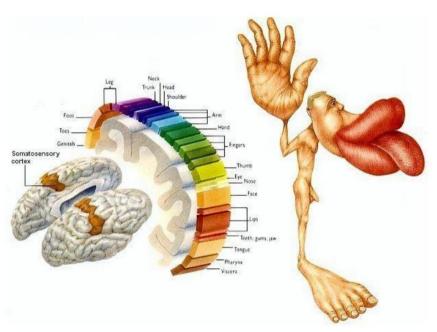
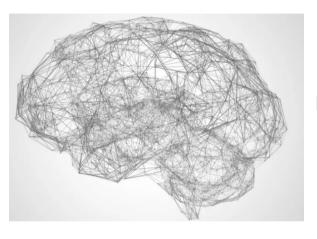
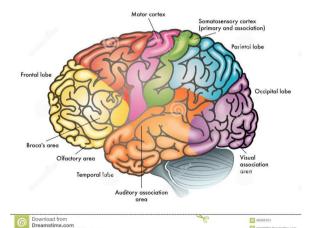
# 7. 네트워크와 집단



Penfield Brain Map

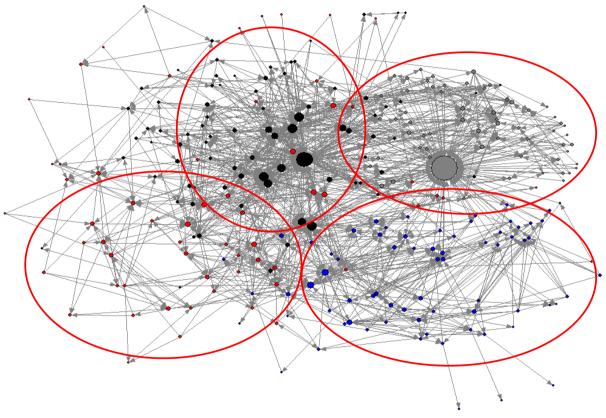


**Brain Neural Network** 



Rob3000 | Dreamstime.com

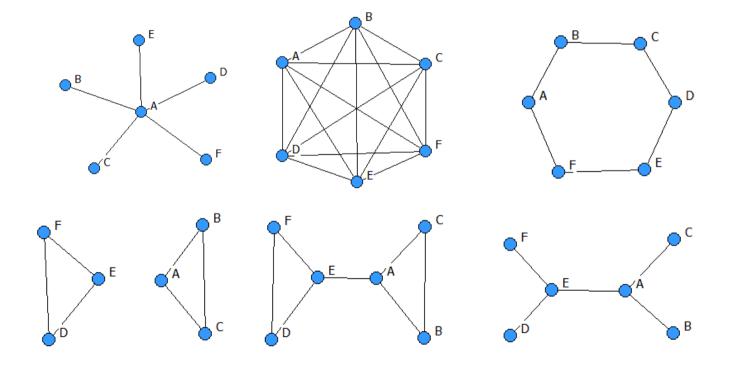
**Brain Functional Areas** 



예쁜꼬마선충의 Neural Network (tie strength > 2) (data: celegans\_net.xlsx)

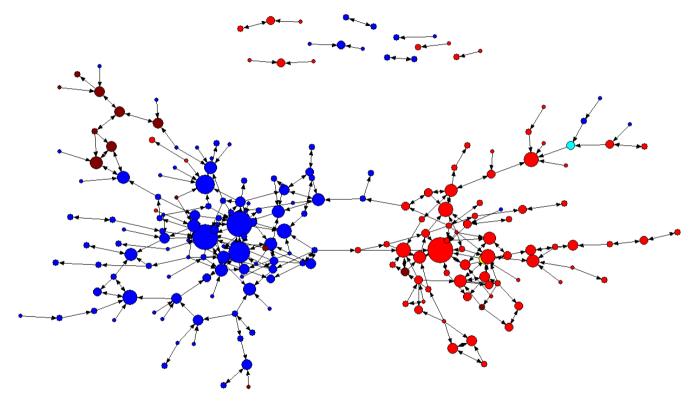
## 네트워크와 하위집단

- 아래에 있는 여러가지 네트워크를 보고 복수의 집단으로 구성된 네트워크는 어떤 것인지 찾아보자
- 여러 집단으로 구성된 네트워크와 그렇지 않은 네트워크 간에는 어떤 특징적 차이가 있는지 생각해 보자



