 연결망에는 같은 속성을 가진 사람들이 끼리끼리 모이는 현상이 존재

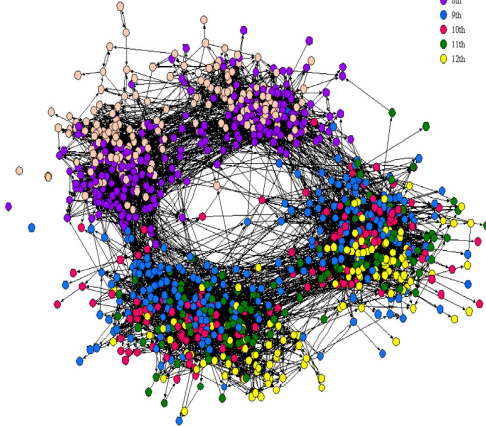
4

네트워크 분석을 통해 알 수 있는 것들

예)미국 고등학교 학생들의 네트워크

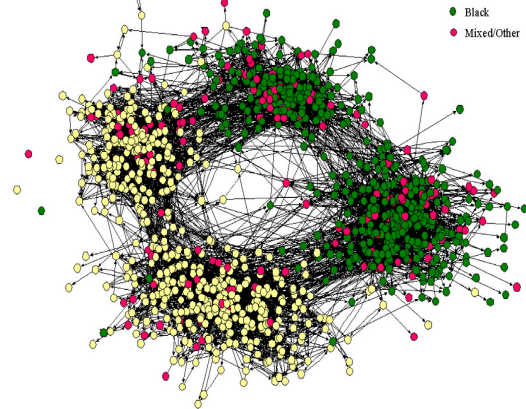
The Social Structure of "Countryside" School District

Points Colored by Grade



The Social Structure of "Countryside" School District

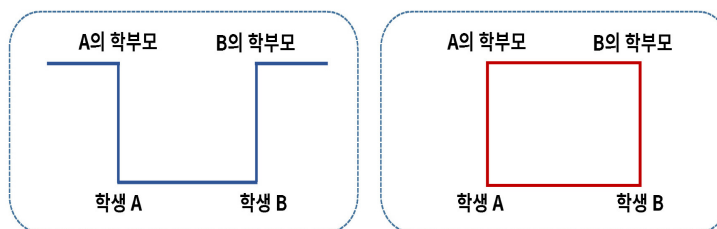
Points Colored by Race



이미지 출처: Moody, J. (2001). "Race, school integration, and friendship segregation in America." *American journal of Sociology*, 107(3), 679-716.

5

사회 연결망 분석



- 왼쪽 연결망: 학생 A와 B는 친구, 그들 부모들은 모르는 사이, 열려 있음
- 오른쪽 연결망: 학부모끼리 서로 연결, 닫혀 있음
- 학생들이 비행청소년이 될 확률이 어떤 연결망에서 더 높은가?
 - 답은 학부모끼리 정보가 흐르지 않는 열린 연결망
- 소셜 네트워크 분석의 정의 (Mitchell, 1969)
 - (사람으로 구성된)연결망의 특성을 통해 연결망에 포함된 사람들의 사회적 행위를 설명
 "...social network is ... one way of understanding behavior ..."

Mitchell, J. C. (Ed.). (1969). *Social networks in urban situations: analyses of personal relationships in Central African towns*. Manchester University Press.

6

네트워크 시각화

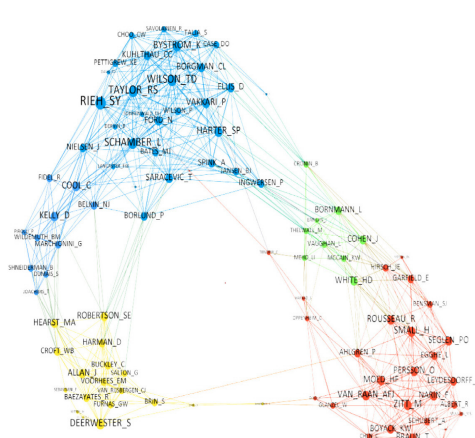
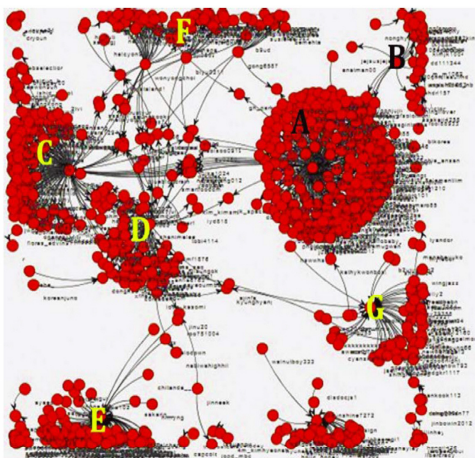
- 네트워크 분석의 첫 단계는 시각화
- 객체 사이의 관계를 2차 평면에 표현
 - 노드 (node or vertex) - 점
 - 엣지 (edge or link) - 선
- 결속집단 (clique)
- 클러스터 (cluster)
- 경로 (path)
- 거리 (length)
- 중심성 (centrality)



이미지 출처: <https://www.interaction-design.org/literature/article/how-to-display-complex-network-data-with-information-visualization>

7

Cluster, Community Detection



Song, M., Kim, M. C., & Jeong, Y. K. (2014). Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter. *Intelligent Systems, IEEE*, 29(2), 18-26.
Jeong, Y. K., & Song, M. (2016). Applying content-based similarity measure to author co-citation analysis. *ICConference 2016 Proceedings*.

8

중심성 Centrality

- 중심에 위치한 정도 측정
- 중심성 측정
 - 특정 한 노드가 다른 많은 노드들과 얼마나 연결되어있는지
 - 한 노드가 다른 모든 노드들에 도달하려면 몇 단계나 거쳐야하는
 - 예) 친구 망에서, 친구로 선택된 횟수
- 권력과 영향력 개념과 연결되어 가장 많이 쓰이는 지표
 - 대부분 중심성이 높은 개인은 특별한 사회 경제적 지위를 갖는 사람
 - 조직의 경우 중심성이 높을수록 생존율이 높거나 기업 성과가 좋게 나타남
- 한 노드가 그 주위의 다른 점들과 직접 연결된 정도가 높을수록 그 노드의 지역 중심성(local centrality)은 높아짐
- 한 노드가 연결망 전체의 연결 구조에서 전략적으로 중요한 자리를 차지할수록 전체 중심성(global centrality)은 높아짐

