

# Social Network Analysis (사회연결망분석) 의 이해

- 데이터로 보는 사회 현안

한국인사이트연구소  
김덕진 부소장



# 김덕진

빅데이터 분석/IT컨설턴트  
 한국인사이트연구소 부소장  
 세종사이버대학교 컴퓨터공학과 겸임교수  
 경희대 경영대학원 콘텐츠MBA 출강

## TV 및 라디오 출연분 (2021)

KBS1R 빅데이터로보는세상 고정출연  
 MBC 손에잡히는경제 IT패널  
 KBS2TV 통합뉴스룸ET 고정출연  
 TBS 명랑시사 "명랑한 트렌드" 코너 진행  
 유튜브 "삼프로TV" 빅사이즈 코너 진행

## 과거출연분

KBS T타임 "데이터분석 SHOW" SBS 생활경제 "건강의 법칙 NOW"  
 MBN 황금알 빅데이터전문가 SBSCNBC 블루베리 빅데이터읽어주는남자  
 TVN 쿨까당 생활정보분석 KBS1라디오 경제투데이 "궁금한 IT트렌드"  
 생생라디오매거진 "매거진 플러스" 이주향의 인문학산책 "인문학카페"

## 언론보도

전자신문 [인터뷰] 소셜네트워크 분석 김덕진 컨설턴트 머니투데이 [인터뷰] 김덕진  
 컨설턴트 "데이터의 가치 무궁무진 사이언스타임즈, 전문가에게 듣는다  
 "SNS 여론은 진짜 여론일까?"  
 KBS 스페셜 "고급 정보인가 악성 루머인가... '짜라시'의 실제  
 더PR "당신을 속이는 다크패턴 7"

## 주요 프로젝트 & 업무분야

빅데이터 분석 및 소셜미디어 컨설팅  
 2021 OO기업 신사업 발굴 전략컨설팅  
 2021 skt 구독경제 플랫폼 연구  
 2021 스타트업 분석 서비스 구축  
 2020 국가인권위원회 / 한국관광공사 빅데이터 전략분석  
 2019 프로스포츠험브 종목별 소셜빅데이터 분석  
 2018 사회적참사특조위 / 대기업 빅데이터 분석  
 2017 기관/지자체 전략수립  
 2016 통신사 / 외국계IT기업 SNS 트렌드 전략분석  
 2016 OO기관 / KBS 소셜미디어 분석  
 2015 병원 소셜미디어 활용 컨설팅  
 2013-4 온라인 콘텐츠 분석  
 2012 대선 빅데이터 분석 (KBS 추적 60분)  
 2011 H그룹 조직네트워크 분석  
 2011 총선 OOO당 캠프 소셜 분석  
 2010 3개사 지식 네트워크 분석

## 강연 및 외부교육

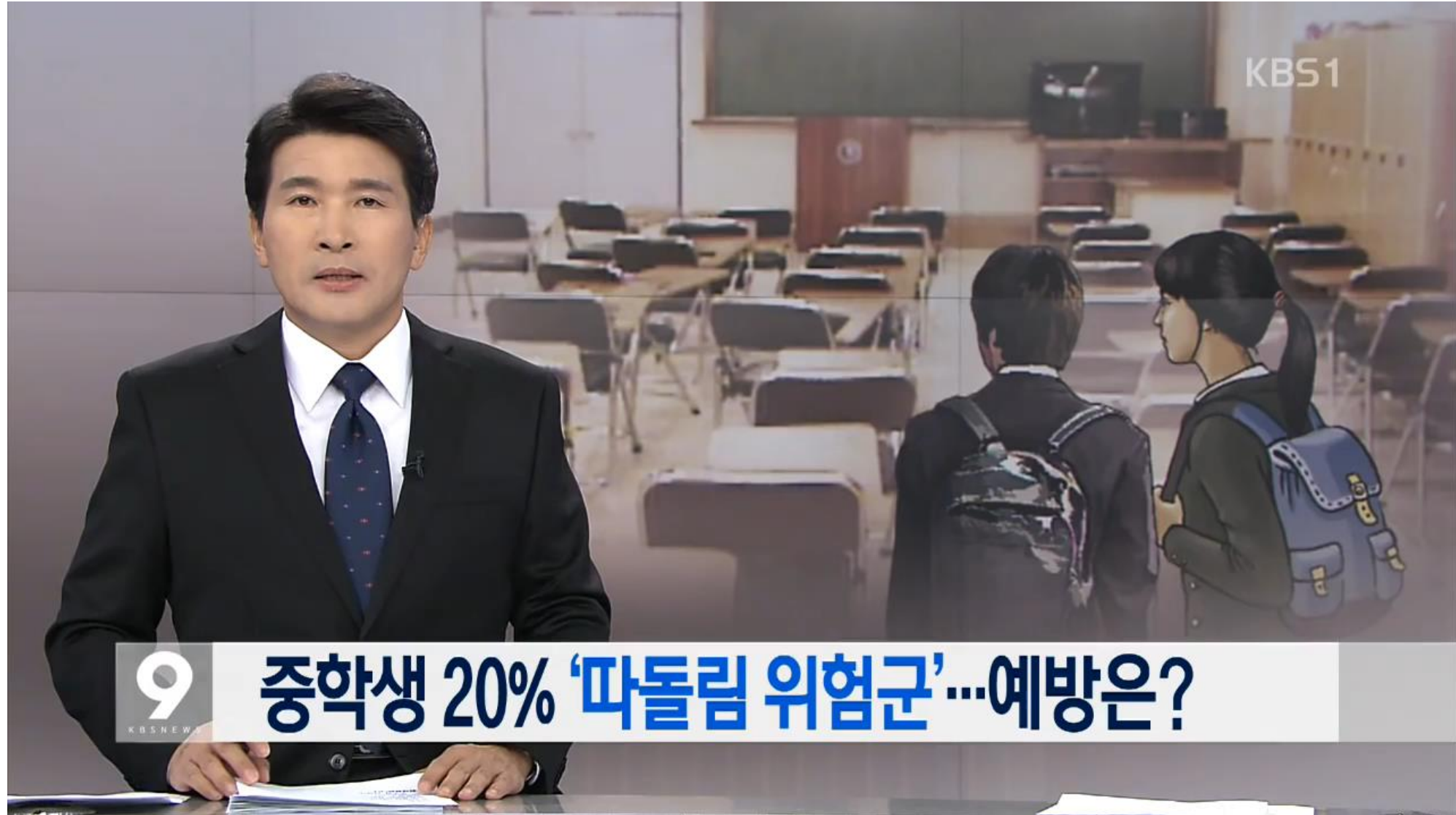
소셜 미디어 교육 및 IT 트렌드 강연  
 디지털 전환과 적용 (DT / DX)  
 4차산업혁명시대와 빅데이터의 이해  
 인공지능트렌드와 윤리  
 소셜빅데이터 분석 강의  
 뉴미디어의 전략적 활용  
 시대와 세대의 이해 (MZ 세대)

## 주요 강의처

유한김벌리/포스코 전사대상 교육  
 서울대학교 사회학과 학기수업  
 공공기관 강연  
 서울시립대, 고려대, 명지대 특강  
 삼성생명, 현대차, 기아차 등 기업체 강의

# 데이터분석을 통한 왕따막기?

데이터 분석을 통해 왕따를 막을 수 있을까?



# 데이터분석을 통한 왕따막기?

데이터 분석을 통해 왕따를 막을 수 있을까?

KBS NEWS

## “교실은 정글” 중학생 20% ‘왕따’ 가능성 높다

2015. 03. 11

### ◆ ‘따돌림 위험’ 분석해봤더니... ‘불안정’ 관계가 무려 58%

요즘 학부모님들의 가장 큰 걱정은 ‘집단 따돌림’ 이른바 왕따 문제일 겁니다.

문제는 이같은 따돌림이 잘 드러나지 않기 때문에 피해학생이 얘기하기 전에는 선생님이나 학부모가 파악하기 어렵고, 일단 피해가 드러났을 때는 걸잡을 수 없이 확대

된 **‘따돌림 위험’이 있는지, 미리 알 수 있다면**

는 이릅니다.

.....

“교실은 정글입니다. 관계맺기에 어려움을 겪는 아이들은 아무리 상담해도 실제 교실에서 살아남기는 정말 힘들어요. ‘왕따’프로그램이 그런 부분에 주목해야 됩니다.”

이번 프로그램에 참여했던 교사의 말입니다. 교사들은 이 연구 결과 ‘따돌림’ 위험이

**학생들을 미리 알아서 맞춤 지도가 가능했다는 점**

**문제를 예방하는 방향**야 한다는 것이 전문가들의 지적입니다.





# 데이터 분석을 통한 왕따막기?

## 데이터 분석을 통해 왕따를 막을 수 있을까?

### 기존 인성 검사



- 나는 친구들과 어울리기를 좋아한다
- 나는 최근 외롭다고 느낀다
- 나는 최근 우울하다고 느낀다

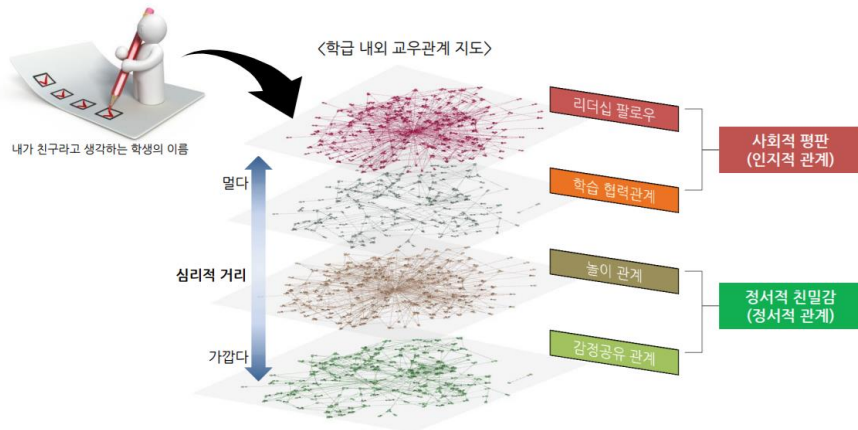
내가 생각하는 나의 모습

### 다면적 관계망 진단 방식



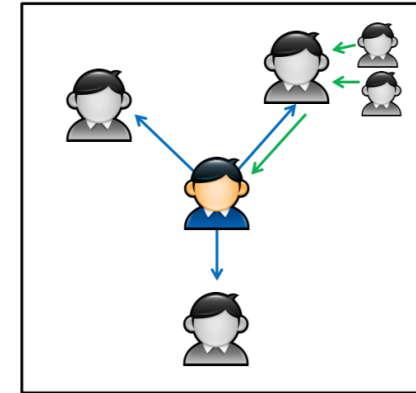
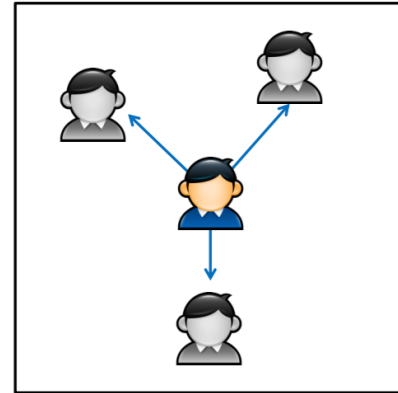
다른 학생들이 생각하는 나의 모습

긍정적인 설문을 통해 학급 내의 학생들 사이에 형성된 다층적인 관계를 조사합니다.

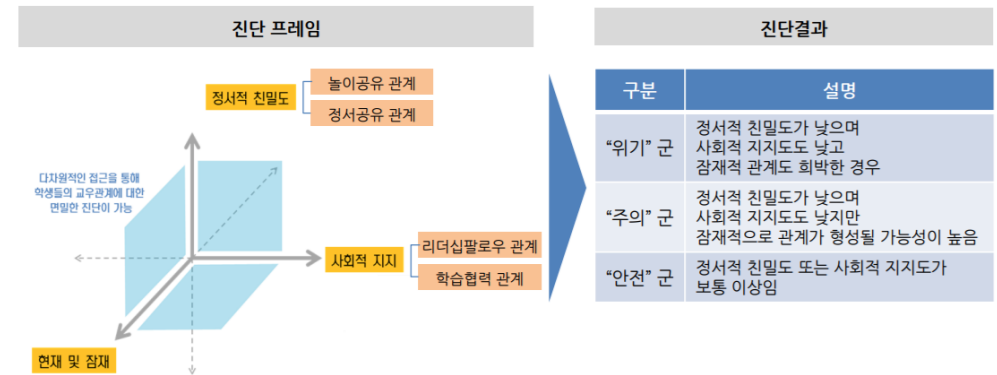


SNA 기법을 통해 “안정적인 관계”의 정도를 측정합니다.

- 내가 얼마나 많이 지목했는가? X
- 얼마나 많이 지목 받았는가? O
- 누구에게 지목 받았는가? O
- 얼마나 서로 지목했는가? O



8개의 관계망을 고려하여 학급 내 “관계적 약자” 위치에 처한 학생을 발견합니다.

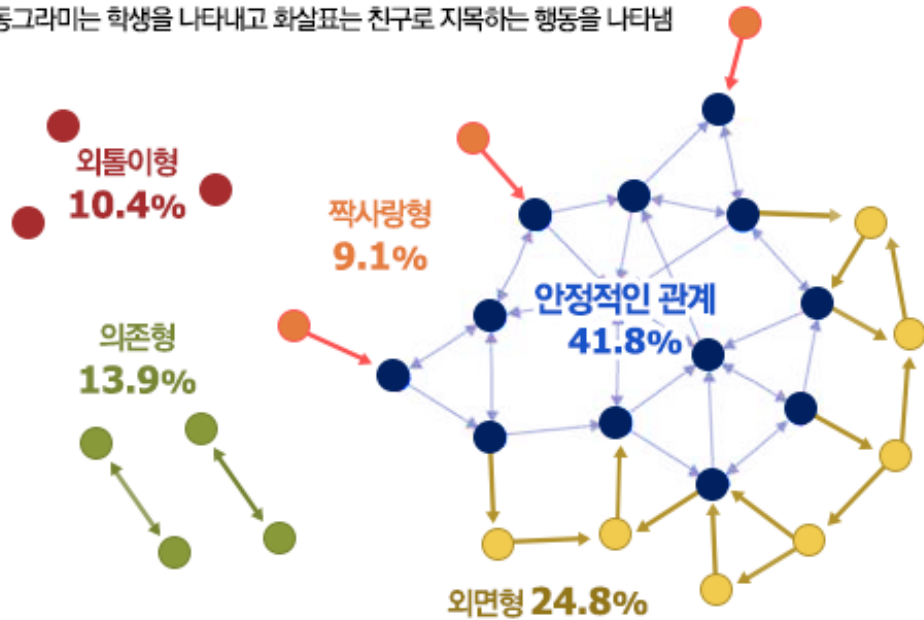


# 데이터분석을 통한 왕따막기?

데이터 분석을 통해 왕따를 막을 수 있을까?

## 2014년 중학교 교우관계 분석으로 나타난 따돌림 현황

9개 중학교 5,470명의 교우관계 진단  
 동그라미는 학생을 나타내고 화살표는 친구로 지목하는 행동을 나타냄

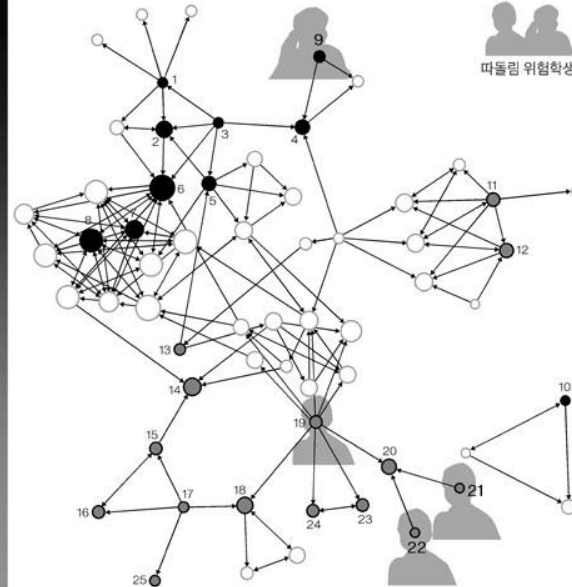


- 의존형: 아무와도 연결되어 있지 않고 오직 두 학생이 서로만 유일하게 선택
- 외면형: 쌍방으로 연결된 경우가 전혀 없는 경우
- 짝사랑형: 해당 학생을 선택한 학생은 아무도 없는 경우
- 외톨이형: 아무도 선택하지 않고 선택 받지 않은 경우

출처: (주)사이람 / 서울대사회발전 연구소 SNCC 공동연구 KBS

## 인천 OO초등학교 6학년 1반 25명 교우관계 보니

카톡 친구로 본 관계망  
 ※원이 클수록, 화살표가 많을수록 교우관계가 원만함  
 숫자는 조사학년 학생 25명의 일련번호



자료:사회연계망 분석업체 사이람

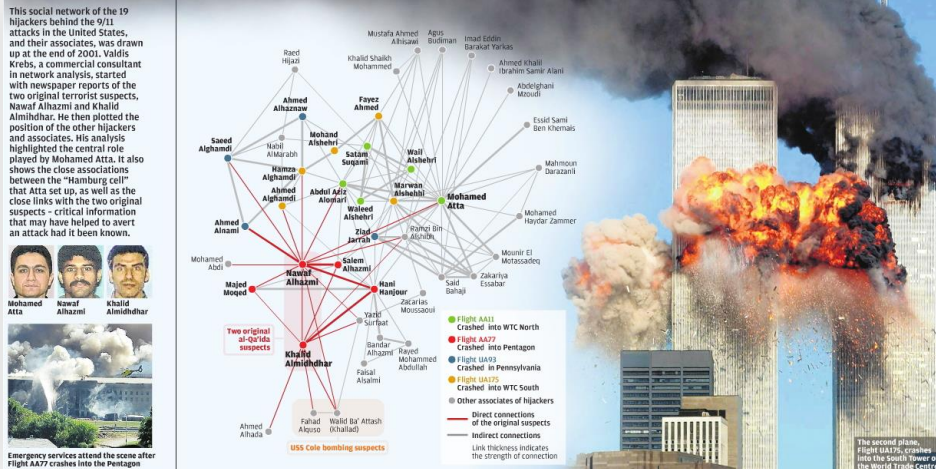
### 외톨이 학생 찾아내는 교우관계망 분석

- 예시 질문
- ▶반장 선거를 한다면 받고 싶은 친구는 누구인지 적어주세요
  - ▶수업시간에 조를 짜서 발표한다면 같은 조가 되고 싶은 친구를 적어주세요
  - ▶함께 앉고 싶은 친구를 순서대로 적어주세요
  - ▶시험 끝나고 함께 어울리고 싶은 친구는 누구인가요
  - ▶카카오톡 스마트폰 메신저로 자주 어울리는 친구는 누구인가요



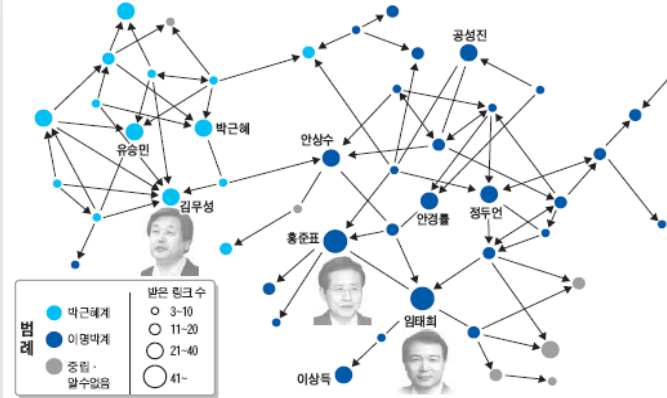
# 관계를 본다는 것? 숨겨진 맥락을 볼 수 있다는 것?