### 네트워크와 하위집단

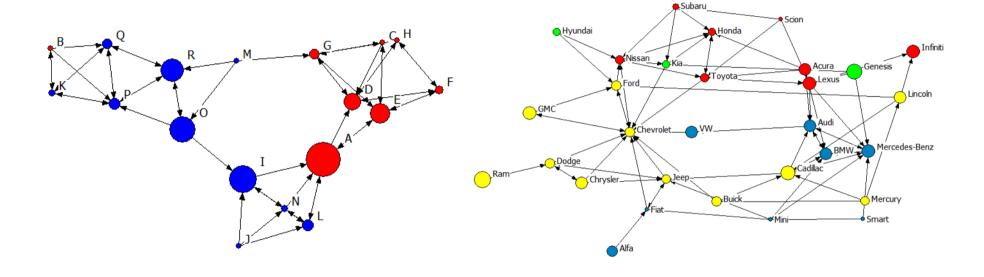
- 하위집단에 대한 다양한 approach
  - ① 명목집단 (nominal group): 성, 연령대, 장르, 산업, 토픽 등과 같이 미리 정해진 범주로 분류된 집단 ⇒ 지난 주에 속성에 따른 E-I Index를 배웠음
  - ② <u>실질집단 (real group)</u>: <u>노드 간 연결을 기반으로 분류한 집단 ⇒ 연결 밀도가 상대적으로 높은 노드의 집합 ⇒ 이번주에 배울 내용</u>
  - ③ 역할집단 (role-based group): 네트워크에서 같은 역할을 수행하는 노드의 집합 ⇒ 9. 위치와 역할에서 배움
- 하위집단으로 무엇을 알 수 있을까?
  - 네트워크의 구조적 특성과 기능 → 갈등, 경쟁, 조정 등 (non-overlapping groups; membership sharing)
  - 연결의 패턴이 유사한 노드들을 어떻게 묶을 수 있을까? (예: 우리 회사에는 어떤 파벌이 존재하는가? 고객들이 같이 구매하는 상품은 어떤 것일까? 같이 검색하는 단어들을 어떻게 분류할 수 있을까?...)
  - 전체 네트워크가 어떤 기제(mechanism)에 의해 구조화(분절, 통합, 분류 등)되는가?
- 다음에 제시되는 몇 개의 네트워크를 살펴보면서 하위집단의 의미를 생각해 보자



Social network(acquaintance network) of drug users in Hartfort (colored by race)

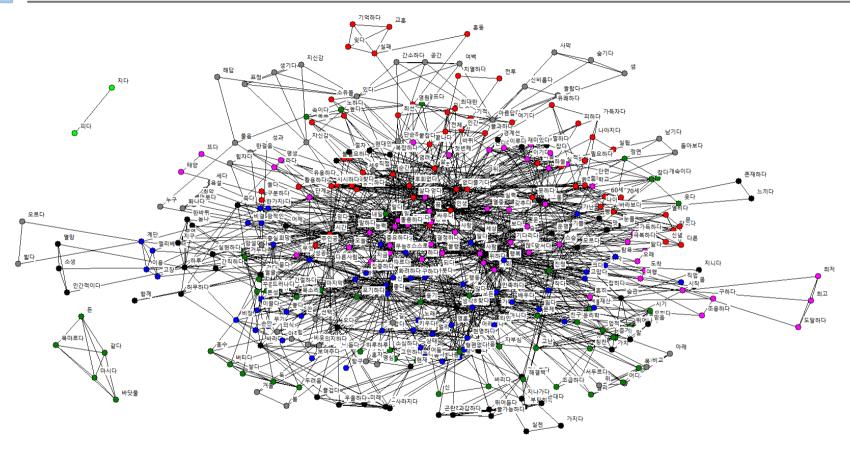
How many subgroups do you see?