9

주요 중심성 지표

- Degree Centrality: 연결정도(degree)로 지역 중앙성을 측정할 수 있음
 - 내향 중심성 (in-degree centrality), 외향 중심성 (out-degree centrality)
- Closeness centrality: 인접 중심성 (근접 중심성), 한 노드의 전체 중심성을 표현하는 지표
 - 다른 노드들과 인접성(closeness) 혹은 거리(distance)로 측정
 - 두 노드 간의 거리는 두 노드를 연결하는 최단거리, 즉 경로 거리를 측정함
 - 근접 중심성은 중요한 노드일수록 다른 노드까지 도달하는 경로가 짧을 것이라는 가정을 바탕으로 함
 - 경로 거리의 합이 가장 작은 노드가 전체 중심성이 가장 높은 네트워크 전체의 중심을 차지하는 노드
- Betweenness Centrality: 사이 중심성
 - 한 노드가 연결망 내의 다른 노드들 ‘사이에’ 위치하는 정도
 - 최단 경로 위에 위치하면 할 수록 그 노드의 사이 중심성은 높아짐
 - 다른 노드들 사이에서 브로커 역할을 하는 정도
 - 예) 남들이 다른 사람들에게 도달하기 위해 ‘나’를 거쳐야하는 정도, 뚜쟁이는 사이 중심성이 높은 사람.

10

Prestige Centrality

- 위세 중심성
 - 연결된 상대방의 중요성에 가중치를 주는 중심성 지표
- 다른 여러 행위자와 관계를 맺은 강자와 연결되면 자신의 영향력을 증가시킬 수 있음
 - 자신이 지위나 권력이 높은 사람과 많이 접촉할 수록 본인의 위세가 높아진다는 개념
- 보나시치 권력 지수 또는 보나시치 중심성 지수라고도 함
 - Bonacich power centrality
 - Katz prestige
 - Eigenvector centrality
- PageRank
 - Google에서 개발 (Brin & Page, 1998)
 - 전체 웹을 복잡한 네트워크로 간주, 개별 웹 페이지를 네트워크의 구성요소인 노드, 링크를 엣지로 가정
 - 이용자들은 자신들이 좋은 페이지라고 생각하는 웹 페이지를 링크

Brin, S., & Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18), 3825-3833.

11

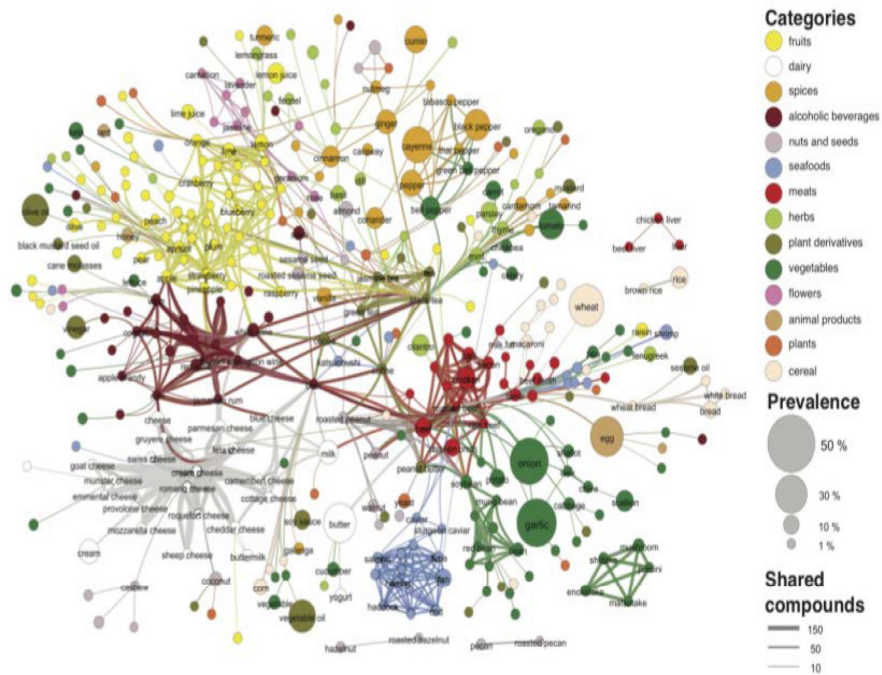


FIG.2: The backbone of the flavor network

Ahn, Y. Y., Ahnert, S. E., Bagrow, J. P., & Barabási, A. L. (2011). Flavor network and the principles of food pairing. *Scientific reports*, 1, 196.

12

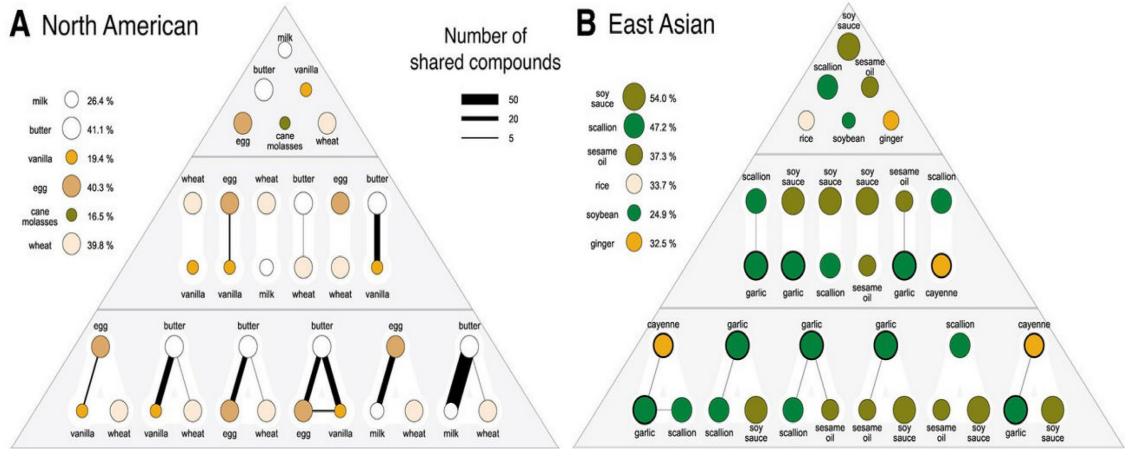


FIG.4: Flavor principles

Ahn, Y. Y., Ahnert, S. E., Bagrow, J. P., & Barabási, A. L. (2011). Flavor network and the principles of food pairing. *Scientific reports*, 1, 196.