## How two names and a sheaf of newspaper cuttings revealed the 9/11



## 조선일보, 2008. 8 18대 의원 네트워크 조사

한나라당내 의원간 네트워크 지도

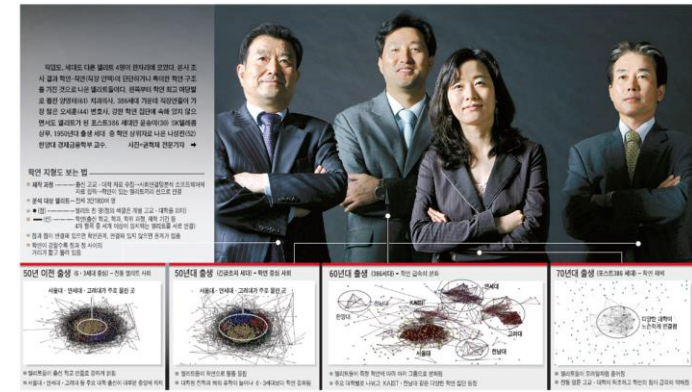


## 최고 '마당발'은 임태희·홍준표 의원

한국 사회 파워 엘리트 대해부

## 그들은 지금 해체되고 있다

중앙일보, 2005. 9



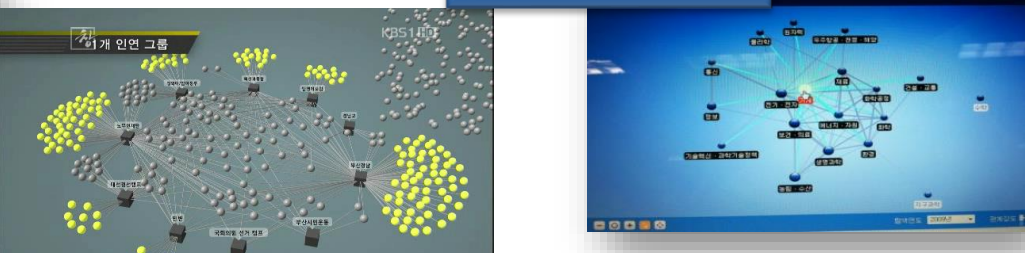
### 통신업종 지형도 '패밀리링' 드러내

7기업 파워 엘리트 관계망 집중 분석

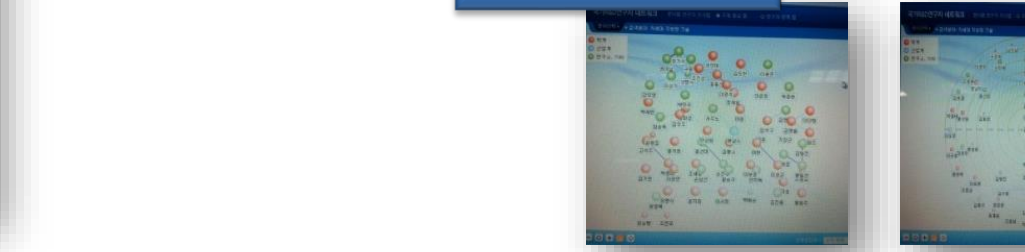
SW기업 CEO '나홀로링' 특징

IBM-HP 중심 다국적 기업들 국내 대기업보다 친밀도 높아

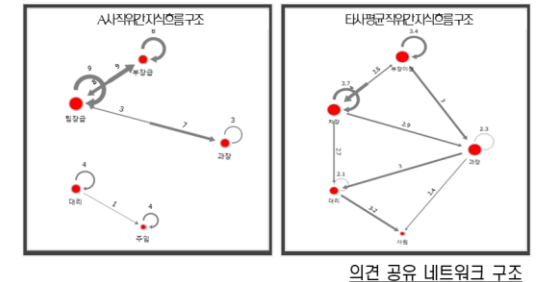
## 분야별 연구자 지식지도



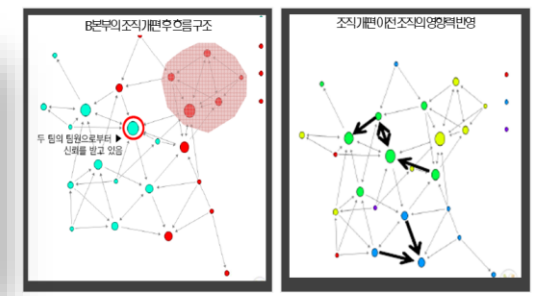
## R&D 분야간 과학지도



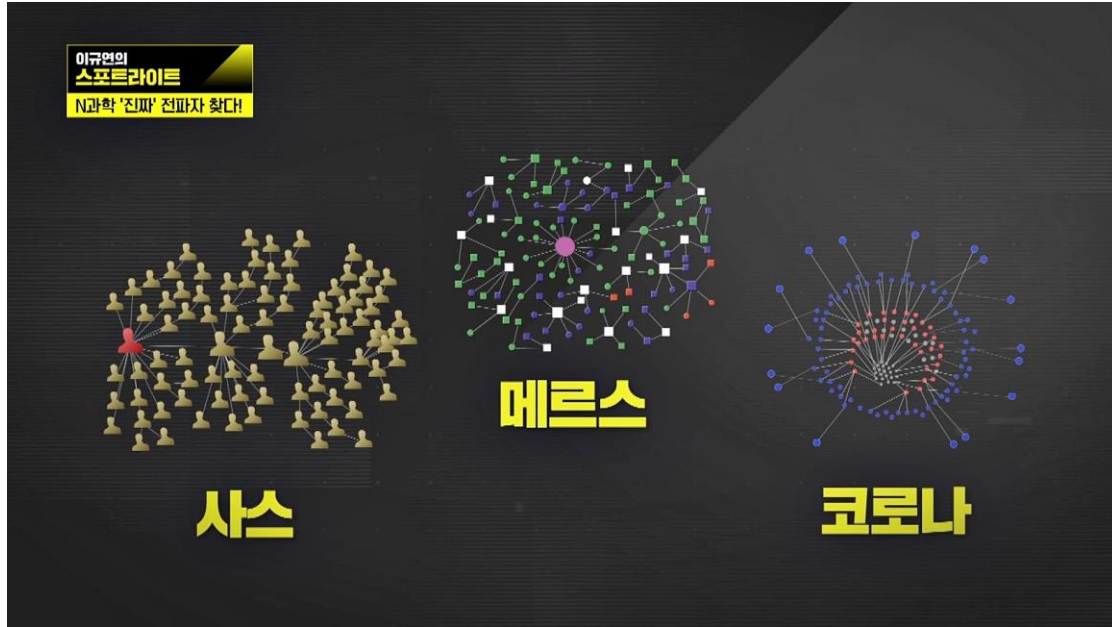
## 직위권 지식은 네트워크 구조



## 의견 공유 네트워크 구조



# 관계를 본다는 것? 관계의 경로를 볼 수 있다는 것?



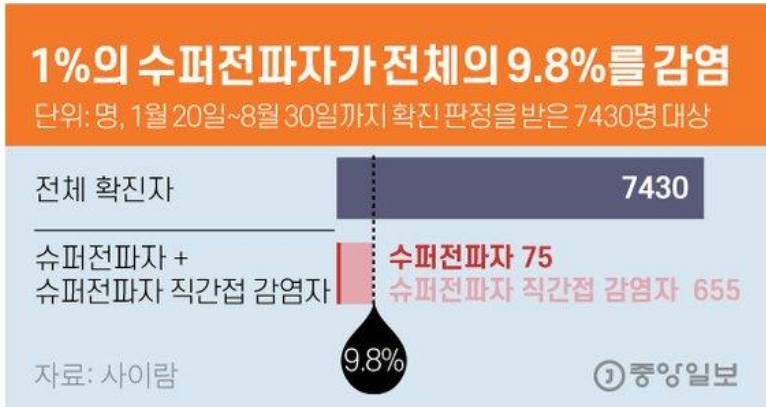
### 1차 확진자 어디서 많이 나왔나

단위: 명, ( )안은 %

※1월 20일~8월30일, 수도권 1차감염자 3071명 대상

고위험 시설 상위 5곳	
유흥주점(클럽, 룸살롱)	166 (5.4)
방문판매	141 (4.6)
뷔페	5 (0.2)
실내집단운동(GX)	3 (0.1)
노래연습장	2 (0.1)
중위험 시설 상위 5곳	
종교시설	1422 (46.3)
카페	57 (1.9)
실내체육시설	56 (1.8)
학원(300인 미만)	50 (1.6)
공연장	38 (1.2)
비위험 시설 상위 5곳	
집회	168 (5.5)
요양시설	109 (3.6)
금융	106 (3.5)
직장	104 (3.4)
병원	85 (2.8)

자료: 사이람      ① 중앙일보



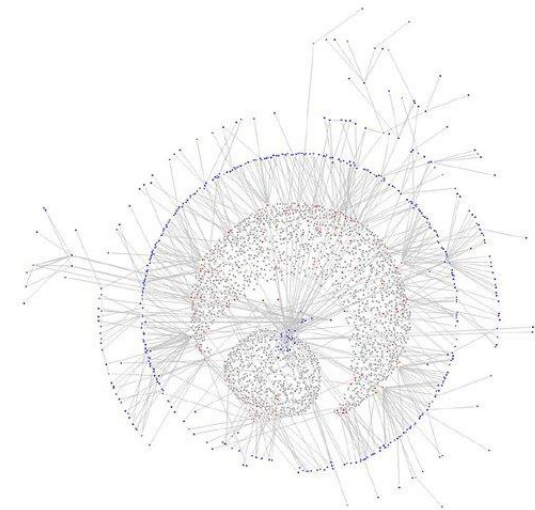
## 슈퍼전파자 1%가 n차감염자 59%에 영향 끼쳤다

중앙일보 | 입력 2020.10.14 00:02 | 업데이트 2020.10.14 01:31

자면보기 ①

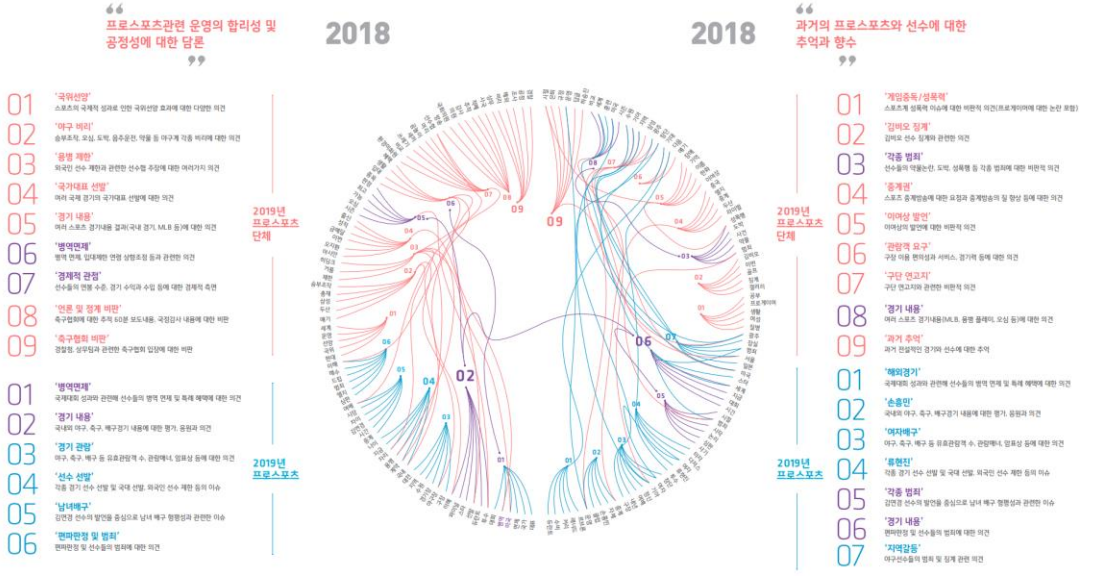
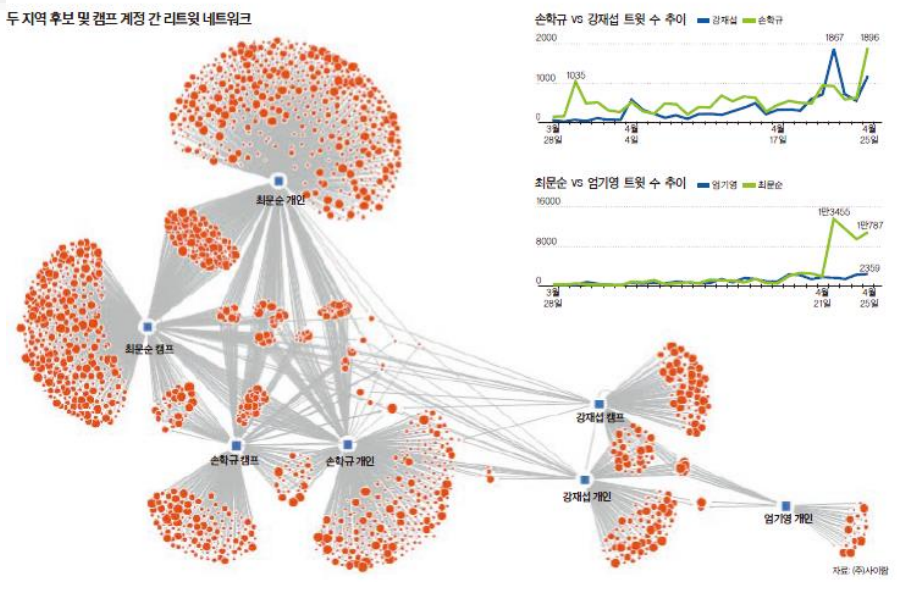
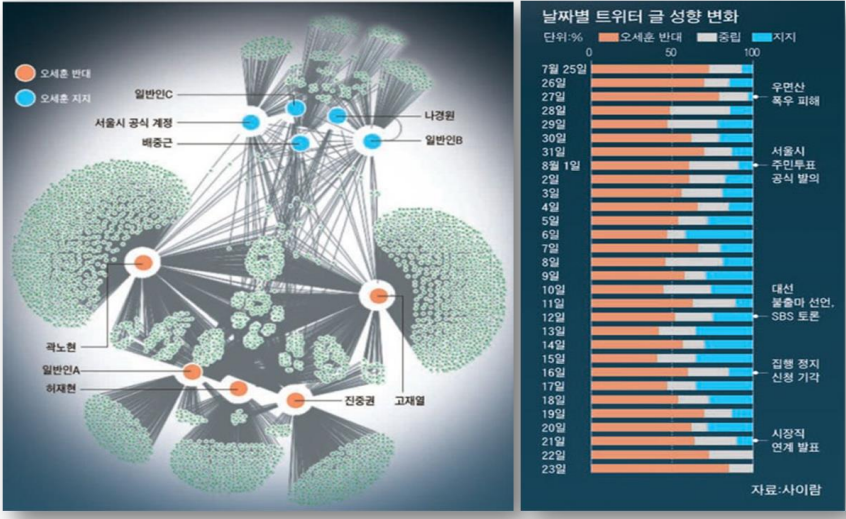
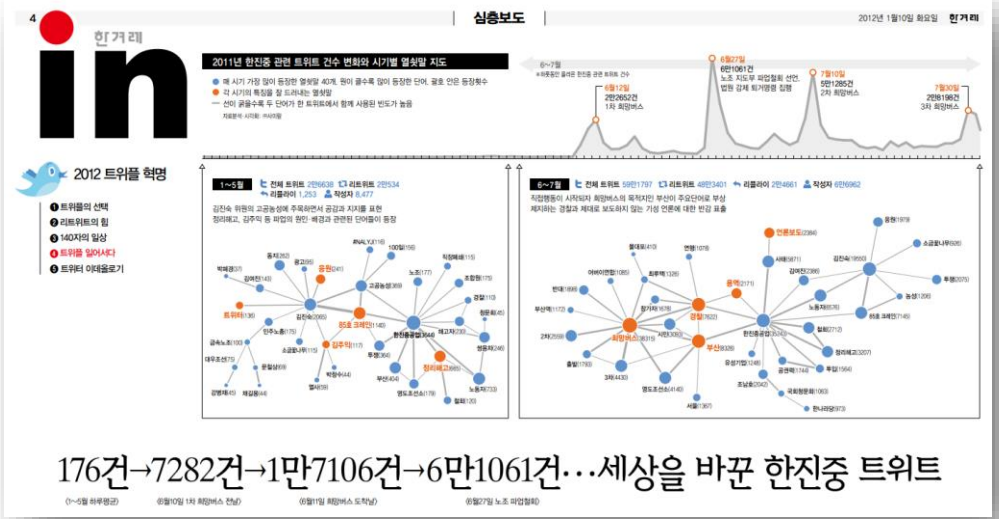
김태윤 기자 김경진 기자

바이러스는 '감염'으로 흔적을 남긴다. 감염의 경로를 추적하면 바이러스가 어떤 틈을 파고들어 확산하는지, 어떤 방역 대책이 필요한지 알 수 있다. 중앙일보는 빅데이터 분석업체인 사이람과 함께 지난 10개월간 대한민국을 강타한 신종 코로나바이러스 감염증(코로나19)의 감염 네트워크를 분석했다. 국내 첫 확진자가 발생한 지난 1월 20일부터 8월 30일까지 수도권에서 확진 판정을 받은 7430명이 분석 대상이다. 방역당국과 지방자치단체의 '확진자 접촉 추적 데이터'를 활용했다.