[drugnet.##h & drugattr.##h]

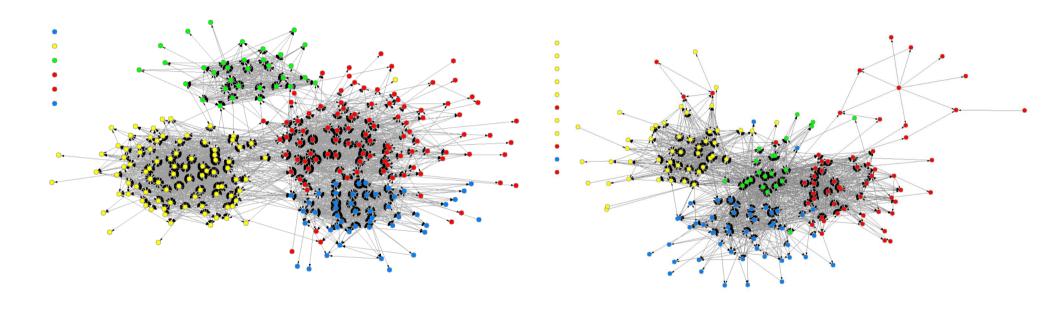


Advice Network (colored by gender)
[interpersonal.##h & Interpersonal\_att.##h]

Car Google Search Network (colored by country) [auto\_node.##h & auto\_att.##h]



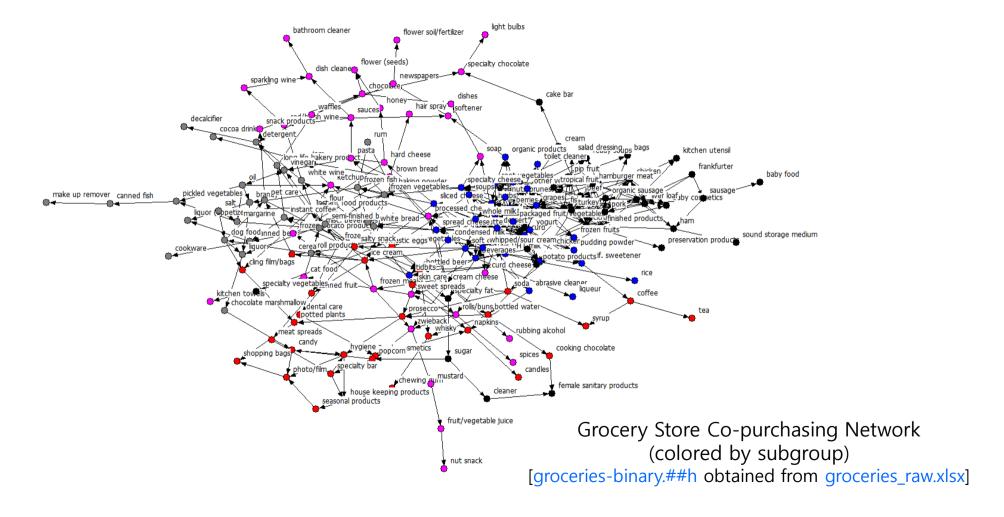
Word Network from 120 Famous Quotes (colored by subgroups) [quotes\_word.##h]



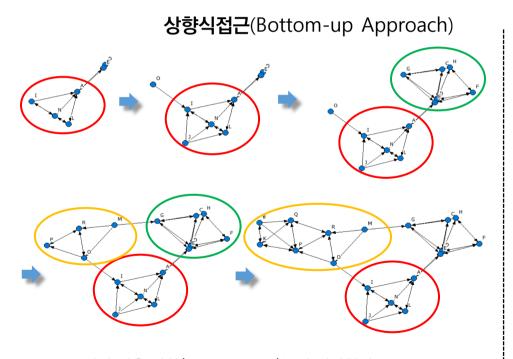
Employee Email Networks: Tow Divisions in a Multidivisional Company