13

C Co-occurrence in recipes

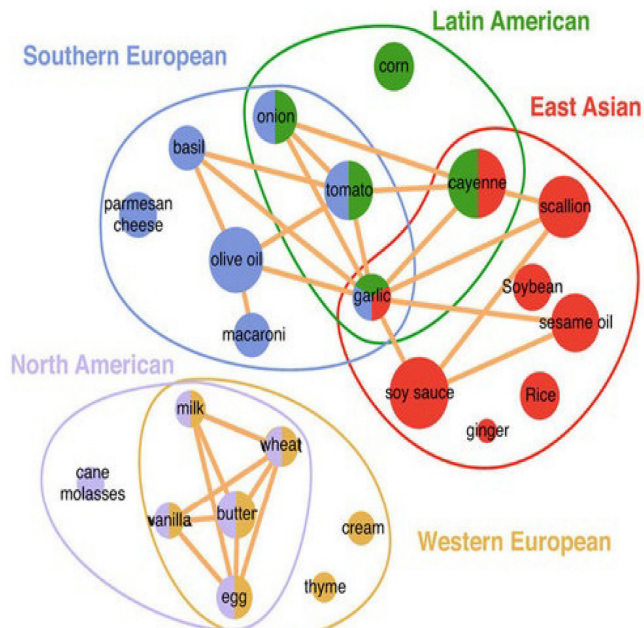
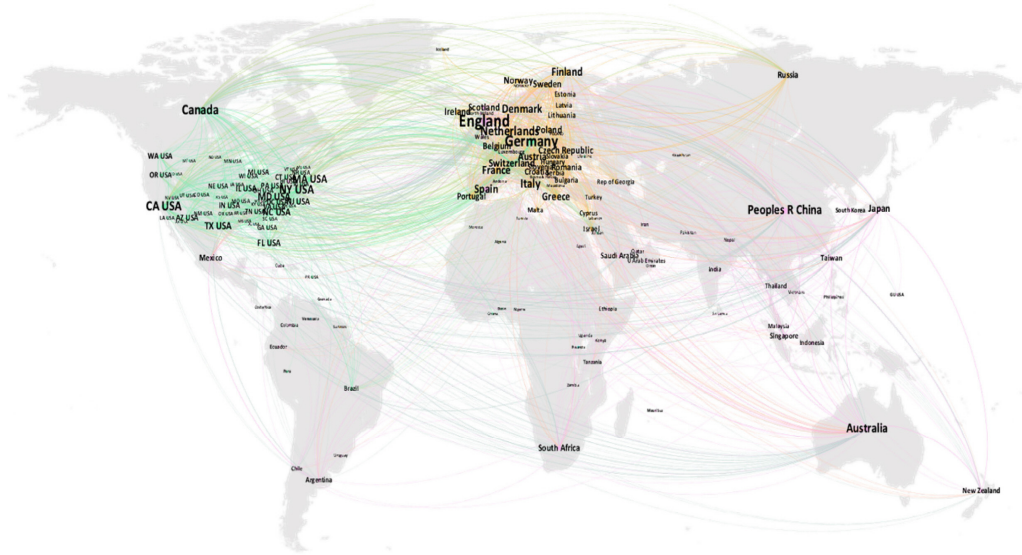


FIG.4: Flavor principles

Ahn, Y. Y., Ahnert, S. E., Bagrow, J. P., & Barabási, A. L. (2011). Flavor network and the principles of food pairing. *Scientific reports*, 1, 196.

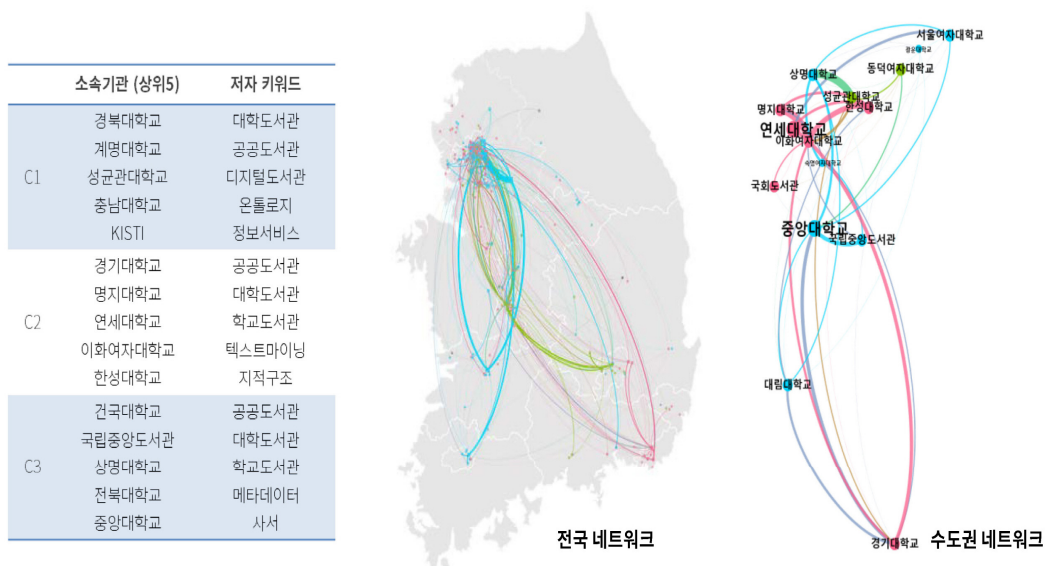
14

지리 정보의 결합 디지털 인문학 분야의 협업 네트워크



15

문헌정보학 분야의 국내 협업 네트워크

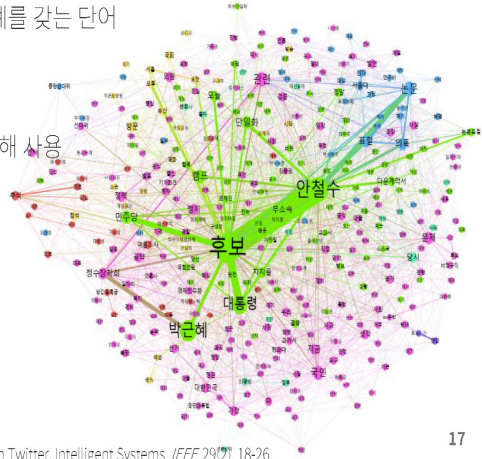


정유경. (2017) 국내 문헌정보학분야 학술지의 협업 주제 분석을 위한지리적 접근에 대한 실험적 연구. *한국정보관리학회:학술대회논문집* 24, 19-22.