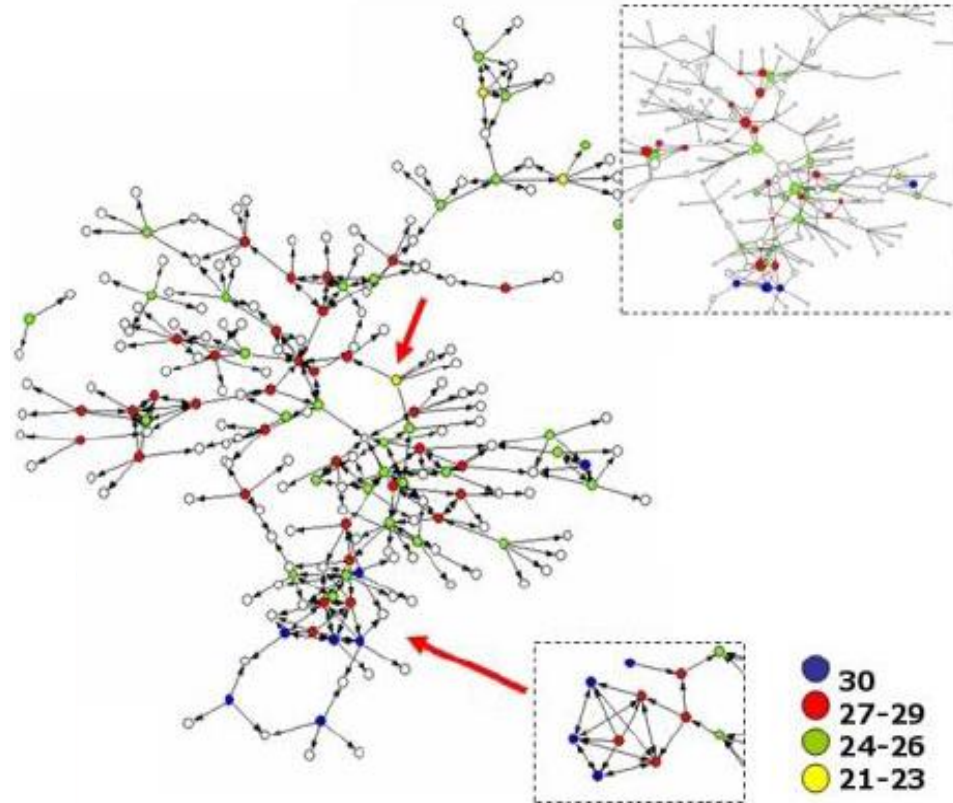
# 온라인상의 관계를 본다는 것? 여론의 흐름을 볼 수 있다는것?





## Social Network Analysis?

개인 및 집단들 간의 관계를 **노드**와 **링크**로서 모델링 하여  
그 위상구조, 확산/진화과정을 계량적으로 분석하는 방법론

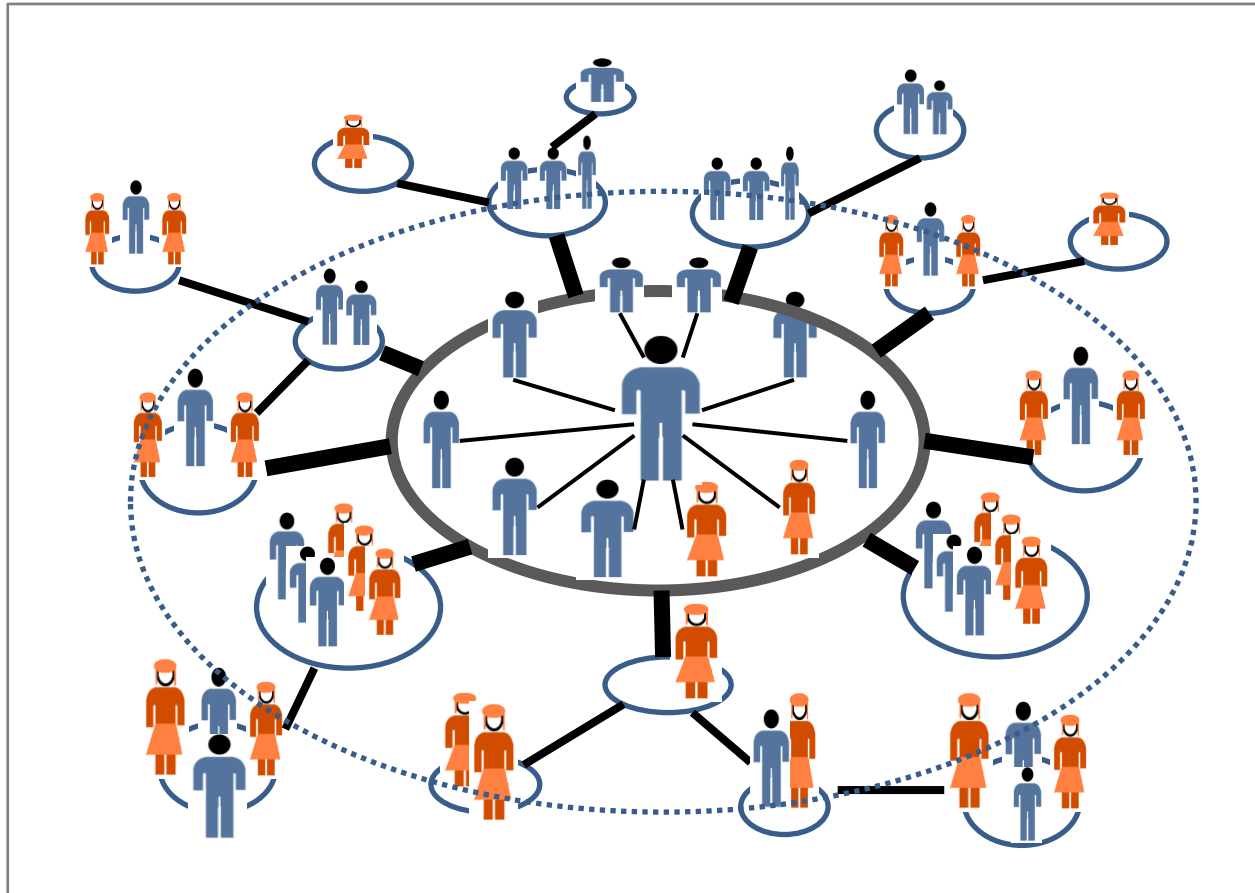


# 1. 소셜 네트워크 분석 개요



## Social Network Analysis?

사람들 사이에 존재하는 다양한 관계



혈연

결혼

지연

학연

커뮤니케이션

통화

금융거래

모임참여

...

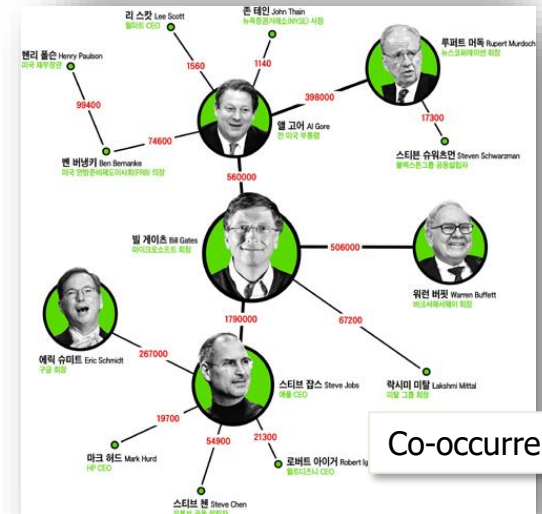
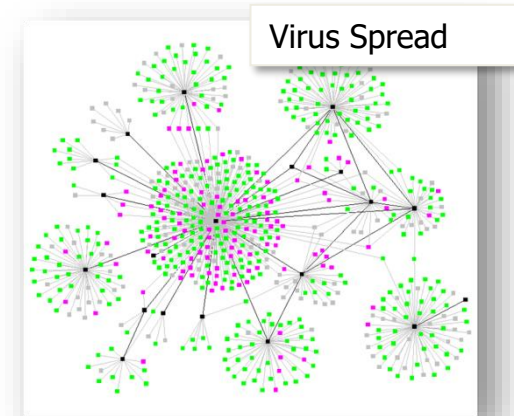
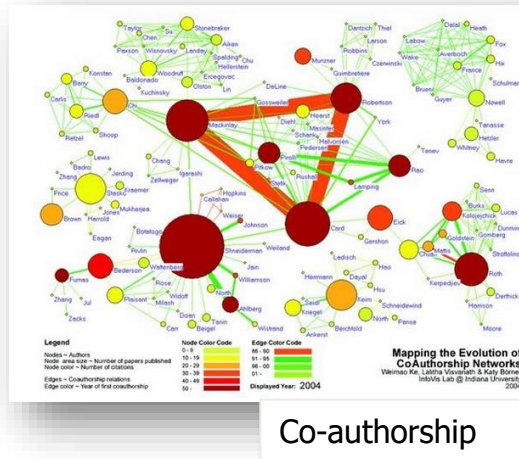
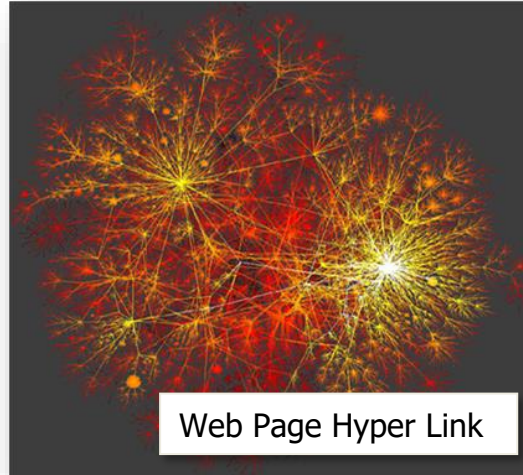
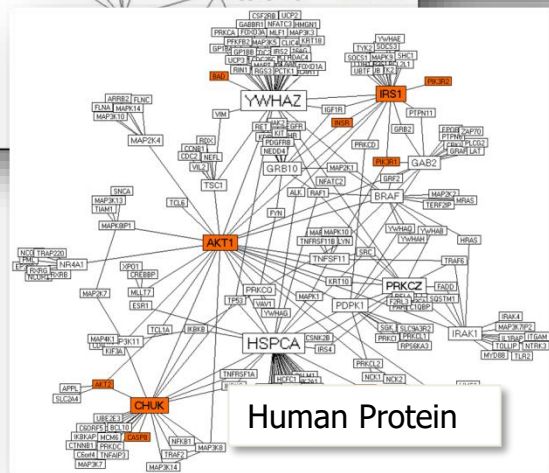
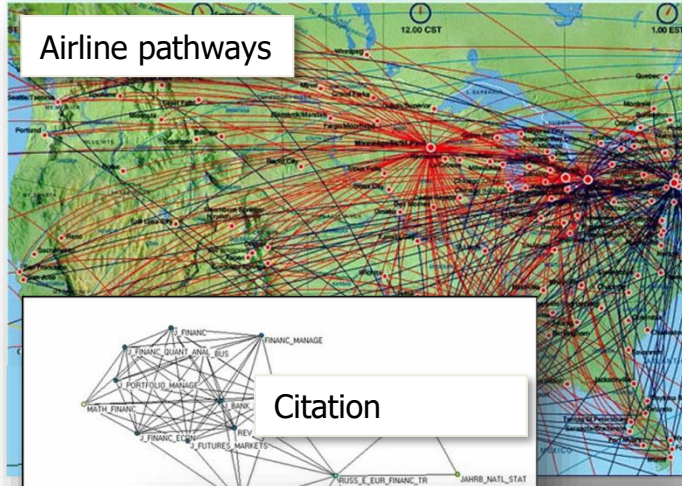


# 1. 소셜 네트워크 분석 개요



## Social Network Analysis?

지식정보/물품/상품 등 다양한 개체 간의 관계





# 1. 소셜 네트워크 분석 개요



## Social Network Analysis?

“Network is the core structure and dynamics mechanism of human, society and nature. ”

Hyper-Connectivity (초연결사회)

Sensor Network (IoT)

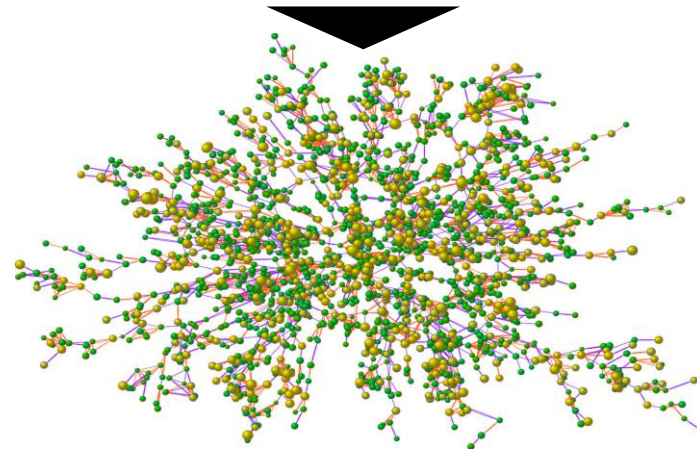
Social Network (Facebook, Kakao)

Financial Network

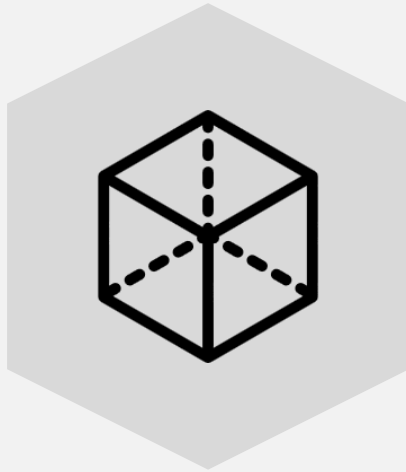
Knowledge Network

Brain Network

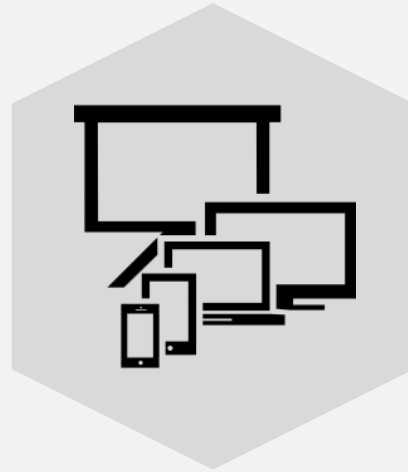
Criminal Network



# BIG DATA



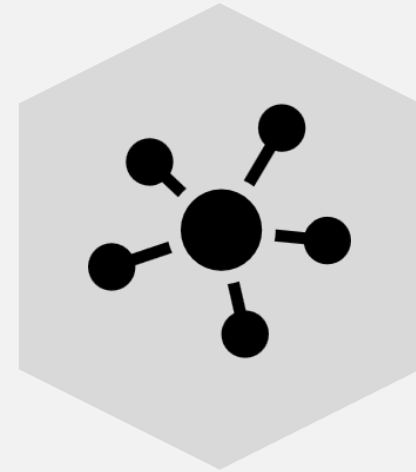
Volume



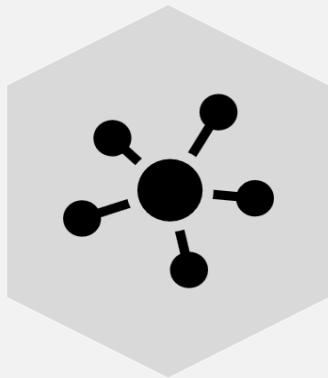
Variety



Velocity

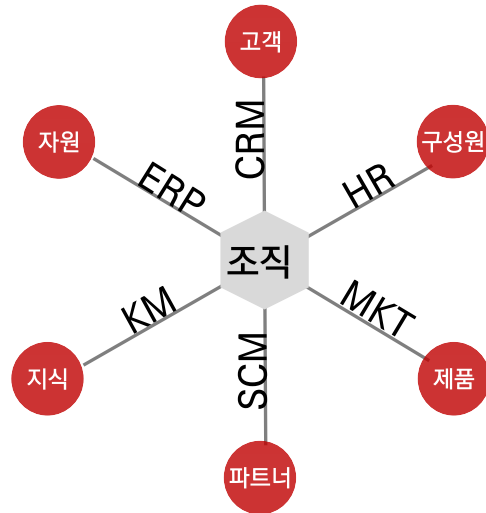


Complexity

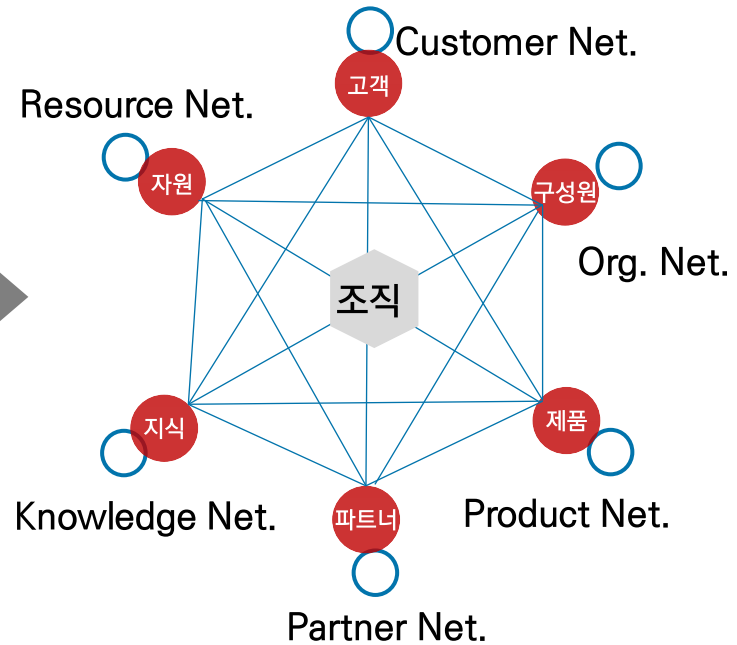


Complexity

### Traditional Management Model



### Network Management Model



기존의 조직중심적 관계모델에 조직의 핵심 실체들 간의  
관계모델 차원이 추가되는 데이터

# 어떤 것이 큰 데이터인가?

유저데이터

$N$

행위데이터

$N \times M$

관계데이터

$N^2$



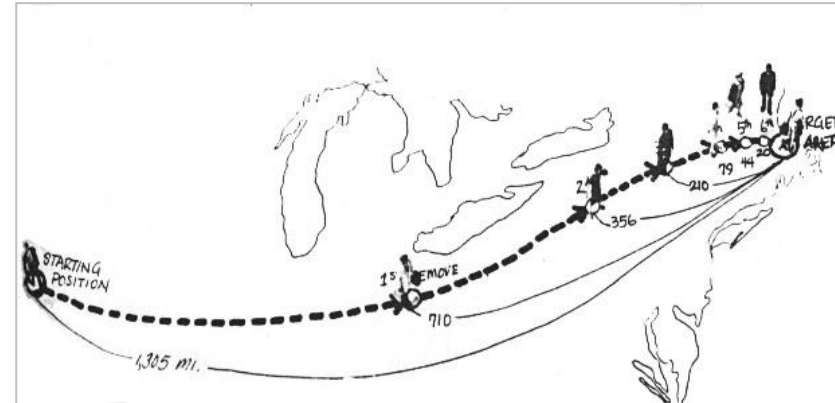


# 1. 소셜 네트워크 분석 개요



## How it happened? Small World

### Six Degree of Separation



- 1967년. 미국의 심리학자인 Stanley Milgram과 Jeffery Traverser
- Starting Point: [Omaha, Nebraska](#) and [Wichita, Kansas](#)
- End Point: [Boston, Massachusetts](#)
- Starting Point에 있는 사람을 무작위 선택하여, End Point에 있는 사람들의 정보를 주고, 그 사람을 알지 못할 경우, 자기가 아는 사람 중 그 사람을 알 것 같은 사람에게 우편물을 전달하도록 실험.
- 총 296개 우편물 중에서 64개가 도착. Average Path Length = 5.5 → 평균 6단계 정도.
- 1991년 John Gaure의 연극 “Six Degrees of Separation” 에서부터 이 용어가 구체적으로 사용되기 시작하였음.