[company.xlsx, IslandDiv\_att.xlsx, CrossDiv\_att.xlsx]

### 네트워크와 하위집단

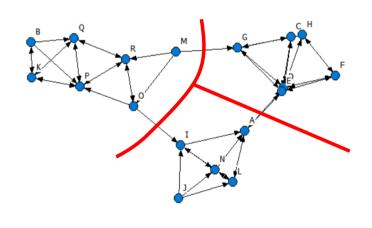


• 두 가지 접근 방법: 상향식 접근(bottom-up approach) & 하향식 접근(top-down approach)



가장 작은 단위(dyad or triad)부터 시작하여 연결밀도가 낮아질 때까지 계속 묶어 나감

하향식 접근(Top-down Approach)



전체 네트워크를 한 눈에 보고 연결밀도에서 두드러진 차이가 나는 곳을 발견하는 방법

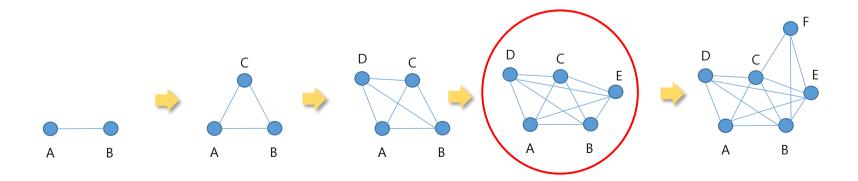
### 2

## 상향식 접근 (Clique)

- 클릭(clique: 파벌?)
  - 네트워크에서 가장 작은 연결의 단위는 이자관계(dyad) → 노드가 하나 추가되면 삼자관계(triad)
  - 클릭: 셋 이상의 노드로 구성된 최대 완전 서브 그래프 (maximal complete subgraph)
  - 완전(Complete): 클릭 내의 모든 노드들이 직접 연결되어 있어야 함
  - 최대(Maximal): 노드를 추가했을 때 완전성이 깨지면 안 됨

상향식 접근은 개념만 배우고 실습은 하지 않습니다

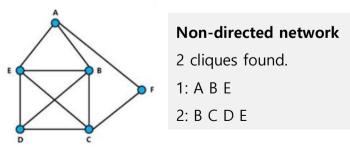
• 아래 그림에서 클릭은?

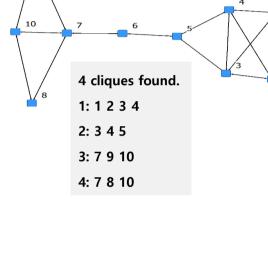


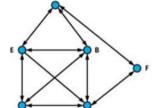
### 2

## 상향식 접근 (Clique)

- 다음 그림에서 클릭을 찾아보자
  - {1, 2, 3, 4}, {3, 4, 5}, (7, 8, 10}, (7, 9, 10}
  - {1, 2, 3}, {5, 6}, (7, 8, 9, 10}은 왜 클릭이 아닐까?
  - 하나의 노드가 여러 개의 클릭에 포함될 수도 있음
- Directed Data의 경우
  - 호혜성(reciprocity) 요건 추가 → 클릭 내에 포함된 노드들 간의 완벽한 호혜적(쌍방향) 연결







#### **Directed network**

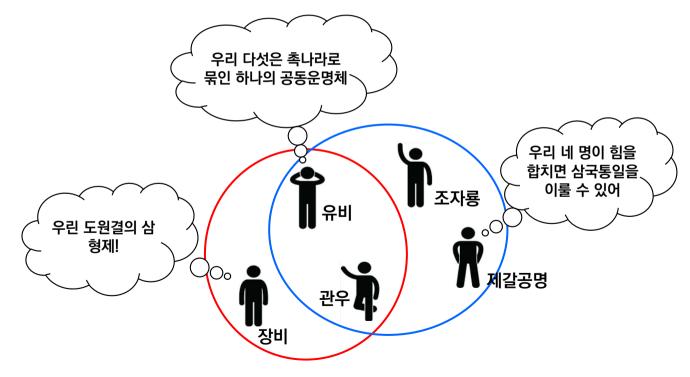
3 cliques found.