16

네트워크 분석의 활용 공기어 분석 Co-word Analysis

- 두 개의 단어가 주어진 문헌의 분석 단위 안에서 함께 출현하는 횟수를 기반으로 단어의 연관성 및 문헌 집단의 특성을 파악
 - 공기: 두 단어가 같은 문맥에서 함께 나타나는 현상, Co-occurrence
 - 공기어: 같은 문맥 안에서 함께 나타나 서로 밀접한 의미관계를 갖는 단어
- 텍스트로부터 동시에 출현하는 단어의 쌍을 추출
- 단어 쌍을 네트워크화 하여 분석
 - 해당 텍스트로부터 주제어와 그들 간의 관계를 파악하기 위해 사용
 - 단어군집을 분석하기에 용이
- 단어 쌍
 - source, target, weight
 - (대통령, 후보), 30

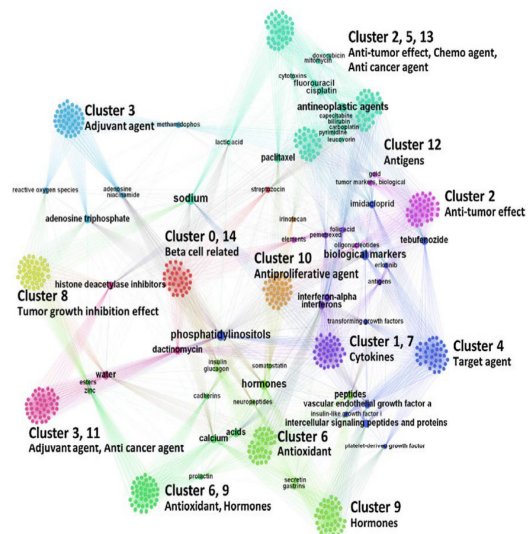
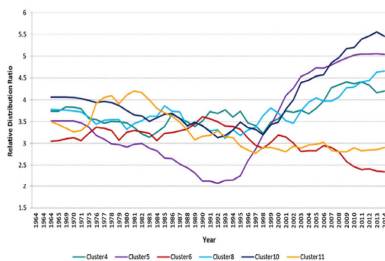


Song, M., Kim, M. C., & Jeong, Y. K. (2014). Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter. Intelligent Systems, IEEE, 29(2), 18-26.

17

클러스터 분석과 시계열 자료의 결합

	Drug 1	Drug 2	Drug 3	Drug 4	Drug 5
Cluster 0 (beta cell related)	Insulin	Glucose	Glucagon	Calcium	Hormones
Cluster 1 (Cytokines)	Interleukin-1	Cytokines	Interleukin-1	Interleukin-1	Tamoxifen
Cluster 2 (Anti-tumor effect)	Fluorouracil	Cyclophosphamide	Fluorouracil	Fluorouracil	Cholesterol
Cluster 3 (Adjuvant agent)	Cisplatin	Docetaxel	Erlotinib	Docetaxel	Zinc
Cluster 4 (Target agent)	Erlotinib	Erlotinib	Erlotinib	Erlotinib	Erlotinib
Cluster 5 (Chemo agent)	Fluorouracil	Fluorouracil	Fluorouracil	Fluorouracil	Fluorouracil
Cluster 6 (Anticancer agent)	Docetaxel	Docetaxel	Docetaxel	Docetaxel	Docetaxel
Cluster 7 (Anti-folate)	Pemetrexed	Pemetrexed	Folic acid	Folic acid	NS-398
Cluster 8 (Tumor growth inhibition effect)	Reactive oxygen species	Merferrin	Oxygen	Superoxides	Glutathione
Cluster 9 (Hormones)	Gastrin	Cholecystokinin	Peptides	Tretinoin	Bombesin
Cluster 10 (Antiproliferative agent)	VEGF-rt	Everolimus	Sunitinib	Sunitinib	Sunitinib
Cluster 11 (Anti-cancer agent)	Streptozotocin	Streptozotocin	Streptozotocin	Streptozotocin	Streptozotocin
Cluster 12 (Antigens)	Glucagon-like peptide 1	Glucagon-like peptide 1	Glucagon-like peptide 1	Glucagon-like peptide 1	Glucagon-like peptide 1
Cluster 13 (Anti-cancer agent)	Fluorouracil	Cisplatin	Cisplatin	Cisplatin	Cisplatin
Cluster 14 (Neurotransmitter)	Calcium	Melatonin	Melatonin	Melatonin	Melatonin



Jeong, Y. K., Heo, G. E., Kang, K. Y., Yoon, D. S., & Song, M. (2016). Trajectory analysis of drug-research trends in pancreatic cancer on PubMed and ClinicalTrials.gov. Journal of Informetrics, 10(1), 273-285.

18

청소년 관련 이슈 분석 연구 사례1 (강주연, 이이든, 2019)

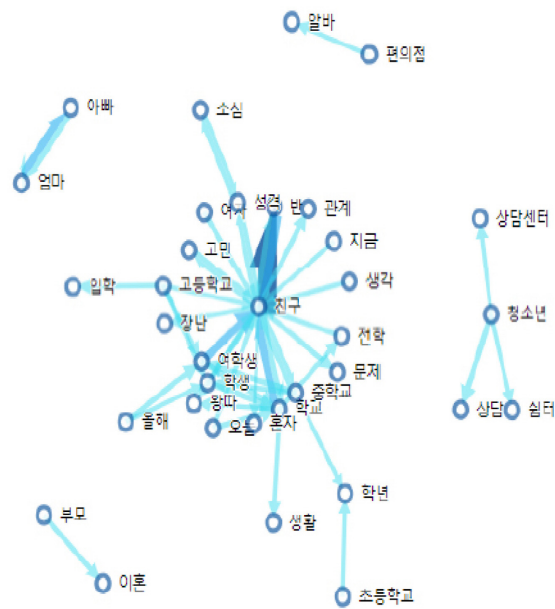
- “네이버 지식iN을 활용해 ‘청소년상담1388’ 키워드가 포함된 빅데이터를 추출하고 분석하여 ‘청소년상담1388’에 관한 의미 있는 결과를 도출하고, 효과적인 청소년 상담서비스 제공을 위한 기초자료를 만드는 데 목적”
- 연구 데이터 및 방법
 - 46,859건의 데이터
 - 사전처리와 품사분석 등의 전처리과정을 통해 단어추출, N-gram과 의미연결망을 통해 데이터를 분석, 시각화
- 연구결과
 - 학교생활 관련, 가정 관련, 심리적 상황을 표현하는 단어, 근로와 관련된 단어 등 생활환경 관련 단어 발견
 - 의미연결망 분석 - 4개의 클러스터가 형성
 - 그룹1: 청소년의 학교와 교육관계에 관련된 키워드 그룹2: 청소년의 학업과 관련된 키워드
 - 그룹3: 청소년의 가족 환경 변화와 관련된 키워드 그룹4: 청소년의 근로와 관련된 키워드 순으로 크기가 형성
 - 비정형 데이터를 통해서도 공공서비스를 이용하는 이용자들의 이슈 영역을 파악할 수 있음을 확인

강주연, 이이든. (2019). 텍스트 마이닝을 활용한 '청소년상담 1388'빅데이터 분석: 2011~2018 네이버 지식 iN 을 중심으로. *청소년상담연구*, 27(2), 127-1473.