# 1. 소셜 네트워크 분석 개요

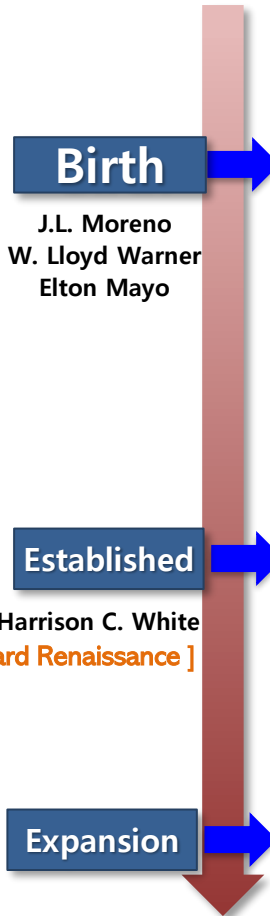
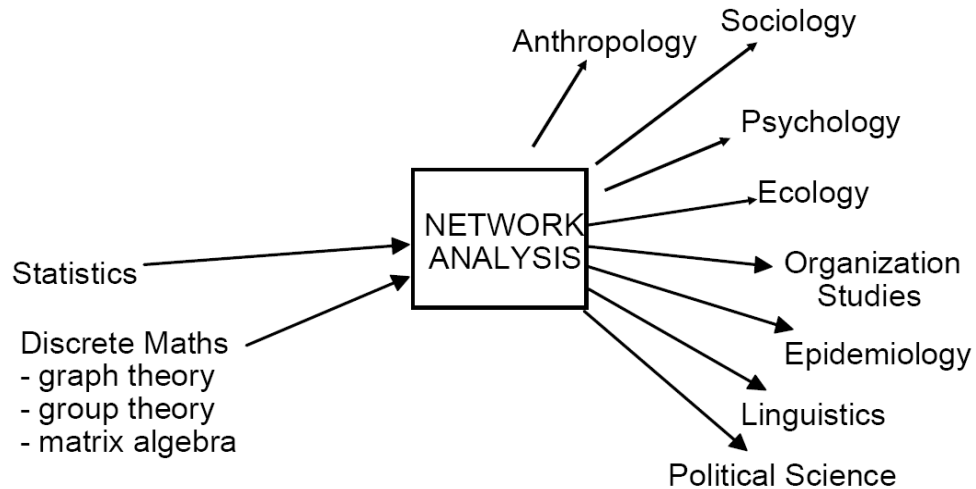
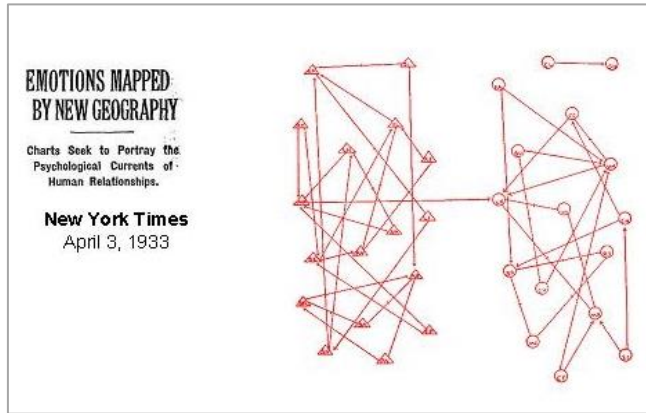


## Historical Backgrounds

SNA의 시작은 1930년대로 거슬러 올라가며, 70~80년대 부흥기를 거쳐 오늘날까지 급속하게 발전하고 있습니다.



(Jacob Moreno)

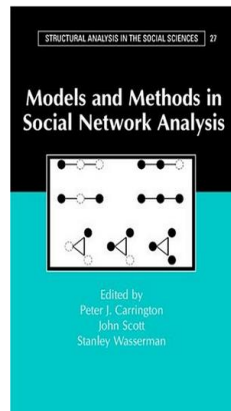
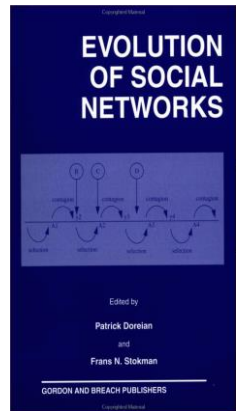
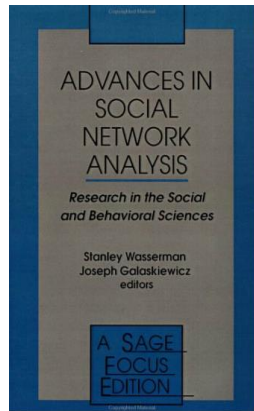
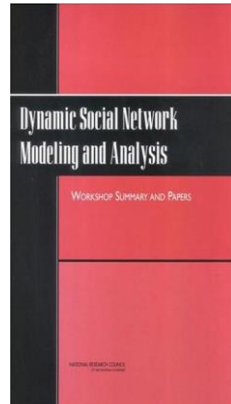
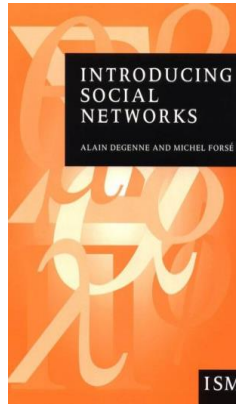
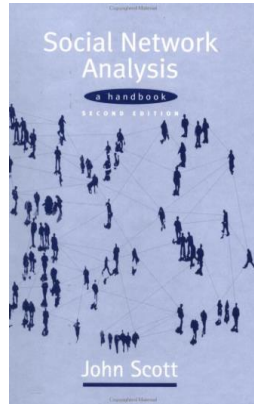


- 1736- Euler
- 1900s- Durkheim, Simmel
- 1930s- Sociometry
  - Hawthorne studies
- 1940s Psychologists
  - Clique formally defined
- 1950s Anthropologists
  - Kinship analysis
- 1960s Graph Theorists
- 1970s Rise of Sociologists
  - Small Worlds, Strength of weak ties
- 1980s IBM computation
  - Computer programs developed
- 1990s Ideas spread
  - UCINET released, spread of network analysis to multiple fields, social capital.
- 2000s "re-invented" by Physicists ("network Science")
  - Scale free, small world

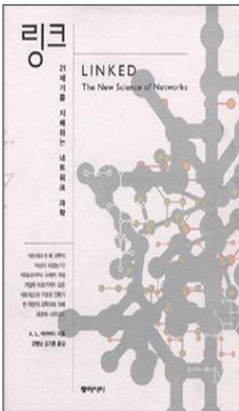
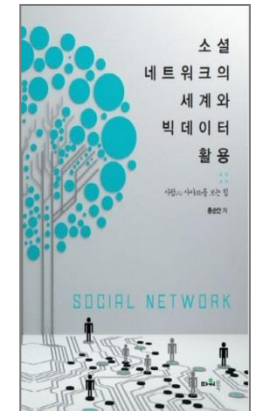
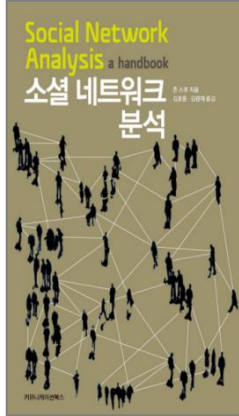
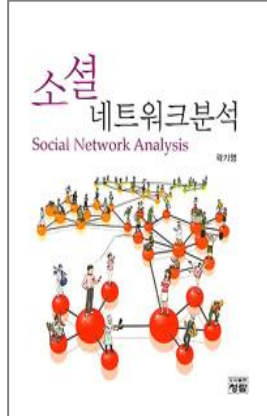
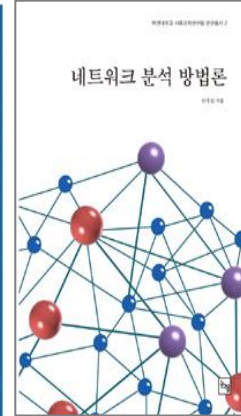
# 1. 소셜 네트워크 분석 개요



## Text Books(해외)



## Text Books(국내)





# 1. 소셜 네트워크 분석 개요



## Computer Programs

### Early Programs on particular SNA purpose

SOCAPAC I,	1971 by Leinhart
BLOCKER,	1971 by Heil & H. White, upgraded in 1976
SOCK,	1972 by Alba & Gutmann
COMPLT,	1972 by Gutmann
CATIJ,	1973 by Bernard & Killworth
CONCOR,	1975 by Breiger, Boorman & Arabie
STRUCTURE,	1976 by Ronald Burt
NEGOPY,	1975 by Richards
SONET,	1978 by Seidman & Foster
CENTER,	1979 by Freeman
GRADAP,	1981 by a group in Netherlands
COBLOC,	1981 by Carrington & Heil

These early programs vary widely. They were concentrated with groups, positions, centrality, kinship structure and distributions of structural properties

### General purpose SNA programs



**SONIS,** 1983 by Pappi & Kappelhoff

**UCINET,** 1983 by Borgatti

**STRUCTURE,** 1983 changed

**GRADAP,** 1983 changed

All four programs made an attempt to include the full range of network analytic procedure



\* Pajek 슬로베니아어로 "거미"란 뜻

**PAJEK,** 1996 by Vladimir Batagelj, Slovenia

\* 슬로베니아 (離散/전산) 수학자,



\* 서울대 경제학 전공. 사회학 박사(수료)

**Netminer,** 2001 by 김기훈 박사, Korea

By providing a standard approach to data analysis, these programs played a role in the development of the social network research community.

Batagelj - Professor of Discrete and Computational Mathematics, University of Ljubljana, Discrete Mathematics, Network Analysis, Algorithms, Optimization, Graph Theory, 대용량 네트워크 데이터의 수학적 처리 및 시각화 기능 구현자.

# 1. 소셜 네트워크 분석 개요



## NetMiner Reputation

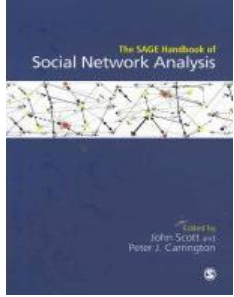


Table 38.4 Scores for some general packages

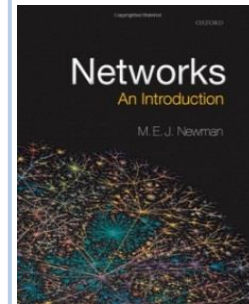
	Functionality						Support		User-friendly	
	Data	Vis.	Desc.	Proc.	Stat.	Dyn.	Doc.	Help		
MultiNet	+ -	+	+ -	+	-	0	+	++	+	
NetMiner 3	++	++	++	++	+ -	+ -	+	+	++	
ORA	++	++	+	++	+	+	+	+	+ -	
Pajek	+	++	+	++	0	+ -	+	0	+ -	
statnet/sna	++	+	++	++	++	+	++	+	+ -	
UCINET + NetDraw	++	++	++	++	+	0	++	+	+	

Mark Huisman and van Duijn (2011).

**The Sage Handbook of Social Network Analysis.**

Ch38. A Reader's Guide to SNA Software p.594 Scores for some general packages

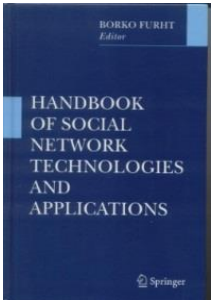
John P. Scott & Peter J. Carrington(Eds.) SAGE Publications Ltd.



Name	Availability	Platform	Description
Pajek	Free	W	Interactive social network analysis and visualization
Net Workbench	Free	WML	Interactive network analysis and visualization
Netminer	Commercial	W	Interactive social network analysis and visualization
InFlow	Commercial	W	Interactive social network analysis and visualization
UCINET	Commercial	W	Interactive social network analysis
yEd	Free	WML	Interactive visualization
Visone	Free	WL	Interactive visualization
Graphviz	Free	L	Visualization
NetworkX	Free	WML	Interactive network analysis and Python library
JUNG	Free	WML	JAVA library for network analysis and visualization
igraph	Free	WML	C/R/Python libraries for network analysis
GTL	Free	WML	C++ library for network analysis
LEDA/AGD	Commercial	WL	C++ library for network analysis

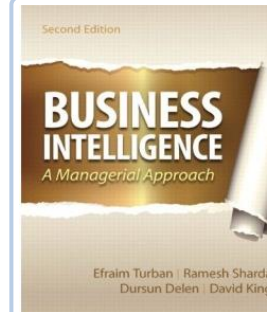
Mark Newman (2010). **Networks: an introduction** p.277 Table 9.1: A selection of software implementing common network algorithms. Oxford University Press.

### 1.5.3 SNA Data Processing Tools



NETMINER is an innovative software tool for Exploratory Analysis and Visualization of Network Data. It can be used for general research and teaching in social networks. This tool allows researchers to explore their network data visually and interactively, helps them to detect underlying patterns and structures of the network. Especially, it can be effectively applied to various business fields where network-structural factors have great deal of influences on the performance (e.g., intra and inter-organizational financial Web criminal/intelligence informetric telecommunication distribution transportation networks). Statistically, this program supports many standardized computer methods, including descriptive statistics, ANOVA, correlation, and regression.

Borko Furht (2010). **Handbook of Social Network Technologies and Applications** p.19 SNA Data Processing Tools. Springer Press.



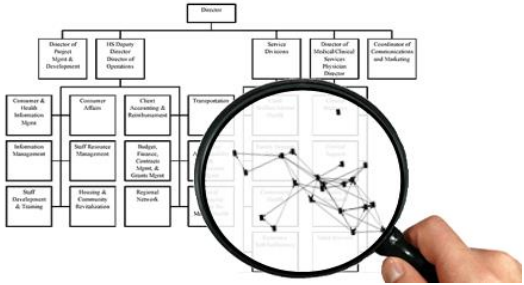
“SNA software tools include Business-oriented social network tools such as InFlow and NetMiner.”

Efraim Turban, Ramesh Sharda, Dursun Delen, David King, Janine E. Aronson (2011). **Business Intelligence: A Managerial Approach(2nd Edition)** p.250 6.8 ONLINE SOCIAL NETWORKING: BASICS AND EXAMPLES. Prentice Hall Press

# 1. 소셜 네트워크 분석 개요



## 통계분석 VS SNA



	통계 접근	네트워크 접근
기본 시각	환원주의(원자론)	전체주의(구조론)
분석 대상 데이터	개체의 속성 데이터	개체간 관계 데이터
데이터 수집방법	(샘플링) 표본집단	(basically) 모집단
핵심 분석방법	속성간 상관관계 (속성의 유사성)	구조적 위치 (관계패턴의 유사성)
분석 목적	경향성 파악	구체적 지점 확인
적용 사례	감염자 수/병 증상	감염경로/전파력



## 2. 데이터 모델링



### 데이터 모델링

Network = Node set + Link set

Node set



Link set



Network



## 2. 데이터 모델링



### 데이터 모델링

#### 1) Node Data

		1	2	3	4	5	6
		Education	Gender	Job-ranking	Department	Duration	Age
1	John	"Master Degree"	"Male"	1.0	"Finance"	21.0	45.0
2	Thomas	"Bachelor Degree"	"Male"	5.0	"Marketing"	8.0	35.0
3	Anna	"HS Graduate"	"Female"	7.0	"Marketing"	0.5	27.0
4	James	"Bachelor Degree"	"Male"	5.0	"Marketing"	9.0	37.0
5	Peter	"Bachelor Degree"	"Male"	7.0	"Sales"	3.0	28.0
6	Mary	"Bachelor Degree"	"Female"	7.0	"Sales"	2.0	26.0
7	Michael	"Bachelor Degree"	"Male"	6.0	"Marketing"	5.0	29.0
8	David	"Bachelor Degree"	"Male"	3.0	"Marketing"	13.0	36.0
9	Anthony	"Bachelor Degree"	"Male"	4.0	"Marketing"	13.0	35.0
10	Bobby	"Bachelor Degree"	"Male"	7.0	"Sales"	2.0	25.0
11	Robert	"Doctoral Degree"	"Male"	2.0	"Marketing"	15.0	41.0
12	Susan	"Master Degree"	"Female"	5.0	"Marketing"	8.0	34.0
13	Steven	"Bachelor Degree"	"Male"	4.0	"Sales"	12.0	38.0
14	Charles	"Doctoral Degree"	"Male"	5.0	"Marketing"	10.0	33.0
15	Ashley	"Master Degree"	"Female"	7.0	"Marketing"	4.0	30.0
16	Richard	"Bachelor Degree"	"Male"	7.0	"Marketing"	1.0	30.0
17	Jessica	"Doctoral Degree"	"Female"	2.0	"Sales"	16.0	42.0
18	Elizabeth	"2YR College"	"Female"	3.0	"Marketing"	13.0	43.0
19	Laura	"HS Graduate"	"Female"	7.0	"Sales"	3.0	24.0
20	Jennifer	"Master Degree"	"Female"	6.0	"Sales"	5.0	31.0
21	Jackson	"Bachelor Degree"	"Male"	6.0	"Sales"	6.0	28.0
22	Julia	"Bachelor Degree"	"Female"	6.0	"Sales"	7.0	33.0



#### 노드 고유식별이 가능해야 함(ID)

- ID는 Text와 숫자 모두 가능
- 단, "사람 이름"이나 "키워드" 를 ID로 활용할 경우 동명이인이나 동음이의어가 없는지 확인해야 함



#### 노드의 속성정보를 포함할 수 있음

- 분석 목적에 따라 다양하게 구성할 수 있음
- 성별, 지역, 연령, 소속조직, 만족도, 찬/반 태도 등

## 2. 데이터 모델링



### 데이터 모델링

#### 2) Link Data

##### Direction

###### Directed

ex) a가 b에게 소문을 전달하다.



###### Undirected

ex) a와 b는 친구이다.



##### Weight

###### Weighted

ex) a와 b가 하루에 4번 통화하다.



###### Unweighted

ex) 웹페이지 a와 b가 연결되어 있다.



#### ✓ 링크는 방향성을 가질 수 있음

- 방향성이 있는 데이터는 Source Node와 Target Node를 구분함
- 필요에 따라 방향성이 있는 데이터를 방향성이 없는 데이터로 변환할 수 있음 (Symmetrize)

#### ✓ 링크는 가중치(값)를 가질 수 있음

- 가중치가 없는 데이터는 1과 0으로만 구분함
- 필요에 따라 가중치가 있는 데이터를 가중치가 없는 데이터로 변환할 수 있음 (Dichotomize)

## 2. 데이터 모델링



### 데이터 모델링

#### 2) Link Data

##### Link Attribute

ex) a가 b와 미팅을 한다.