1. A B E

2. B D E

3. B C D

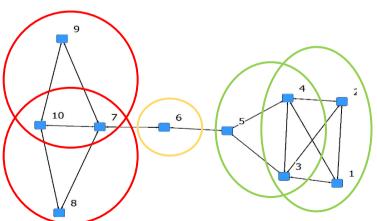
- Comembership
  - 클릭들은 노드를 공유할 수 있음 ⇒ 노드를 많이 공유하는 클릭들은 하나의 하위집단을 간주할 수 있음

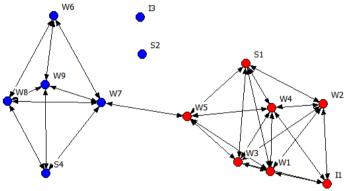


## 상향식 접근 (Clique)

- Comembership에 따라 하위집단을 묶어보자
  - 클릭1과 클릭2는 두 노드 3, 4를 공유 → 묶을 수 있지 않을까?
  - 클릭3과 클릭4는 두 노드 7, 10을 공유 → 묶을 수 있지 않을까?
  - 클릭1과 클릭3은 공유하는 노드 없음 공유 → 묶기는 어려울 것 같은
  - 따라서… 하위집단 1 = {클릭1, 클릭2} = {1, 2, 3, 4, 5}… 하위집단

• Clique 방식은 개념적으로만 알아보고 실습은 하지 않습니다!





#### 4 cliques found.

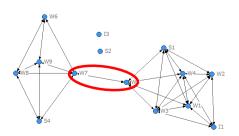
1: 1 2 3 4

2: 3 4 5

3: 7 9 10

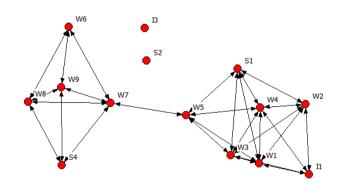
4: 7 8 10

- 하향식 접근의 기본 논리
  - 약한(옅은) 고리를 찾아라 ⇒ 전체 네트워크에서 단단히 묶여 있는 곳(밀도가 높은 지점)과 느슨하게 연결된 곳(밀도가 낮은 곳)
     을 찾아냄 ⇒ 밀도가 낮은 지점이 집단 사이에 분리가 있는 곳
  - 아래 그림에서 옅은 고리(다른 곳보다 밀도가 낮은 지점)는??
  - 앞쪽에 제시된 다양한 네트워크 그림을 보면서 각 네트워크별로 몇 개의 하위집단으로 나뉘는 게 좋을지 생각해 보자.
- Faction 방법: 가장 직관적인 방법으로 전체 네트워크를 분석자가 원하는 수만큼 하위집단으로 나눔
  - 기본 알고리즘: 무작위로 원하는 집단 수만큼 분류 → 노드를 반복적으로 이동 → 하위집단 '내'의 밀도(density)를 최대화하고 하위집단 '간'의 밀도를 최소화 → 반복 멈춤 → final proportion correct 확인 ⇒ 하위집단 수 확정
  - Final proportion correct에 대한 일률적인 기준은 없으나 대략 0.7~0.8 정도면 기준 충족한 것으로 간주할 수 있음
  - Final proportion correct를 참고하여 분석자가 가장 적합하다고 판단하는 하위집단 수를 선택