19

〈표 6〉 ‘청소년상담1388’ 검색결과에 대한 N-gram

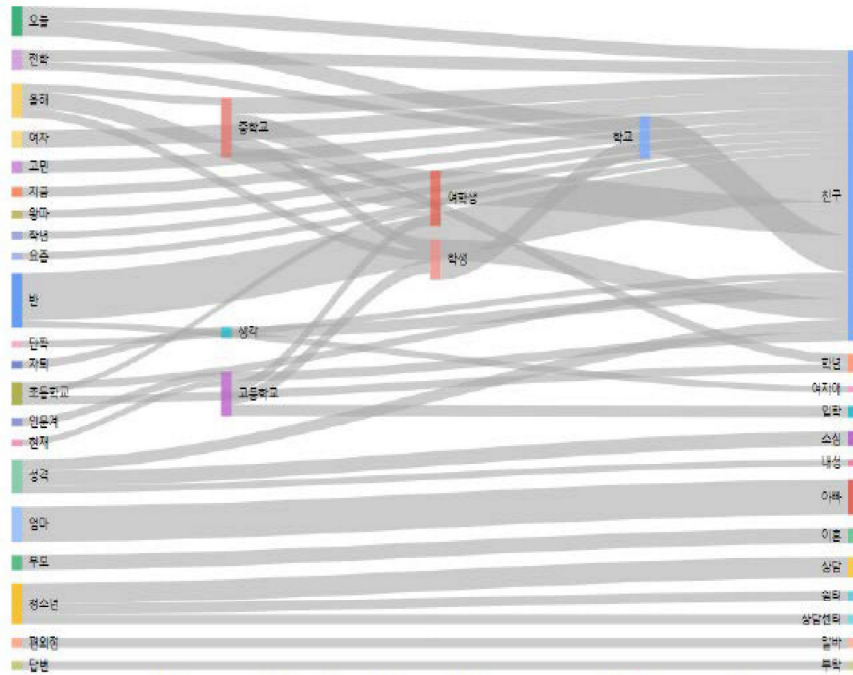
	키워드	→	키워드	빈도
1	반		친구	391
2	학교		친구	302
3	엄마		아빠	289
4	여학생		친구	262
5	청소년		상담	173
6	중학교		여학생	173
7	학생		친구	164
8	친구		반	157
9	학생		학교	157
10	아빠		엄마	154
11	여학생		학교	147
12	친구		학교	147
13	올해		여학생	139
14	여자		친구	134
15	친구		관계	131
16	오늘		학교	129
17	중학교		친구	128
18	친구		문제	124
19	부모		이혼	122
20	성격		소심	122
21	같은반		친구	110
22	친구		성격	106
23	고민		친구	101
24	친구		장난	100
25	고등학교		입학	99



(그림 2) '청소년상담1388' 검색결과에 대한 N-gram 네트워크 결과

강주연, 이이든. (2019). 텍스트 마이닝을 활용한 '청소년상담 1388' 빅데이터 분석: 2011~2018 네이버 지식 iN 을 중심으로. *청소년상담연구* 27(2), 127-147.

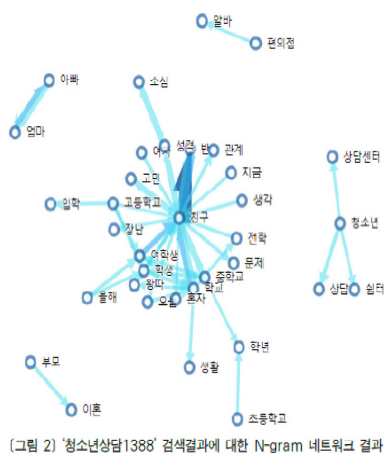
20



(그림 3) '청소년상담1388' 검색결과에 대한 N-gram 네트워크워드트리 결과

강주연, 이이든. (2019). 텍스트 마이닝을 활용한 '청소년상담 1388'빅데이터 분석: 2011~2018 네이버 지식 iN 을 중심으로. *청소년상담연구*, 27(2), 127-1473.

21



(그림 2) '청소년상담1388' 검색결과에 대한 N-gram 네트워크 결과

〈표 7〉 ‘청소년상담1388’ 검색결과에 대한 관련 키워드 간 연결중심성 분석결과

2011년 ~ 2018년											
순 위	단어	중심성 지수	순 위	단어	중심성 지수	순 위	단어	중심성 지수	순 위	단어	중심성 지수
1	친구	0.087	11	반	0.017	21	고민	0.012	31	질문	0.009
2	학교	0.042	12	여학생	0.016	22	혼자	0.011	32	학원	0.009
3	엄마	0.026	13	지금	0.016	23	선생	0.011	33	이유	0.009
4	공부	0.022	14	사람	0.015	24	왕따	0.011	34	남자	0.009
5	부모	0.020	15	여자	0.014	25	돈	0.010	35	맨날	0.008
6	생각	0.019	16	중학교	0.013	26	스트레스	0.010	36	애기	0.008
7	오늘	0.018	17	고등학교	0.013	27	기출	0.010	37	마음	0.008
8	아빠	0.018	18	성격	0.013	28	문제	0.009	38	현재	0.008
9	학생	0.018	19	알바	0.012	29	상당	0.009	39	아이	0.008
10	집	0.017	20	전학	0.012	30	시간	0.009	40	남자애	0.008

강주연, 이이든. (2019). 텍스트 마이닝을 활용한 '청소년상담 1388'빅데이터 분석: 2011~2018 네이버 지식 IN 을 중심으로. *청소년상담연구*, 27(2), 127-1473.