- 미팅 목적
- 미팅 소요시간
- 미팅에 대한 만족도(5점 만점)



##### 방향성이나 가중치 외 속성값을 가질 수 있음

- 링크의 속성 데이터를 구성할 수 있음  
하지만 반드시 구성할 필요는 없으며 없는 경우도 많음
- 텍스트와 숫자 모두 가능
- 관계의 목적, 소요시간, 강도 등
- 분석 과정에서 특정 속성을 가진 링크만 추출할 때 활용할 수 있음

※ 어떤 것이 가중치이고 어떤 것이 속성이 될지는 연구자의 선택에 따라 달라짐



# 1. 소셜 네트워크 분석 개요



## Link Data

### “링크 데이터”란?

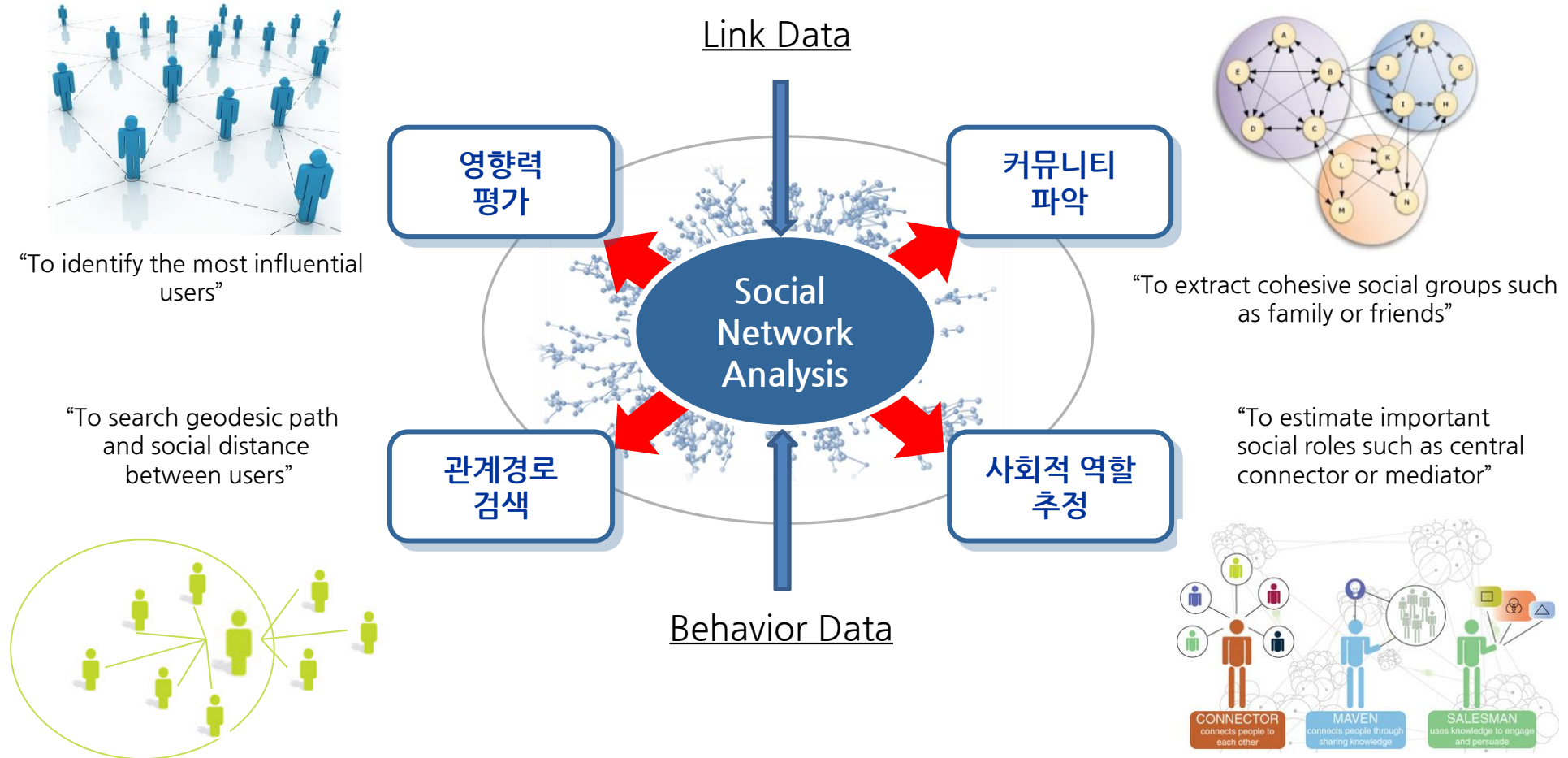
관계의 유형	예시	데이터 소스
Social Relations	(친분관계) A와 B는 서로 알고 지내는 사이이다.	설문조사
	(가족관계) A와 B는 결혼해서 C를 낳았다.	주민등록등본
	(온라인 관계) A와 B는 페이스북/트위터 친구이다.	친구 목록
Interaction	(커뮤니케이션) A는 B에게 전화/이메일/카톡을 보냈다	통화/이메일/카톡 내역
	(거래) A회사는 B회사에게 물품을 팔았다	세금계산서 발행내역
	(협업) 저자 A와 저자 B는 함께 논문을 작성했다	공저 목록
Flows	(자금의 흐름) A는 B에게 돈을 부쳤다.	통장거래내역
	(정보의 흐름) A는 B의 논문을 인용했다.	인용논문 목록
	(자원의 흐름) A 부서에서 일하던 김씨는 B 부서로 이동했다.	인사이동 경로
	(인구이동) A 도시에서 B 도시로 이사 갔다.	전출입 신고내역
	(물류) A 도시에서 B 도시로 택배를 발송했다.	배송차량 이동경로
Similarities	(속성 유사성) A와 B는 모두 박보검과 서강준을 좋아한다.	페이스북 “좋아요” 로그
	(이벤트 참여) A와 B는 모두 같은 컨퍼런스와 행사에 참여했다.	이벤트 참여 목록
	(공동구매) A와 B는 모두 기저귀와 맥주를 함께 구매했다.	구매내역
	(공동출현) 키워드 A와 B는 154건의 논문에 함께 등장했다.	논문 본문

# 1. 소셜 네트워크 분석 개요



## Core Value of SNA

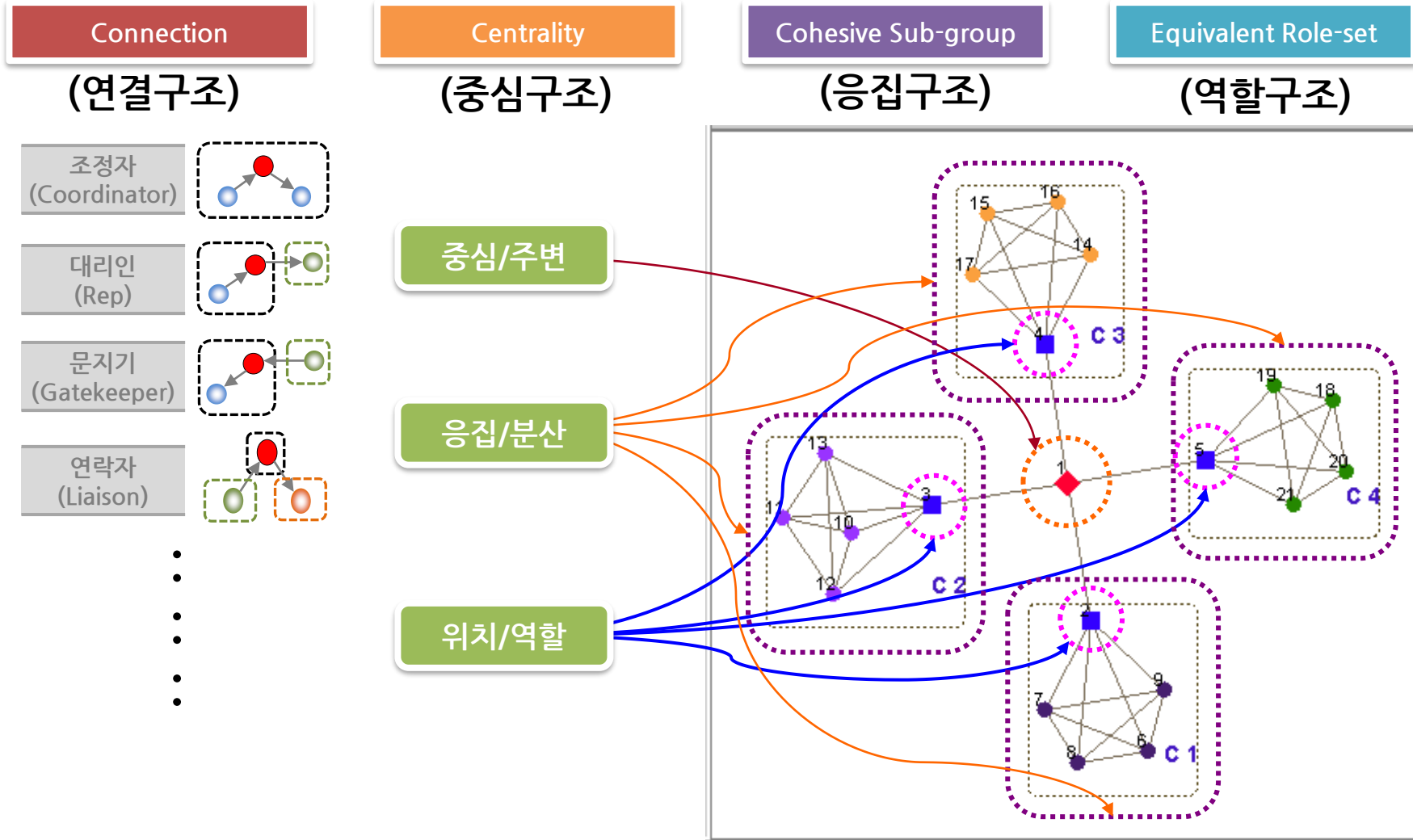
From **Big** to **Micro Target!**



# 1. 소셜 네트워크 분석 개요



## 네트워크 분석의 주요 주제



## 2. 데이터 모델링



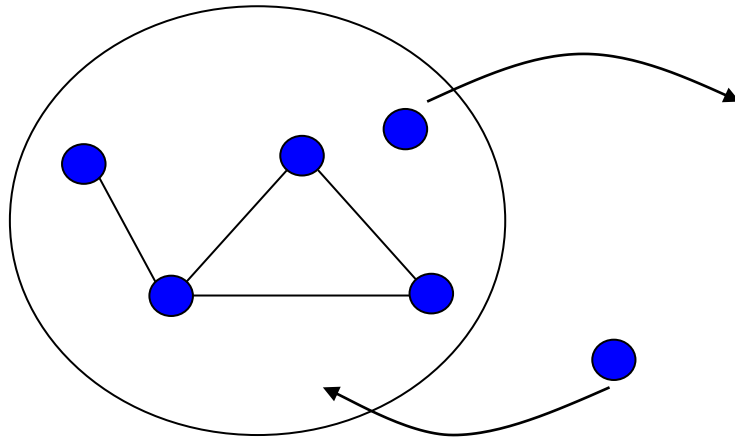
### 데이터 수집 방법

#### 1) 노드 데이터 수집



네트워크 데이터의 특성상 전수조사를 권장하며, 불가피할 경우 하위조직 단위의 표본조사 진행

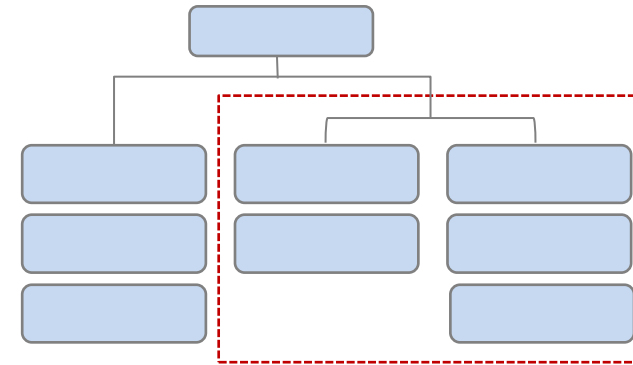
##### 전수조사



##### ✓ 노드 범위 제한

- A 학교 1학년 학생
- 2000년~2010년에 게재된 논문

##### 표본조사



##### ✓ 하위조직 단위 선정

- “A학교 1학년” 대신 “1학년 2반”으로 축소
- 2000년~2010년에 게재된 논문 중 A 카테고리



## 2. 데이터 모델링



### 데이터 수집 방법

#### 2) 링크 데이터 수집

##### ① 설문조사

- 구성원 또는 조직 간 관계를 조사할 경우에 주로 활용
- NGQ(Name Generator Questions): 응답자가 문항에 적합한 대상을 직접 지목하는 방식의 설문조사임
- 대부분 실명조사로 진행됨

##### 예시

##### II. 주민 관계망 조사

1. 현재 마을계획단 구성원 중, 우리 등과 관련해 유용한 정보(행정서비스, 신청방법, 각종 행사 등)를 자주 제공해주는 사람은 누구입니까? 최대 10명까지 기재할 수 있습니다.

	이름
①	
②	
③	
④	
⑤	



##### 조사하고자 하는 관계의 의미를 구체적으로 제시

- 함께 점심을 먹거나, 쉬는 시간에 즐거운 시간을 보내는 친구는 누구인가요?
- 함께 영화를 보거나 놀이공원에 놀러 가고 싶은 친구는 누구인가요?

## 2. 데이터 모델링



### 데이터 수집 방법

#### 2) 링크 데이터 수집

##### ① 설문조사

수집해야 할 관계 속성 및 응답환경 고려

속성	구분	예시
지목방법	목록제시형	선택할 대상의 목록을 제시 Q. 함께 영화를 보거나 놀이공원에 놀러 가고 싶은 친구는 누구인가요? 아래 목록에서 선택해주세요.
	자유회상형	선택할 대상을 주관식으로 입력 Q. 함께 영화를 보거나 놀이공원에 놀러 가고 싶은 친구는 누구인가요? 떠오르는 대로 적어주세요.
지목대상 제한여부	선택제한형	선택할 대상의 최소/최대 수를 제한 Q. 함께 영화를 보거나 놀이공원에 놀러 가고 싶은 친구는 누구인가요? 최소 1명 이상 적어주세요. Q. 함께 영화를 보거나 놀이공원에 놀러 가고 싶은 친구는 누구인가요? 최대 7명까지 적어주세요.
	선택개방형	선택할 대상의 수를 제한하지 않음 Q. 함께 영화를 보거나 놀이공원에 놀러 가고 싶은 친구는 누구인가요? 떠오르는 대로 적어주세요.
링크 가중치 유무	링크 강도	선택할 대상 및 링크 강도 정보를 함께 수집 Q. 함께 영화를 보거나 놀이공원에 놀러 가고 싶은 친구는 누구인가요? 가고 싶은 정도를 함께 선택해주세요 ① 매우 가고 싶음 ② 약간 가고 싶음 ③ 보통
	순위	선택할 대상의 링크 강도를 입력 순서로 수집 Q. 함께 영화를 보거나 놀이공원에 놀러 가고 싶은 친구는 누구인가요? 제일 친한 순서대로 적어주세요.

# 2. 데이터 모델링



## 데이터 수집 방법

### 2) 링크 데이터 수집

#### ② 기존 데이터

- 기존에 축적된 로그 데이터, 문헌 데이터 등에서 링크 데이터를 추출할 수 있음

#### 예시

논문ID	SSN/ISBN	발행기관명	논문/챕터	부제목	저자명	발행연도	조목	키워드
1	1225-0120	한국사회학회	소셜네트워크	페이스북	사:윤명희	2013	이 연구	상호작용 의례,복합성,페이스북,자기전시,친밀성,ir
2	1229-2435	한국비블리오	우리나라	출:트위터, 페이	이종문	2011	본 연구	소셜 네트워크,출판사 SNS,트위터,페이스북,미투데이
3	598-5733	사이버커뮤니	트위터와	페이스북	이용자	2012	최근 학	트위터,페이스북,SNS,정치참여,결속적 사회자본,교
4	598-5733	사이버커뮤니	소셜미디어	이용자의 자기	이상숙, 이	2013	본 연구	관계 지향,소셜 미디어,정보 지향,컨조인트 분석,M
5	598-0170	한국인터넷	소셜네트워크	트위터, 페이	홍삼열, 오	2012	2009년	소셜네트워크서비스,,트위터,페이스북,카카오토
6	598-2653	한국지역언	소셜네트워크	싸이월드, 페	김유정	2013	싸이월	소셜네트워크서비스,SNS,이용과 충족이론,싸이월
7	598-6284	한국데이타	온라인	소셜	싸이월드, 페	서보밀	2013	As the
8	1229-134X	한국지역정	지방자치단체	소셜미디어	(박근수	2013	이 연구	지방자치단체,소셜미디어,비인자형 집단지성,SNS,
9	598-3757	한국언론정	소셜네트워크	'페이스북', '김	위근, 최	2012	SNS 이	소셜네트워크서비스(SNS),이용동기,페이스북,트위
10	1738-6195	서울대학교	'소셜미디어	'갈릴레오 모	'반현, 이현	2011	소셜미	소셜미디어,트위터,페이스북,스마트폰,멀티 이용자

#### - 논문 및 키워드 목록 -

논문ID	키워드
1	상호작용 의례,복합성,페이스북,자기전시,친밀성,interaction ritual,multiplicity,faceb
2	소셜 네트워크,출판사 SNS,트위터,페이스북,미투데이,Social Network,Publishing Cc
3	트위터,페이스북,SNS,정치참여,결속적 사회자본,교향적 사회자본,Twitter,Facebook
4	관계 지향,소셜 미디어,정보 지향,컨조인트 분석,Morae,Conjoint Analysis,Informati
5	소셜네트워크서비스,,트위터,페이스북,카카오토,결속,비교분석,SNS
6	소셜네트워크서비스,SNS,이용과 충족이론,싸이월드, 페이스북,트위터,Social Netwo
7	Social Networking Service(SNS),Purpose of Use,Cyworld,Facebook,Twitter
8	지방자치단체, 소셜미디어, 비인자형 집단지성,SNS,CNS(Car Navigation System),Bas
9	소셜네트워크서비스(SNS),이용동기,페이스북,트위터,싸이월드,미투데이,Social Net
10	소셜미디어,트위터,페이스북,스마트폰,멀티 이용자,갈릴레오 모델,인지적 매핑,사회
11	조직공중 관계성,미디어몰입,커뮤니케이션 특성 (관계적 차원,내용적 차원),소셜미
12	소셜네트워크 서비스,페이스북,트위터,사회자본,허위문화,Social Network Service,F
13	페이스북,페이스북 이용 동기,사회적 지지 인식,자의식,FacebookFacebook use m
14	SNS,네트워크구조,싸이월드,페이스북,네트워크 개방성,Web2.0,SNS,Cyworld,Faceb

#### - Linked List -

No	키워드1	키워드2	키워드3	키워드4	키워드5
1	상호작용의례	복합성	페이스북	자기전시	친밀성
2	소셜네트워크	출판사SNS	트위터	페이스북	미투데이
3	트위터	페이스북	SNS	정치참여	결속적사회자본
4	관계지향	소셜미디어	정보지향	컨조인트분석	Morae
5	SNS	트위터	페이스북	카카오토리	결속
6	SNS	SNS	이용과충족이론	싸이월드	페이스북
7	SNS	PurposeofUse	싸이월드	페이스북	트위터
8	지방자치단체	소셜미디어	비인자형집단지성	SNS	CNS(CarNavigationS
9	SNS	이용동기	페이스북	트위터	싸이월드
10	소셜미디어	트위터	페이스북	스마트폰	멀티이용자
11	조직공중관계성	미디어몰입	커뮤니케이션특성	소셜미디어	
12	SNS	페이스북	트위터	사회자본	허위문화
13	페이스북	페이스북이용동기	사회적지지인식	자의식	
14	SNS	네트워크구조	싸이월드	페이스북	네트워크개방성
15	페이스북	프라이버시	개인정보자기결정권	SNS	
16	SNS	페이스북	감성	감정상태	감정전이가설
17	페이스북페이지	FCB그리드모델	메시지전략	크리에이티브전략	MPR