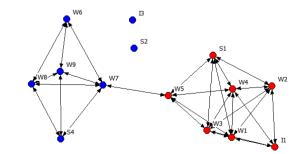
- 다음 예를 가지고 faction 분석의 논리를 이해해 보자
- 아래와 같이 총 14개의 노드가 있는 데이터
- 그림을 확인해 보니 2개로 나누는 것이 좋을 것으로 생각됨 ⇒ 하위집단을 2개 나누어 봄

	I1	13	W1	W2	W3	W4	W5	W6	W7	W8	W9	S1	S2	S4
I1	0	0	1	1	1	1	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W1	1	0	0	1	1	1	1	0	0	0	0	1	0	0
W2	1	0	1	0	1	1	0	0	0	0	0	1	0	0
W3	1	0	1	1	0	1	1	0	0	0	0	1	0	0
W4	1	0	1	1	1	0	1	0	0	0	0	1	0	0
W5	0	0	1	0	1	1	0	0	1	0	0	1	0	0
W6	0	0	0	0	0	0	0	0	1	1	1	0	0	0
W7	0	0	0	0	0	0	1	1	0	1	1	0	0	1
W8	0	0	0	0	0	0	0	1	1	0	1	0	0	1
W9	0	0	0	0	0	0	0	1	1	1	0	0	0	1
S1	0	0	1	1	1	1	1	0	0	0	0	0	0	0
S2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S4	0	0	0	0	0	0	0	0	1	1	1	0	0	0



#### **FACTIONS** Number of factions: Measure of fit: Hamming Input dataset: WIRING ( Initial proportion correct: : 0.5714 1. # of errors = 32 2. # of errors = 323. # of errors = 32Final proportion "correct" : 0.8242 Group Assignments: 1: T1 W1 W2 W3 W4 W5 S1 2: I3 W6 W7 W8 W9 S2 S4 Grouped Adjacency Matrix 5 W3 | 1 1 1 1 1 1 6 W4 | 1 1 1 1 1 1 1 12 S1 | 11111 2 I3 8 W6 1 1 1 10 W8 11 W9 9 W7 13 52 14 S4 | 1 1 1

- UCINET이 반복작업을 하면서 2개의 하위집단으로 나누는 최적화된 방법을 찾음 ⇒ Group Assignments
- Grouped Adjacency Matrix는 Group Assignments에 따라 노드를 재배열한 Matrix
- Final Proportion Correct = 완벽하게 집단이 나누어졌을 때와 비교할 때, 집단 분류가 얼마나 정확한가?
  - 만약 완벽하게 집단이 나누어진다면 왼쪽 매트릭스와 같아야 함 (100% correct)
  - 그러나 실제 나누어진 집단은 왼쪽 매트릭스와 같음
  - 총 182개의 셀 중에 맞는 셀은 150개 (1번 블록: 36개, 2번과 3번 블록 각각: 48개, 4번 블록: 18개)
  - 따라서 final proportion correct =  $\frac{150}{182}$  = 0.8242