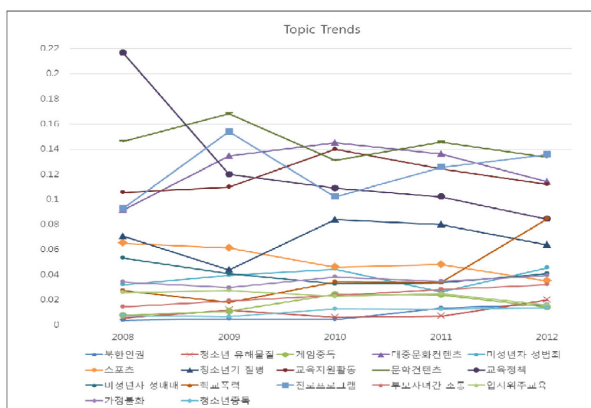
22

청소년 관련 이슈 분석 연구 사례2 (조주연, 김현숙, 조민제, 2018)

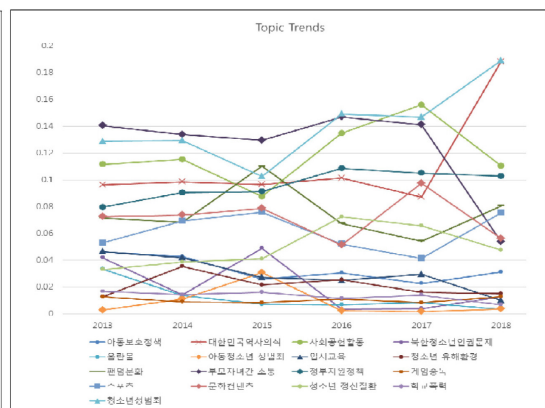
- “인터넷 뉴스 기사에 실린 청소년 문제와 관련된 이슈와 그 동향을 파악”
- 연구방법
 - ‘조선일보’, ‘동아일보’, ‘중앙일보’ 대상, 2008년~2018년 3월, 총 8,110건의 청소년 관련 기사가 분석에 활용
 - 토픽모델링의 일종인 잠재디리클레 할당(Latent Dirichlet Allocation)이 사용
- 연구결과
 - 2008년~2012년 주요 토픽: 문학컨텐츠, 교육정책, 대중문화컨텐츠, 진로프로그램
 - 문학컨텐츠의 상업화가 가속화되던 시기, 엔터테인먼트 토픽이 등장, 문화 키워드와 함께 진로문제가 부상
 - 2013년~2018년 주요 토픽: 주요 토픽은 청소년 성범죄, 부모자녀 간 소통, 사회공헌활동 등
 - 사회적 이슈와 청소년문제가 함께 다루어짐, 이전에 비해 청소년 이슈가 다양화
- “인터넷 뉴스 사이트에서 청소년과 관련하여 어떤 주제가 기사화되어 왔는지 파악하고, 청소년문제를 이해하는데 사회문화적 환경과 교육정책의 변화가 고려되어야 함을 파악”

조주연, 김현숙, 조민재. (2018). 텍스트마이닝을 활용한 청소년 문제의 이슈 변화 분석-2008년~2018년 인터넷 신문기사를 중심으로. 교육혁신연구, 28(4), 461-487.

23



〈그림 III-1〉 2008-2012년 토픽추이(x축 : 연도, y축 : 비중)

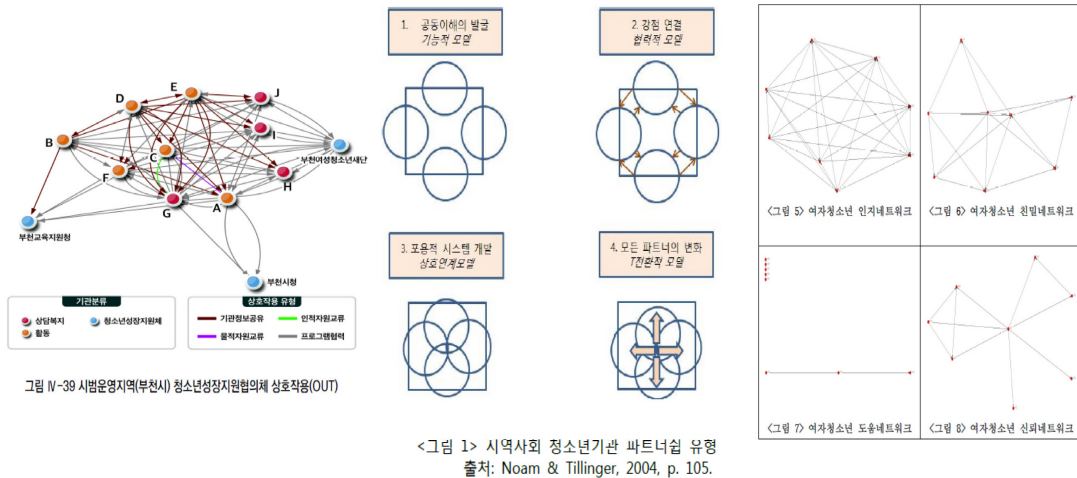


〈그림 III-2〉 2013-2018년 토픽추이(x축 : 연도, y축 : 비중)

조주연, 김현숙, 조민제, (2018). 텍스트마이닝을 활용한 청소년 문제의 이슈 변화 분석-2008년~2018년 인터넷 신문기사를 중심으로, *교육혁신연구*, 28(4), 461-487.

24

청소년 관련 네트워크 분석 사례

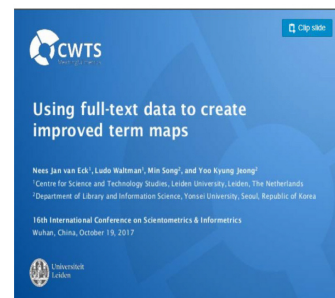


최인재, 이윤주, 송민경, 조윤정. (2019). 지역사회 네트워크를 활용한 청소년 성장지원 정책 추진체계 구축방안 연구 I. 한국청소년정책연구원 연구보고서, 1-446.
 천정용. (2013). 지역사회 청소년 네트워크와 외국의 실천경험 연구. *청소년시설향경*, 11(4), 145-156.
 우수영, 정지윤, & 김재원. (2020). 중장기센터 내 청소년간 사회적 지지 네트워크가 자립의지와 자립역량에 미치는 영향. *청소년학연구*, 27(4), 23-41.