## 2. 데이터 모델링



### 데이터 수집 방법

#### 2) 링크 데이터 수집

##### ③ Web Crawling

- 트위터나 페이스북 등 SNS 내 구성원간 커뮤니케이션 관계를 분석하거나, 웹 페이지간 연결관계를 분석할 때 활용

예시

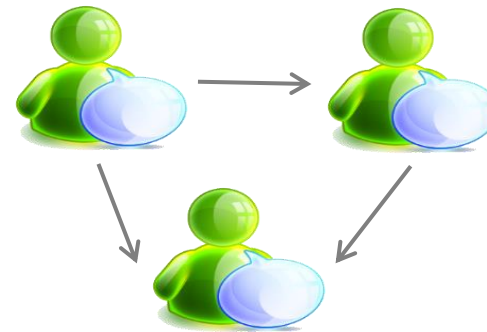
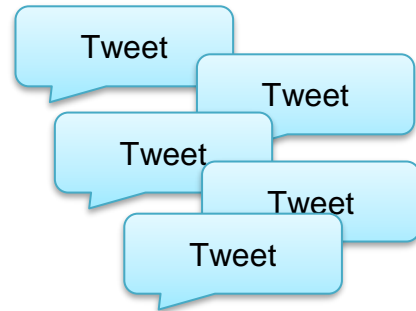


- 수집 키워드 및 기간 설정

- 해당 기간 내 해당 키워드를 포함한 트윗글 수집

- 수집된 트윗글 ID를 기반으로 글을 작성한 이용자 간의 follow/RT/reply 등 관계 데이터 수집

- 분석목적에 적합하지 않은 홍보나 스팸 목적의 글이나 이용자 등을 정제



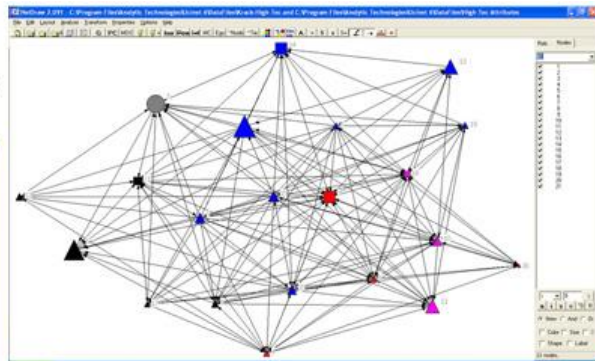


# 3. 분석을 위한 툴



**UCINET Software**

	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0



**TEXTOM**

키워드	수집날짜	수집채널	수집기간	기타
키워드	2017-03-13	페이스북	2017-02-13 ~ 2017-03-13	기본
키워드	2017-03-13	페이스북	2017-03-06 ~ 2017-03-13	기본
키워드	2017-03-10	네이버(블로그, 카페)다음 (블로그, 카페)페이스북트위터	2016-03-10 ~ 2017-03-10	기본
키워드	2017-03-10	네이버(블로그, 카페)다음 (웹, 블로그, 뉴스, 카페, 지식인, 이미지, 동영상) 구글+, 뉴스, 이미지, 동영상, 페이스북	2016-03-10 ~ 2017-03-10	기본

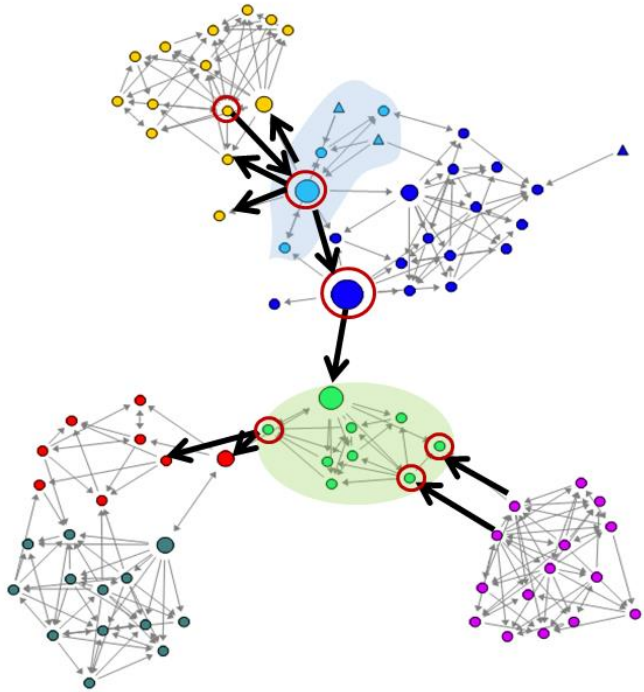
채널	섹션	데이터량	수집량	수집데이터
NAVER	블로그	253	259	VIEW
	뉴스	10	10	VIEW
	카페	213	204	VIEW
	지식인	10	10	VIEW



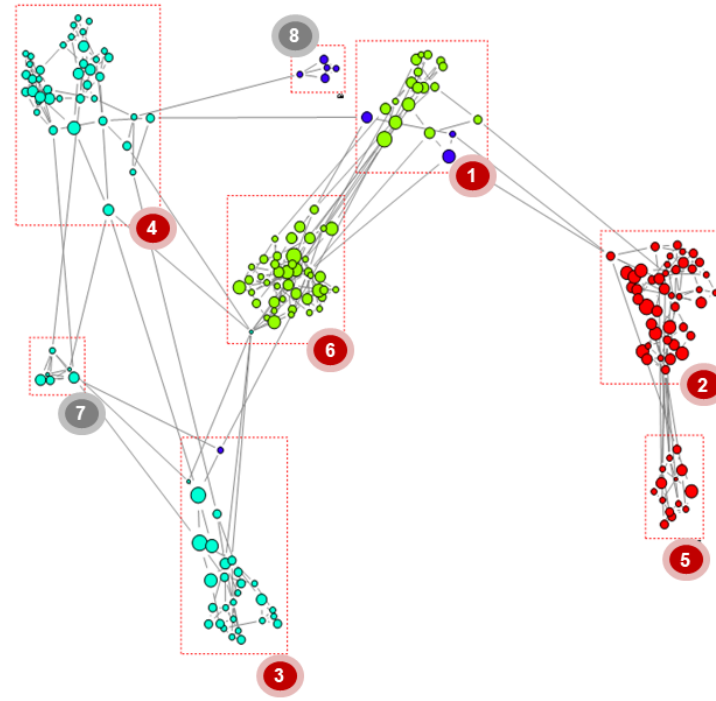
# 분석사례 - 조직네트워크분석

## 특정 부서 내 커뮤니케이션 네트워크

특정 팀, 특정 인물이 부서간 징검다리 역할을 함

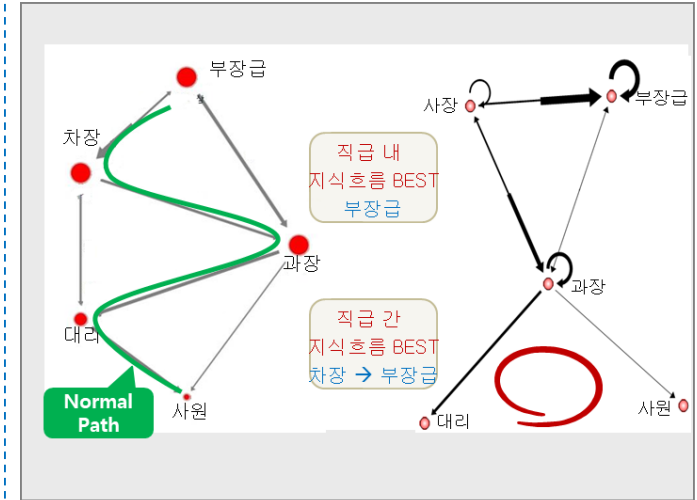


## 비공식네트워크 커뮤니티 추출

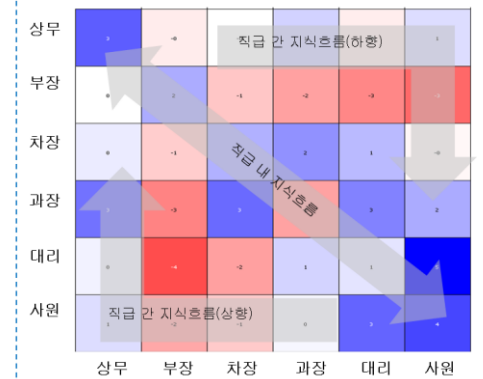


## 직급별 지식흐름 구조

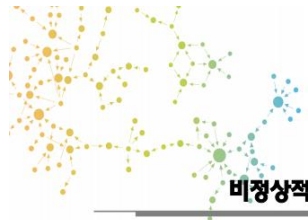
### 직급별 지식흐름 구조



### 직급별 지식흐름 구조 - Matrix Diagram



# 분석사례 - SNS내 비정상적 패턴 분석



## 비정상적 SNS 활동 패턴 그룹 분석

July, 2016

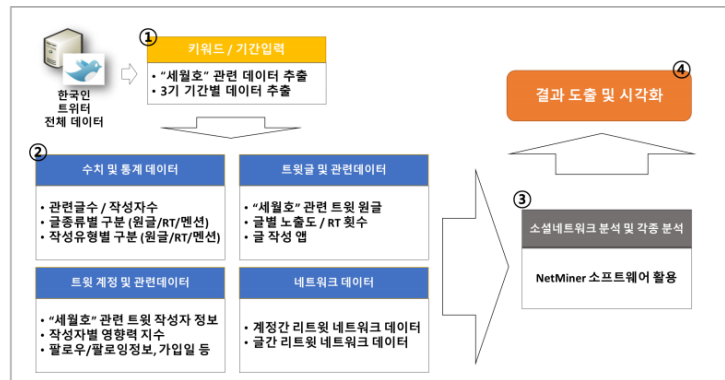
## Table of Contents

- I. 분석 목표
- II. 기초통계 및 분석통계 요약
- III. 비정상적 트윗작성 패턴분석 개요
- IV. 비정상적 트윗작성 패턴분석 알고리즘
- V. 기수별 비정상 트윗 패턴 추출
  - A. 1기 혐의그룹 추출
  - B. 2기 혐의그룹 추출
  - C. 3기 혐의그룹 추출
- VI. 혐의의심그룹활동 변화이상 파악
  - A. 2기 기반 추출그룹 특징
  - B. 2기 기반 추출그룹 행동패턴 종합본

### I. 분석 목표

비정상적 트윗작성 패턴 기반의 혐의 의심그룹 분석  
소셜네트워크 분석기법과 트위터 데이터상의 특징을 활용한 비정상적 커뮤니케이션 패턴을 발견하고 분석함

#### 데이터 추출 및 분석 과정 도식화



### II. 기초 통계 및 분석통계 요약

#### 기초 현황 요약 - “세월호” 관련 트윗 작성 글 수 및 작성자 기수별 분석

- 세월호와 관련한 기수별 글 수는 사전발생 후 열흘간인 1기때 가장 많이 작성되었고, 3기와 2기순으로 글이 작성됨
- 자신의 생각을 직접적으로 작성하는 “원글”의 경우에도 1기>3기>2기 순이었지만, 1인당 평균 작성글 수 및 1인당 원글 수는 2기가 가장 높은 것을 볼 수 있음
- 원글 1건당 평균 RT 및 멘션수가 역시도 2기가 가장 높은 것을 볼 수 있는데, 이는 다른 시기에 비해 2기때는 자신의 생각을 직접적으로 적는 사람들 보다 누군가의 생각을 “전달” 하거나 “전파” 하는 RT와 비율이 높은 것을 알 수 있음

	1기(14.4.16-26)	2기(14.8.19-29)	3기(15.4.16-26)
전체 글수	1,969,383 건	951,130 건	1,007,578 건
전체 작성자수	183,476 명	71,904 명	112,161 명
1인당 평균 작성글 수	10.7 건	13.2 건	9 건
원 글수	319,158 건	117,221 건	133,969 건
원글 작성자수	75,151 명	26,468 명	44,995 명
1인당 평균 원글 수	4.2 건	4.4 건	3 건
원글1건당 평균 RT 및 멘션수	6.2 건	8.1 건	7.5 건

원글 작성자수  
RT 및 멘션작성자수



### I. 분석 목표

비정상적 트윗작성 패턴 기반의 혐의 의심그룹 분석  
소셜네트워크 분석기법과 트위터 데이터상의 특징을 활용한 비정상적 커뮤니케이션 패턴을 발견하고 분석함

#### 분석대상 및 데이터 추출과정

- 분석대상 - 한국인 트위터 데이터 중 “세월호” 라는 키워드를 활용하여 데이터를 추출함
- 분석기간 - 여론의 흐름을 보기 위해 총 3기에 걸쳐 SNS내의 여론변화를 분석
  - 1기 : 2014. 4. 16. ~ 2014. 4. 26. (참사발생 직후)
  - 2기: 2014. 8. 19. ~ 2014. 8. 29. (특별법 합의안 재협상 요구-단식농성 중단)
  - 3기: 2015. 4. 11. ~ 2015. 4. 21. (2주기)

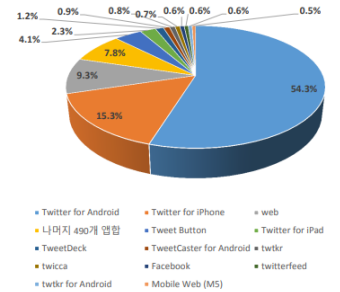
### III. 비정상적 트윗 작성 패턴분석

#### 트위터 글 작성 App이란?

- 트위터 상에 글을 남길 때, 내가 어떠한 앱을 사용하여 작성하였는지를 분석 서버를 통해 확인 할 수 있음
- 일반적인 사용자는 스마트 폰에서 모바일 앱을 활용하거나 인터넷 사이트를 통해 트윗을 작성함 (모바일 공식앱 70%, 공식 홈페이지 10% 정도의 비율)
- 이밖에도 트위터 모바일 앱과 웹사이트를 제외한 “서드파티 앱”을 활용하여 글을 작성하기도 하는데, 이러한 서드파티 앱들은 특정 목적이나 활용성을 위해 사용할 때가 상당수 있음.

#### 1기 (14.4.16-26)때 작성된 트윗글들의 앱 상위 15개 정보 (전체 505개의 앱)

App명	작성글수	비율
Twitter for Android	1,070,168	54.3%
Twitter for iPhone	301,275	15.3%
web	183,635	9.3%
Tweet Button	79,825	4.1%
Twitter for iPad	45,303	2.3%
TweetDeck	23,874	1.2%
TweetCaster for Android	17,735	0.9%
twtkr	15,958	0.8%
twicca	14,172	0.7%
Facebook	12,762	0.6%
twitterfeed	11,874	0.6%
twtkr for Android	11,248	0.6%
Mobile Web (M5)	9,986	0.5%
Linkis.com	9,871	0.5%
ttbird	8,971	0.5%





# 분석사례 - SNS내 비정상적 패턴 분석

## III. 비정상적 트윗 작성 패턴분석

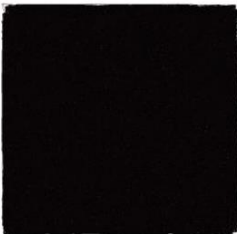
TweetDeck 에 대한 이슈제가 : 2013년 9월 뉴스타파에서 국정원이 트윗덱을 사용하여 같은 시간에 글을 여러 개를 작성하거나 동시에 RT를 하는것을 발견함 (봇 프로그램)  
[http://www.youtube.com/watch?v=gBr8\\_YqPY4&feature=youtu.be&t=16m12s](http://www.youtube.com/watch?v=gBr8_YqPY4&feature=youtu.be&t=16m12s)

**Why 트윗덱?**

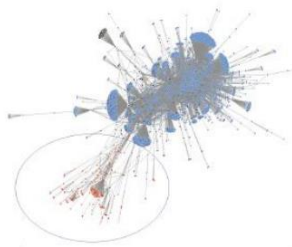
- 여러 개의 계정 동시 관리 가능 (동시 글쓰기 및 동시 RT가능)
- 시간 예약 트윗 가능
- 특정 트윗 알람 가능
- 리스트 및 멤버 관리 가능

## V. 기수별 비정상 트윗 패턴 추출- 1기

RT네트워크 시각화 (1기)



(1회이상 RT관계 네트워크)  
 유저수 : 157,070  
 RT관계수 : 1,209,200  
 너무나 복잡한 관계들로 얽혀있어 패턴을 찾아내기 어려움



(10회 이상 RT관계 네트워크)  
 유저수 : 3,589  
 RT관계수 : 7,192  
 전형적인 단계별 RT의 패턴을 보여줌. 특징들에 대해 영향력자를 중심으로 단계별로 퍼져나가는 패턴을 보여줌

## III. 비정상적 트윗 작성 패턴분석

뉴스타파 외에도 비정상적인 트윗작성패턴을 분석하기 위해 만지일보 및 한겨레에서도 일명 "조장-조원" 패턴의 RT네트워크 분석을 시도함

과거 언론사 비정상적 SNS커뮤니케이션 패턴 분석 시도내용

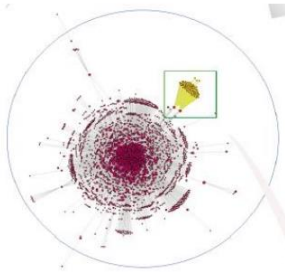
RT패턴 및 계정 활동내역을 기반으로 한 비정상적패턴 분석시도  
 만지일보  
<http://www.ddanzi.com/Notice/1009442>  
 일명 '조장-조원' 패턴을 분석하려 시도함

한겨레에서도 이러한 RT네트워크 분석을 추가로 시도함  
<http://www.hani.co.kr/art/economy/514136.html>

비정상적인 패턴을 찾아 분석한 결과, 뉴스타파와 유사한 패턴이 발견되었습니다. 이는 언론사와 관련된 계정들이 특정 계정을 팔로우하거나 팔로잉하는 패턴이 관찰되었습니다. 이는 '조장-조원' 패턴의 일종으로, 특정 계정이 다른 계정을 팔로우하거나 팔로잉하는 패턴을 보여줍니다. 이는 '조장-조원' 패턴의 일종으로, 특정 계정이 다른 계정을 팔로우하거나 팔로잉하는 패턴을 보여줍니다. 이는 '조장-조원' 패턴의 일종으로, 특정 계정이 다른 계정을 팔로우하거나 팔로잉하는 패턴을 보여줍니다.