25

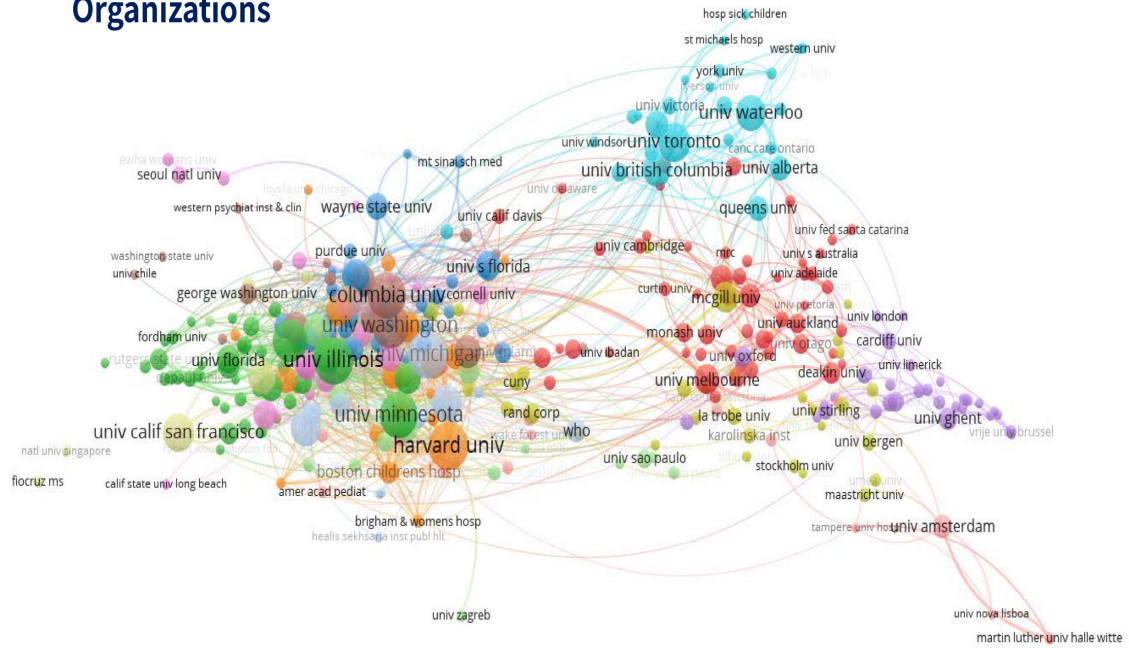
파일럿 분석 청소년 정책 관련 연구동향 네트워크 분석

- 청소년 정책 관련 연구 논문
 - Web of Science에서 'youth', 'adolescent', 'policy'로 검색
 - 연도: 1990~2020
 - 수집 건수: 4,025건
- 네트워크 구축
 - 연구기관 네트워크: 공저 기관, 국가 대상
 - 단어 동시 출현 네트워크(공기어 네트워크)
 - 논문에 부여된 WOS 키워드를 대상
 - 서지데이터로부터 직접 추출한 용어
- 네트워크 시각화 도구
 - VOSviewer (<https://www.vosviewer.com/>)