## V. 기수별 비정상 트윗 패턴 추출- 1기

RT네트워크 시각화 (1기)



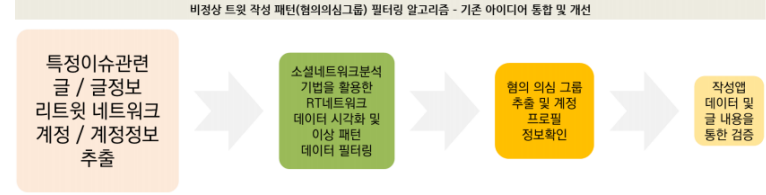
단계별 RT네트워크 중 중심과 떨어진 네트워크를 별도 추출하였더니 녹색네트워크 같이 특정 계정에게만 RT를 하는 이상패턴을 발견함



비정상적 SNS 커뮤니케이션 의심 패턴 추출

## VI. 비정상적 트윗 작성 패턴 필터링 알고리즘

- 기존 언론과 연구 등에서 사용한 비정상 트윗 추출 방식은 전체 트윗 데이터를 대상으로 전수조사에서 시작하기 때문에 연구 범위가 방대하여 시간이 오래 걸리고, 관계정보들을 추가로 작성하거나 일부 정보만 확인할 수 있어 한계점을 지니고 있음
- 본 분석에서는 지금까지 시도된 다양한 조사방식과 함께 소셜네트워크 분석기반의 RT네트워크 시각화 및 패턴 데이터 필터링을 통해 비정상적 커뮤니케이션 패턴을 추출하고 검증하는 알고리즘을 기획함



## V. 기수별 비정상 트윗 패턴 추출- 1기

비정상적 SNS커뮤니케이션 의심패턴 검증 - APP 사용패턴 분석

- 추출된 99명이 어떤 앱을 통해 트위터를 하는지 사용 앱을 분석해 보니 데이터 유실로 출처를 알수없는 3명을 제외한 96명 모두가 TweetDeck 을 사용하여 글을 작성함

아이디	사명명	직업명	아이디	사명명	직업명	아이디	사명명	직업명
101	TweetDeck	7	102	TweetDeck	7	103	TweetDeck	7
104	TweetDeck	7	105	TweetDeck	7	106	TweetDeck	7
107	TweetDeck	7	108	TweetDeck	7	109	TweetDeck	7
110	TweetDeck	7	111	TweetDeck	7	112	TweetDeck	7
113	TweetDeck	7	114	TweetDeck	7	115	TweetDeck	7
116	TweetDeck	7	117	TweetDeck	7	118	TweetDeck	7
119	TweetDeck	7	120	TweetDeck	7	121	TweetDeck	7
122	TweetDeck	7	123	TweetDeck	7	124	TweetDeck	7
125	TweetDeck	7	126	TweetDeck	7	127	TweetDeck	7
128	TweetDeck	7	129	TweetDeck	7	130	TweetDeck	7
131	TweetDeck	7	132	TweetDeck	7	133	TweetDeck	7
134	TweetDeck	7	135	TweetDeck	7	136	TweetDeck	7
137	TweetDeck	7	138	TweetDeck	7	139	TweetDeck	7
140	TweetDeck	7	141	TweetDeck	7	142	TweetDeck	7
143	TweetDeck	7	144	TweetDeck	7	145	TweetDeck	7
146	TweetDeck	7	147	TweetDeck	7	148	TweetDeck	7
149	TweetDeck	7	150	TweetDeck	7	151	TweetDeck	7
152	TweetDeck	7	153	TweetDeck	7	154	TweetDeck	7
157	TweetDeck	7	158	TweetDeck	7	159	TweetDeck	7
160	TweetDeck	7	161	TweetDeck	7	162	TweetDeck	7
163	TweetDeck	7	164	TweetDeck	7	165	TweetDeck	7
168	TweetDeck	7	169	TweetDeck	7	170	TweetDeck	7
173	TweetDeck	7	174	TweetDeck	7	175	TweetDeck	7
178	TweetDeck	7	179	TweetDeck	7	180	TweetDeck	7
183	TweetDeck	7	184	TweetDeck	7	185	TweetDeck	7
188	TweetDeck	7	189	TweetDeck	7	190	TweetDeck	7
193	TweetDeck	7	194	TweetDeck	7	195	TweetDeck	7
198	TweetDeck	7	199	TweetDeck	7	200	TweetDeck	7
203	TweetDeck	7	204	TweetDeck	7	205	TweetDeck	7
208	TweetDeck	7	209	TweetDeck	7	210	TweetDeck	7
213	TweetDeck	7	214	TweetDeck	7	215	TweetDeck	7
218	TweetDeck	7	219	TweetDeck	7	220	TweetDeck	7
223	TweetDeck	7	224	TweetDeck	7	225	TweetDeck	7
228	TweetDeck	7	229	TweetDeck	7	230	TweetDeck	7
233	TweetDeck	7	234	TweetDeck	7	235	TweetDeck	7
238	TweetDeck	7	239	TweetDeck	7	240	TweetDeck	7
243	TweetDeck	7	244	TweetDeck	7	245	TweetDeck	7
248	TweetDeck	7	249	TweetDeck	7	250	TweetDeck	7
253	TweetDeck	7	254	TweetDeck	7	255	TweetDeck	7
258	TweetDeck	7	259	TweetDeck	7	260	TweetDeck	7
263	TweetDeck	7	264	TweetDeck	7	265	TweetDeck	7
268	TweetDeck	7	269	TweetDeck	7	270	TweetDeck	7
273	TweetDeck	7	274	TweetDeck	7	275	TweetDeck	7
278	TweetDeck	7	279	TweetDeck	7	280	TweetDeck	7
283	TweetDeck	7	284	TweetDeck	7	285	TweetDeck	7
288	TweetDeck	7	289	TweetDeck	7	290	TweetDeck	7
293	TweetDeck	7	294	TweetDeck	7	295	TweetDeck	7
298	TweetDeck	7	299	TweetDeck	7	300	TweetDeck	7
303	TweetDeck	7	304	TweetDeck	7	305	TweetDeck	7
308	TweetDeck	7	309	TweetDeck	7	310	TweetDeck	7
313	TweetDeck	7	314	TweetDeck	7	315	TweetDeck	7
318	TweetDeck	7	319	TweetDeck	7	320	TweetDeck	7
323	TweetDeck	7	324	TweetDeck	7	325	TweetDeck	7
328	TweetDeck	7	329	TweetDeck	7	330	TweetDeck	7
333	TweetDeck	7	334	TweetDeck	7	335	TweetDeck	7
338	TweetDeck	7	339	TweetDeck	7	340	TweetDeck	7
343	TweetDeck	7	344	TweetDeck	7	345	TweetDeck	7
348	TweetDeck	7	349	TweetDeck	7	350	TweetDeck	7
353	TweetDeck	7	354	TweetDeck	7	355	TweetDeck	7
358	TweetDeck	7	359	TweetDeck	7	360	TweetDeck	7
363	TweetDeck	7	364	TweetDeck	7	365	TweetDeck	7
368	TweetDeck	7	369	TweetDeck	7	370	TweetDeck	7
373	TweetDeck	7	374	TweetDeck	7	375	TweetDeck	7
378	TweetDeck	7	379	TweetDeck	7	380	TweetDeck	7
383	TweetDeck	7	384	TweetDeck	7	385	TweetDeck	7
388	TweetDeck	7	389	TweetDeck	7	390	TweetDeck	7
393	TweetDeck	7	394	TweetDeck	7	395	TweetDeck	7
398	TweetDeck	7	399	TweetDeck	7	400	TweetDeck	7
403	TweetDeck	7	404	TweetDeck	7	405	TweetDeck	7
408	TweetDeck	7	409	TweetDeck	7	410	TweetDeck	7
413	TweetDeck	7	414	TweetDeck	7	415	TweetDeck	7
418	TweetDeck	7	419	TweetDeck	7	420	TweetDeck	7
423	TweetDeck	7	424	TweetDeck	7	425	TweetDeck	7
428	TweetDeck	7	429	TweetDeck	7	430	TweetDeck	7
433	TweetDeck	7	434	TweetDeck	7	435	TweetDeck	7
438	TweetDeck	7	439	TweetDeck	7	440	TweetDeck	7
443	TweetDeck	7	444	TweetDeck	7	445	TweetDeck	7
448	TweetDeck	7	449	TweetDeck	7	450	TweetDeck	7
453	TweetDeck	7	454	TweetDeck	7	455	TweetDeck	7
458	TweetDeck	7	459	TweetDeck	7	460	TweetDeck	7
463	TweetDeck	7	464	TweetDeck	7	465	TweetDeck	7
468	TweetDeck	7	469	TweetDeck	7	470	TweetDeck	7
473	TweetDeck	7	474	TweetDeck	7	475	TweetDeck	7
478	TweetDeck	7	479	TweetDeck	7	480	TweetDeck	7
483	TweetDeck	7	484	TweetDeck	7	485	TweetDeck	7
488	TweetDeck	7	489	TweetDeck	7	490	TweetDeck	7
493	TweetDeck	7	494	TweetDeck	7	495	TweetDeck	7
498	TweetDeck	7	499	TweetDeck	7	500	TweetDeck	7





# 분석사례 - SNS내 비정상적 패턴 분석

## V. 기수별 비정상 트윗 패턴 추출- 3기

비정상적 SNS커뮤니케이션 의심패턴 검증 - RT글 및 작성글 확인

혐의그룹이 3기때에 RT한 글들을 분석해보니 아래의 글처럼 조장계정이 직접 쓰거나 조장계정이 RT했던 글들을 공통적으로 RT하는 특성을 보임. 나머지 조원 계정들은 1, 2 기때와 마찬가지로 직접 작성한 글은 없으며 조장의 글을 RT하는 공통적인 특성을 보임.

작성자	작성시간	내용	RT 횟수	노출도
po1	12:22:37 AM	김동길김일 > 이문진 링크 걸지말 테다진 http://t.co/5vH2whakOe 나의 상자를 가져오는 세월호 안양을 이해할 수 없습니다.	106	28213
po1	12:41:37 AM	김동길김일 > 죽은 뒤에도 후회할 일을 http://t.co/18a9P32uYf 한 기업인의 자살이 <세월호>보다 더 심하게 대한민국을 괴롭힐 수도	102	26904
po1	12:56:34 AM	*세월호 천막 둘러싸며, 서울시 공무원이 20개 http://t.co/lMPuTUGwED "평화론 천막은 불법" 알렸다가 폭행당한 서울시 공무원이 전한 이야기	101	26687
po1	12:55:18 AM	교사 111명 "학교에 퇴원 요구... 또 천로2? http://t.co/TvNsdn58lg *정부 세월호 사명명 폐기하라???	104	26427
po1	12:54:32 AM	성환중 자살소동과 세월호 반정부단종 http://t.co/1a5Pw1c25U 자살/사상을 촉매제 전례로 학자에서 멀어	100	26248
po1	12:51:58 AM	북한, 세월호 1주기 맞아 3만 장부 성도 목사/이용달하는 남한국민들 http://t.co/WaUQe03y 우리민중여러, *세월호 참사가 한 정부의 부능에서 초래했다???	98	26044

po1	12:38:01 AM	세월호 참사 1주년이네요. 서울시 공무원이 20개 http://t.co/vR2323XaI 이혼 정부 받고 위해 자살로 돌아가시길 꾀하는 참마은 불법" 알렸다가 폭행당한 서울시 공무원이 전한 이야기	101	26987
po1	12:52:42 AM	세월호 참사 1주년을 맞으면서 http://t.co/52KqgPMuacE 국민들의 이해와 역량을 모아서 위기 극복하기를 다시 믿어주세요. 대한민국. 슬프지만 용기있는 고충 밝혀야 합니다.	98	25806
po1	12:53:49 AM	"이런 비극 재발 않도록... 모두 힘 모으자" http://t.co/PjGAM7h20c 경기중, 세월호 참사 1주기 맞아 인산재일교차에서 추모기도회	97	25802
po1	12:23:58 AM	세월호 가족에 대한 호소문! 대한민국 온 상생! 일사 http://t.co/2010년 4월 15일(수) 오전 1시 장소: 평화원이(순천동남) 앞 주차: 평화부대용사단 오시는길: 영동로 순천역을 구 역국은 행동입니다. 오이와 행동여과 전하	6	21277

[3기 혐의 그룹 주요 작성 글 및 RT글 (노출도 순 정렬/ 상위 10개만)]

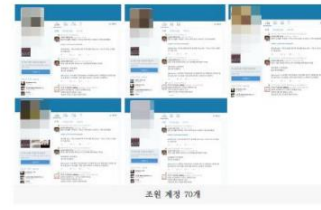
## VI. 혐의의심그룹활동 변화양상 파악 - 2기 (표본) 추출 그룹 특징

표본 2기 추출 그룹 71명 계정 특징

2기 표본 그룹은 타 그룹과 다르게 모든 계정이 2016년 6월 25일 현재까지 존재하고 있었으며, 모두가 가장 최근에 활동한 글이 2016년 4월 5일 조장추정계정에서 작성한 글을 리트윗한것으로 동일하며 그 외에 최근 활동들도 동일한 패턴을 보임



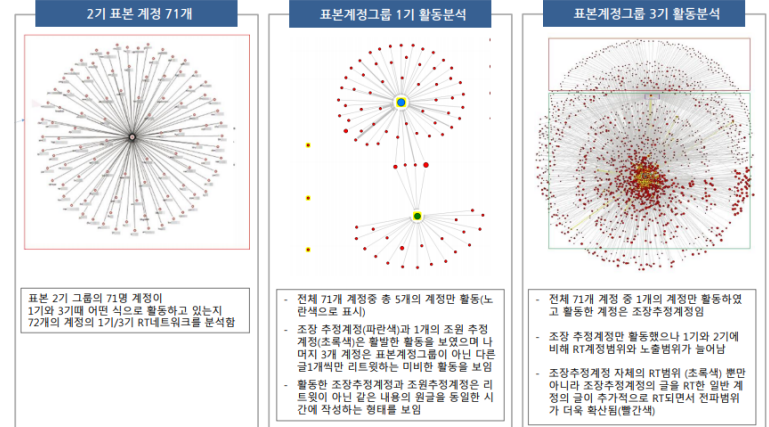
조장추정 계정



표본 계정 70개

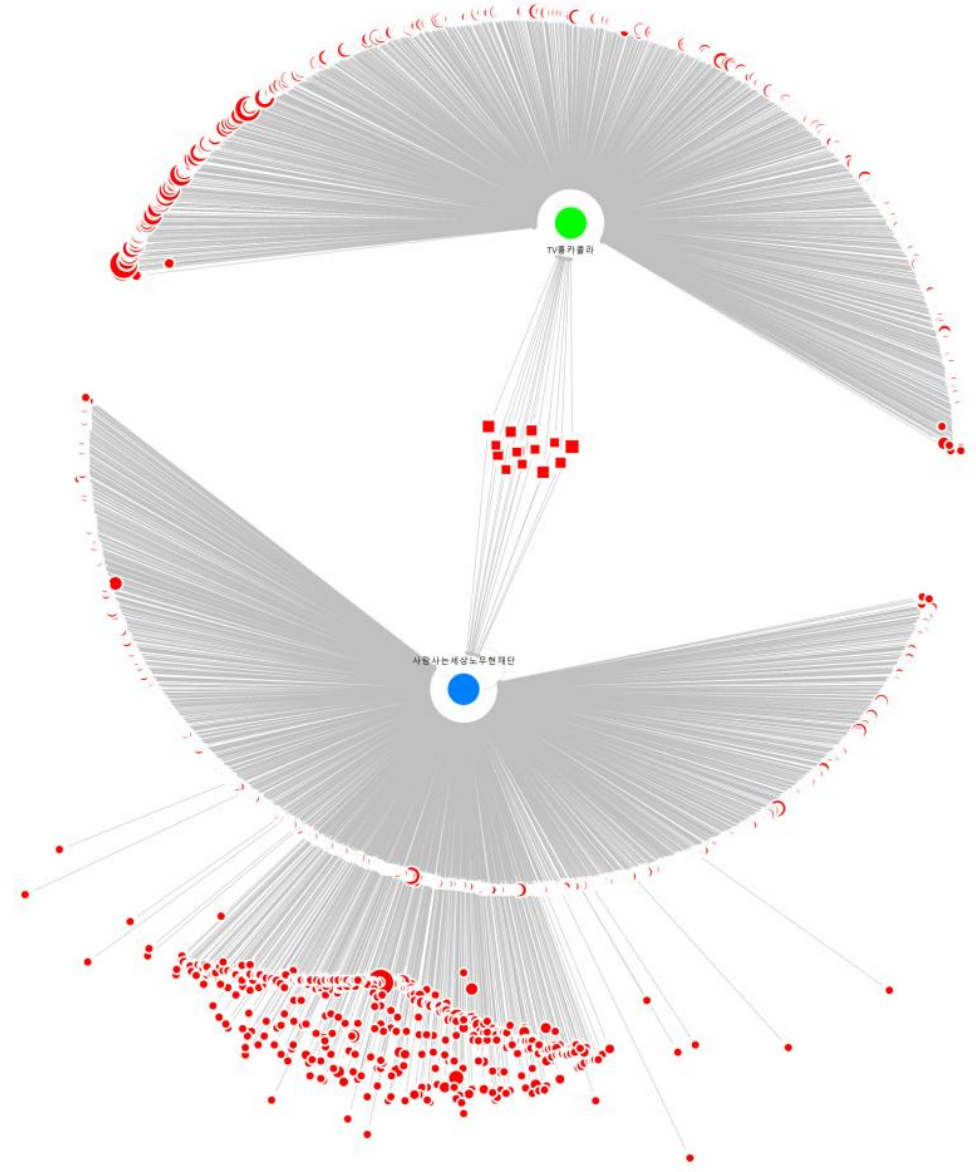
가입일이 2012년 12월로 동일 / 현재까지 존재하고 있음  
계정별로 작성된 글의 패턴 및 시간이 동일하며  
대부분의 내용이 조장계정이 작성한 글을 리트윗하였음

## VI. 혐의의심그룹활동 변화양상 파악 - 표본 그룹 전파력 범위 분석



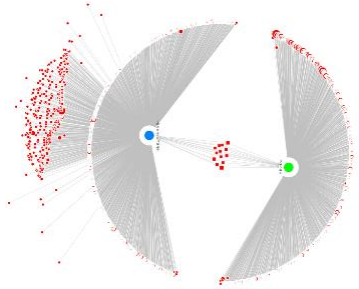
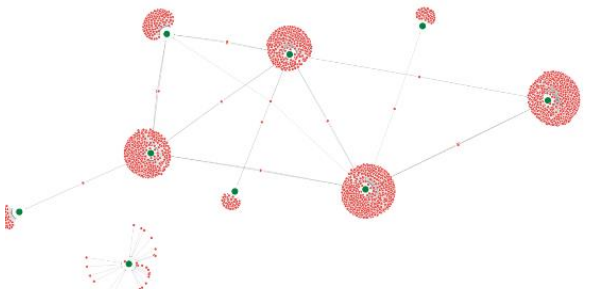
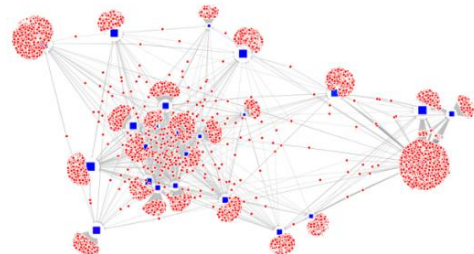
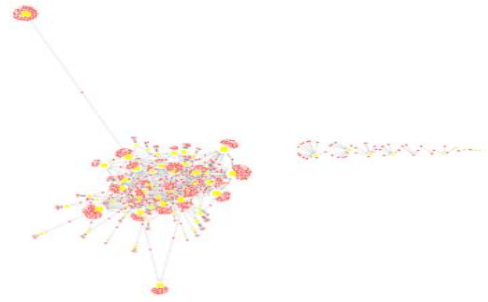
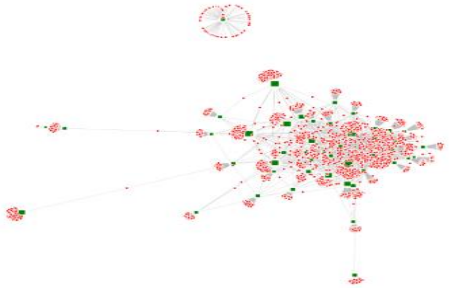
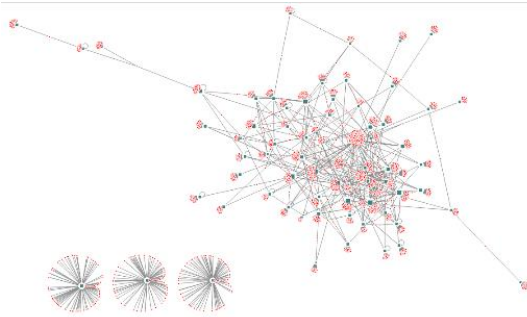
## ● 홍카콜라+알릴레오 유저 커뮤니케이션 패턴