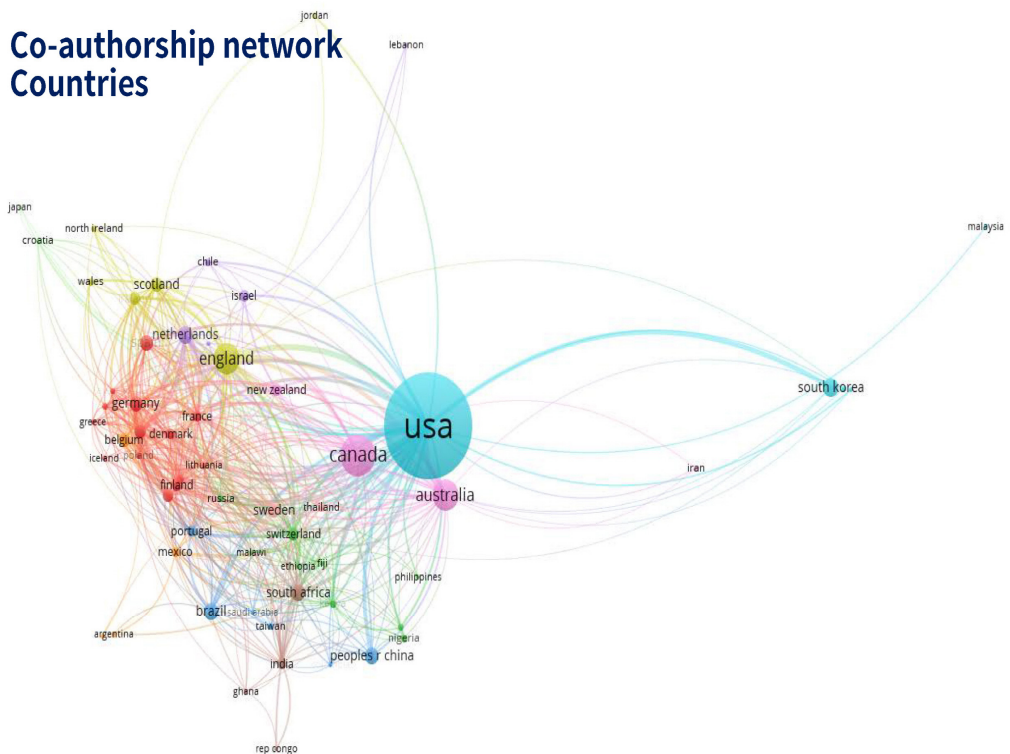


26

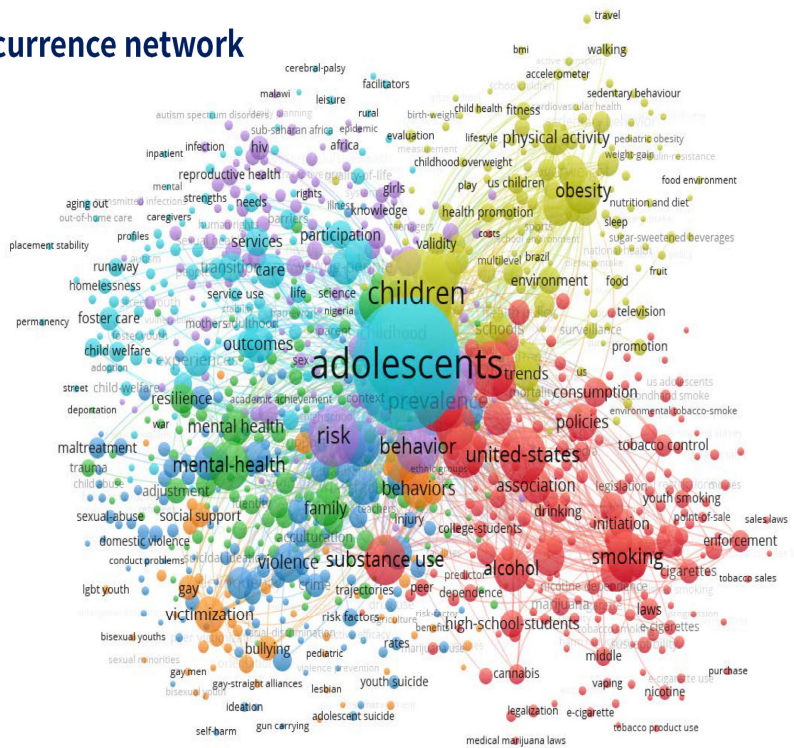
Co-authorship network Organizations



Co-authorship network Countries

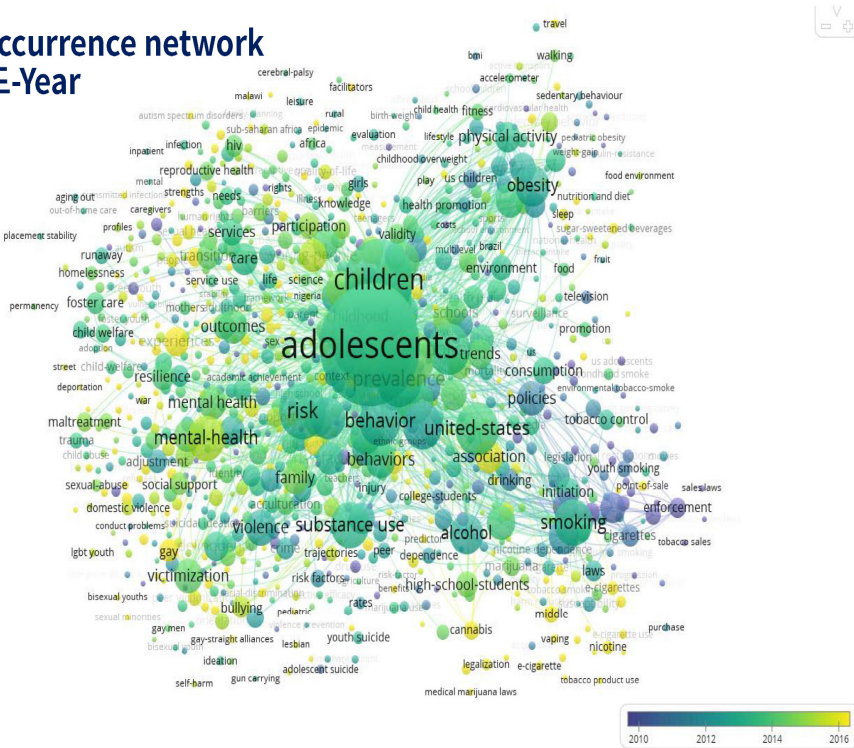


Co-occurrence network 키워드



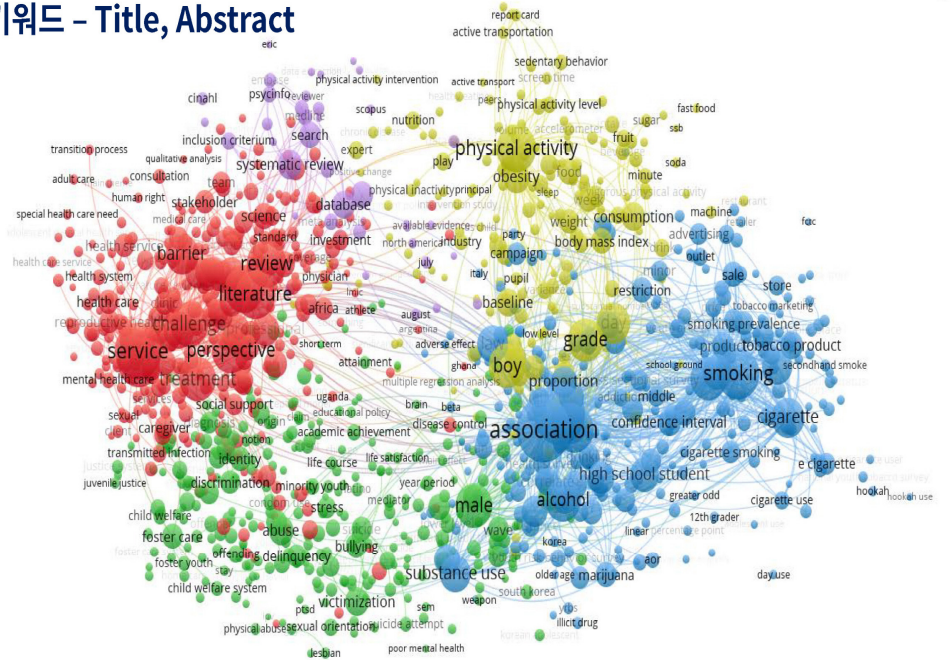
29

Co-occurrence network

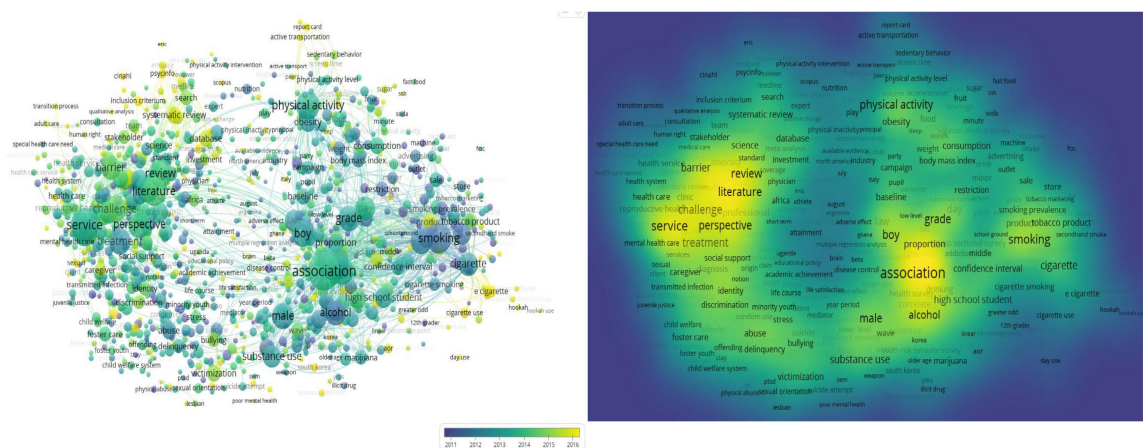


30

Co-occurrence network 키워드 - Title, Abstract



31



32

Thank you
yookyungjeong@gmail.com

MEMO

MEMO

고유과제 콜로키움 자료집

청소년 이슈 관련 네트워크 분석