- 유저 커뮤니케이션 패턴 맵은 오른쪽과 같음
- 초록색원은 홍카콜라, 파란색원은 노무현재단 계정임
- 빨간색 원은 댓글을 작성한 사람들이며 원의 크기는 댓글을 보낸 횟수를 의미함
- 두 네트워크간에 명확하게 커뮤니케이션이 거의 존재하지 않음을 시각적으로 볼 수 있음



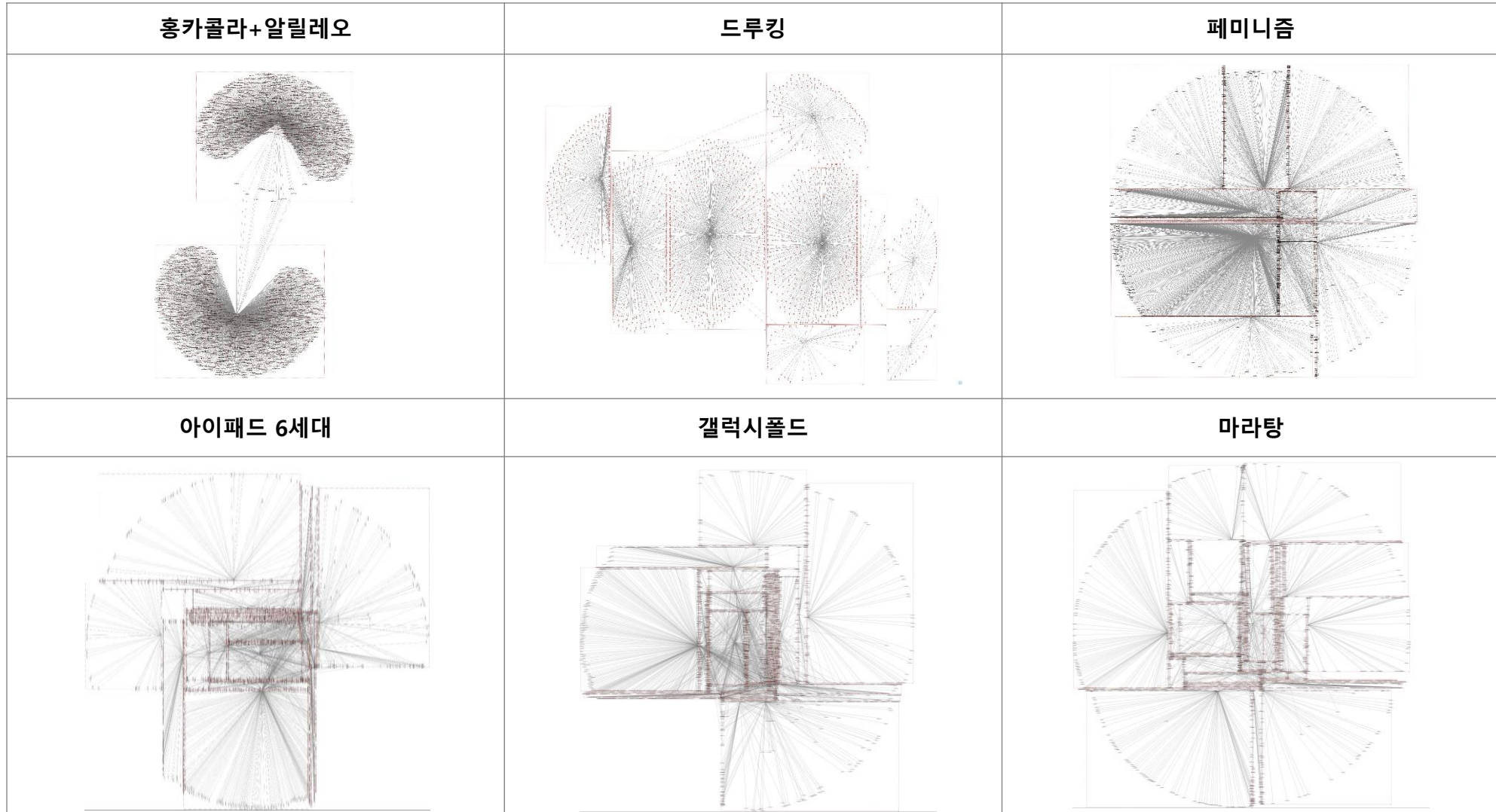
김덕진(2020). 유튜브 필터버블 현상에 대한 실증적 연구  
: 사회연결망분석과 텍스트네트워크 분석을 활용하여

# 분석사례 - 유튜브 커뮤니케이션 패턴 분석

홍카콜라+알릴레오			드루킹			페미니즘		
								
노드 4284	링크 4558	네트워크밀도 0.00022065	노드 2033	링크 1899	네트워크밀도 0.00054163	노드	링크 13766	네트워크밀도 0.0000979
아이패드 6세대			갤럭시폴드			마라탕		
								
노드 7917	링크 5095	네트워크밀도 0.00009785	노드 9303	링크 8910	네트워크밀도 0.00015050	노드 9442	링크 9101	네트워크밀도 0.00012038

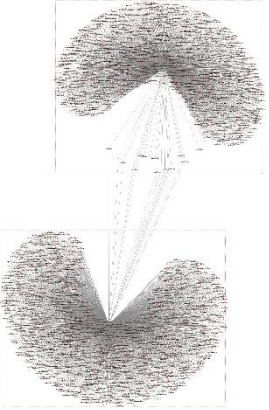
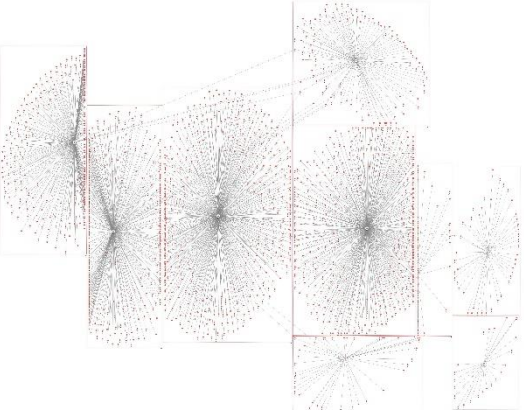
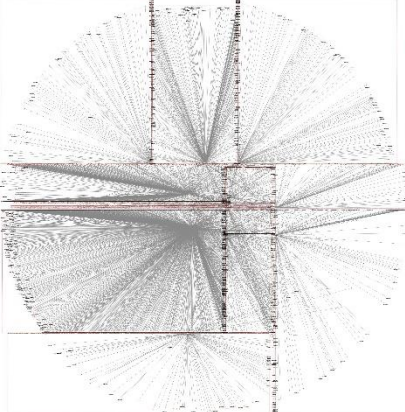
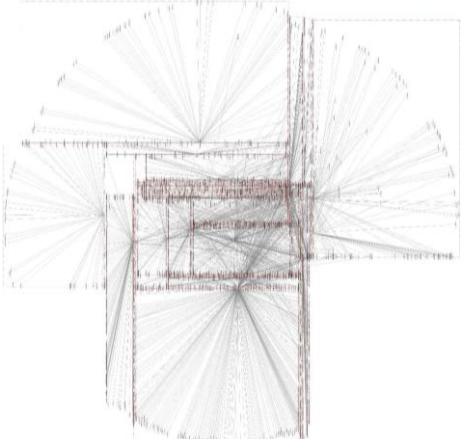
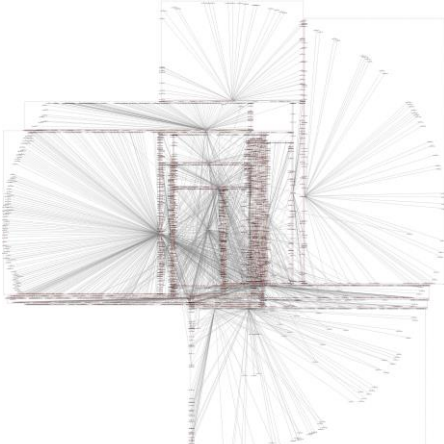
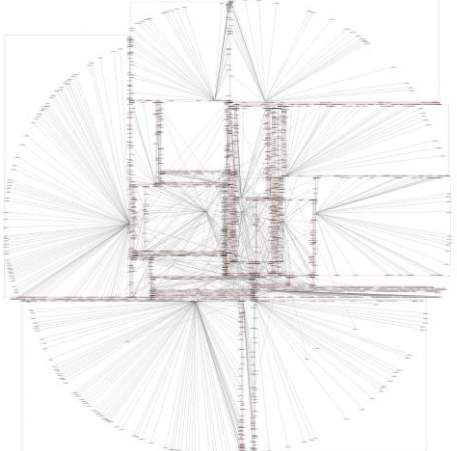


# 분석사례 - 유튜브 커뮤니케이션 패턴 분석





# 분석사례 - 유튜브 커뮤니케이션 패턴 분석

홍카콜라+알릴레오	드루킹	페미니즘
 Network graph showing a central hub with two large, dense clusters of nodes connected to it, one above and one below.	 Network graph showing a central hub with several smaller, dense clusters of nodes connected to it, arranged in a horizontal line.	 Network graph showing a central hub with many smaller, dense clusters of nodes connected to it, arranged in a circular pattern.
아이패드 6세대	갤럭시폴드	마라탕
 Network graph showing a central hub with many smaller, dense clusters of nodes connected to it, arranged in a circular pattern.	 Network graph showing a central hub with many smaller, dense clusters of nodes connected to it, arranged in a circular pattern.	 Network graph showing a central hub with many smaller, dense clusters of nodes connected to it, arranged in a circular pattern.

# 분석사례 - 유튜브 커뮤니케이션 패턴 분석

주요지표	홍+알	드루킹	페미니즘	아이패드 6세대	갤럭시 폴드	마라탕
step #	4544	1861	12730	4694	7740	8767
Best Modularity (0.3-1)	0.49	0.79	0.854	0.882	0.84	0.95
# of communities	<b>2</b>	<b>9</b>	<b>24</b>	<b>41</b>	<b>42</b>	<b>71</b>
평균 E-I Index	-0.988	-0.933	-0.679	-0.767	-0.607	-0.883
평균 SMI	0.995	0.996	0.993	0.996	0.993	0.999
평균 Cohesion Index	162.3	546.0	209.9	656.8	768.9	2262.4
평균 Density	0.001	0.013	0.002	0.076	0.007	0.009
평균 Group Modularity	0.245	0.088	0.036	0.021	0.020	0.013
총 영상수	2	10	100	100	100	100
업로드 유저수	<b>2</b>	<b>9</b>	<b>24</b>	<b>71</b>	<b>49</b>	<b>75</b>

지표명	설명
E-I Index	<p>파티션 내에 속한 노드에 연결된 링크의 수가 파티션 내에 많은지 파티션 밖에 많은지를 측정함</p> <p>-1~1 사이의 값을 가지며 파티션 내의 링크가 많을수록 -1에 가까워지고, 파티션 밖의 링크가 많은 수록 1에 가까움</p>
SMI	<p>파티션 밖의 링크 밀도와 파티션안의 링크 밀도 비율을 계산한 값으로 -1~1사이의 값을 가짐</p> <p>E-I Index와 달리 -1에 가까울수록 파티션 밖의 링크밀도가 높고, 1에 가까울수록 파티션 안쪽의 링크 밀도가 높다는 것을 의미</p>
Cohesion Index	<p>파티션 안의 링크밀도를 파티션 밖의 링크밀도로 나눈 값</p> <p>값이 1일 경우 파티션 내 링크와 파티션 외 링크 밀도가 같다는 의미이고 1보다 클수록 파티션 내 링크 밀도가 높다는 것을 의미함</p>

6개의 네트워크 응집그룹의 특징은 표처럼 정리 할 수 있으며 각각의 값들을 통해 네트워크 응집그룹이 범위안에서 잘 구분되었음을 알 수 있다.

여기서 이 응집그룹을 기존 네트워크의 특징과 비교해보았다. 실제 영상을 올린 업로드 유저수와 Best modularity를 통해 나누어진 커뮤니티 수를 비교해 본 것인데, it주제인 아이패드와 갤럭시폴드를 제외하고는 정치, 사회, 문화생활 분야에서 업로드한 유저수와 커뮤니티 개수가 일치하거나 비슷한 현상을 보여주었다.

이것은 영상을 올린 사람과 댓글을 작성한 유저간의 커뮤니케이션의 형태가 명확히 응집되어서 그룹으로 구분될 수 있음을 이야기하고 이는 한명의 크리에이터가 올린 영상에 대해 댓글을 통해 커뮤니케이션하는 유저가 같은 주제 내의 다른 영상을 올린 크리에이터와는 댓글을 통한 커뮤니케이션이 없다는 것을 반증하는 지표이며, 유튜브 댓글 커뮤니케이션을 통해 실증한 유튜브 필터버블 현상은 존재하고 있음을 증명할 수 있다.

# 분석사례 - 유튜브 커뮤니케이션 패턴 분석

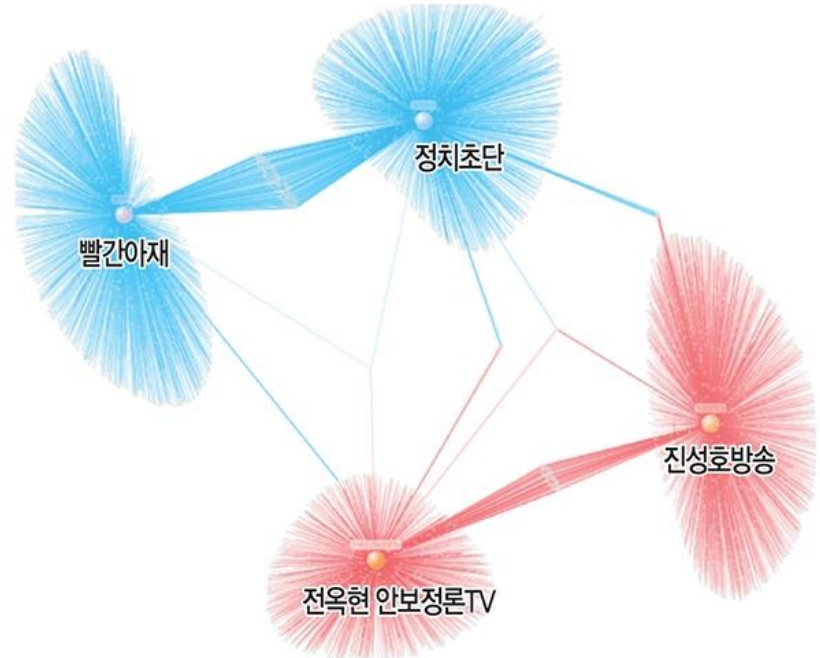
국민일보

## 고삐 풀린 유튜브의 '극과 극' 알고리즘, 갈등 키운다 [이슈&탐사]

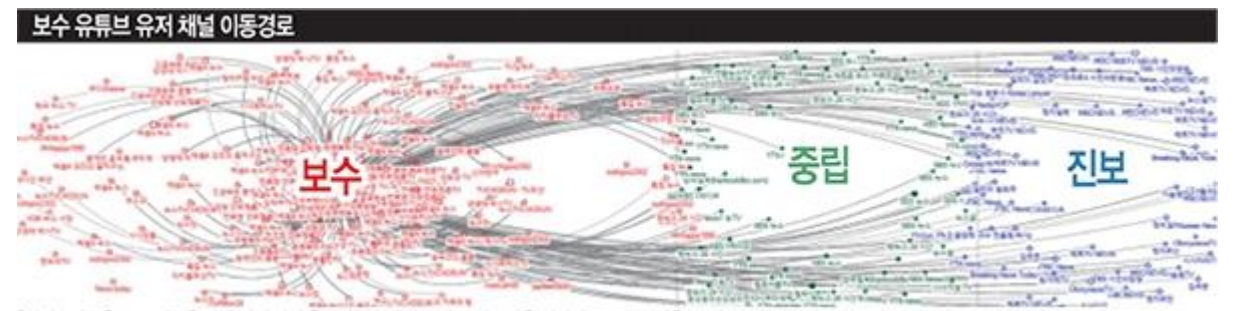
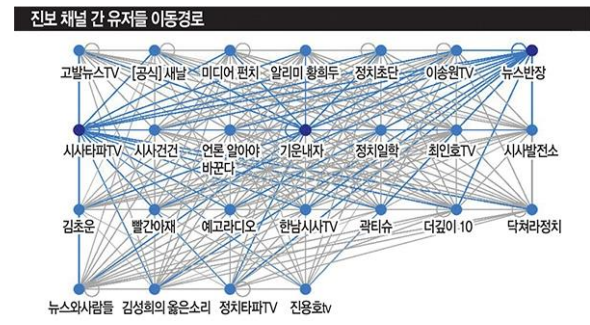
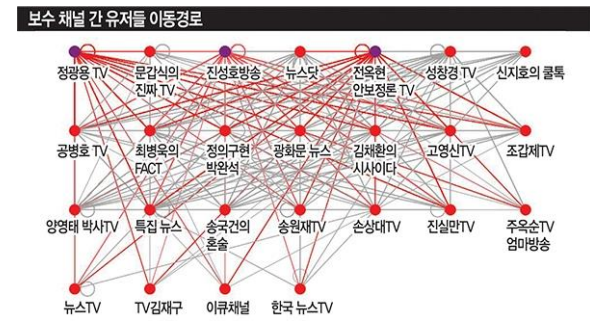
기사입력 2020-12-10 00:02 최종수정 2020-12-10 09:58

| [극단으로 안내하는 알고리즘 해설서-상식이 2개인 나라] ① 극단의 광장에 갇힌 사람들

### 진보·보수 채널 이용자 분석



\*파란색이 진보 채널 이용자들, 붉은 색은 보수 채널 이용자들.



\*진성호방송을 보고 댓글을 단 유저의 직전 출발지를 나타낸 그래프, 대부분은 직전에도 보수 영상을 본 것으로 나타났다.

<https://news.naver.com/main/read.naver?mode=LSD&mid=sec&sid1=102&oid=005&aid=0001389305>



# 감사합니다

한국인사이트연구소 김덕진 부소장

[socialkim@ki.re.kr](mailto:socialkim@ki.re.kr)

다양한 소셜채널에서의 소통 환영합니다 ^^