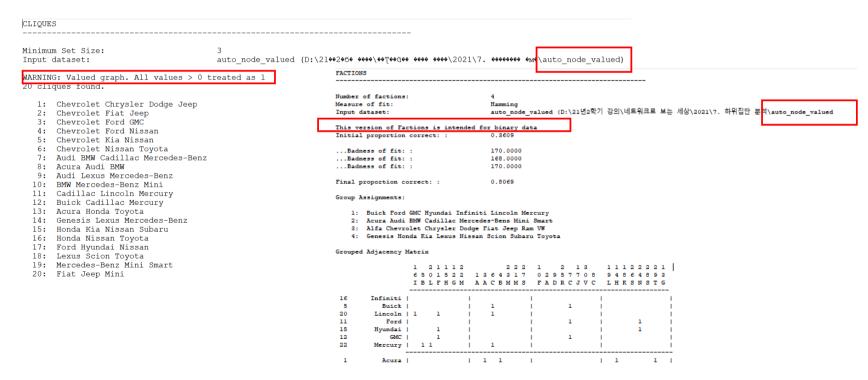


	11	W3	W1	W2	W4	W5	S1	13	W6	W8	W9	W7	52	54
11		1	1	1	1	1	1	0	0	0	0	0	0	0
W3	1		1	1	1	1	1	0	0	0	0	0	0	0
W1	1	1		1	1	1	1	0	0	0	0	0	0	0
W2	1	1	1		1	1	1	0	0	0	0	0	0	0
W4	1	1	1	1		1	1	0	0	0	0	0	0	0
W5	1	1	1	1	1		1	0	0	0	0	0	0	0
S1	1	1	1	1	1	1		0	0	0	0	0	0	0
13	0	0	0	0	0	0	0		1	1	1	1	1	1
W6	0	0	0	0	0	0	0	1		1	1	1	1	1
W8	0	0	0	0	0	0	0	1	1		1	1	1	1
W9	0	0	0	0	0	0	0	1	1	1		1	1	1
W7	0	0	0	0	0	0	0	1	1	1	1		1	1
52	0	0	0	0	0	0	0	1	1	1	1	1		1
54	0	0	0	0	0	0	0	1	1	1	1	1	1	

## 4 기티

#### Valued Data

- Valued data는 일반적으로 subgroup 분석에 적합하지 않음
- Valued data를 그대로 쓰는 경우 → UCINET에서 자동적으로 binary data로 변환 (0보다 큰 값은 자동적으로 1로 dichotomize)
- 따라서 (1) UCINET의 dichotomize 방식을 그대로 따르거나, (2) 그렇지 않을 경우 적절한 수준에서 사전에 dichotomize한 데이터를 활용



## 실습 (Faction)

#### Network → Subgroups → Factions → Input dataset → Number of blocks (원하는 수 입력) → OK

#### rsion 6.732 [32-Bit] Network Visualize Options Help Dyadic measures Whole networks & cohesion Regions e for Social Network Analysis Subgroups Cliques ... N-Cliques ... Ego Networks N-Clan ... Centrality K-plexes Group Centrality K-Plex (legacy) KeyPlayer Core/Periphery Lambda Set Roles & Positions Factions Girvan-Newman Triad Census Louvain method

Factions		×
Parameters Additional		
Input dataset:	WIRING.##H 2 facsets2	
Output sets dataset: Output partition dataset:	WIRING-fac2	
	✓ OK X Cancel ?	<u>H</u> elp

## 실습 (Faction)

- Data: majorminor.xlsx (4년제 대학생의 전공-복수전공 선택 네트워크)
- 3개의 시트가 있습니다. 모두 확인해 봅시다.

id	e-비즈니스	가정관리흐	간호학	건축공학	건축디자인	건축학	건축₹
e-비즈니스	0	0	0	0	0	0	
가정관리흐	0	0	0	0	0	0	
간호학	0	0	0	0	0	0	
건축공학	0	0	0	0	0	0	
건축디자인	0	0	0	0	0	0	
건축학	0	0	0	0	0.540541	0	0.54
건축환경디	0	0	0	0	0	0	
건축환경설	0	0	0	0	0	0	
게임공학	0	0	0	0	0	0	
경영학	0.217391	0.434783	0	0	0	0	
경제학	0	0.359712	0	0	0	0	
경찰법행정	0	0	0	0	0	0	
고분자공학	0	0	0	0	0	0	
공연영상	0	0	0	0	0	0	
공예	0	0	0	0	0	0	
공학교육	0	0	0	0	0	0	
과학교육	0	0	0	0	0	0	
관광외식경	0	0	0	0	0	0	
관광학	0	0	0	0	0	0	
광고/홍보학	0	0	0	0	0	0	
광학공학	0	0	0	0	0	0	
교육학	0	0	0	0	0	0	
그토고하	najorminor_	net mai	orminor_att	t descrip	tion   G	+)	
	,0111111101_	ct IIIaj	omor_att	descrip			

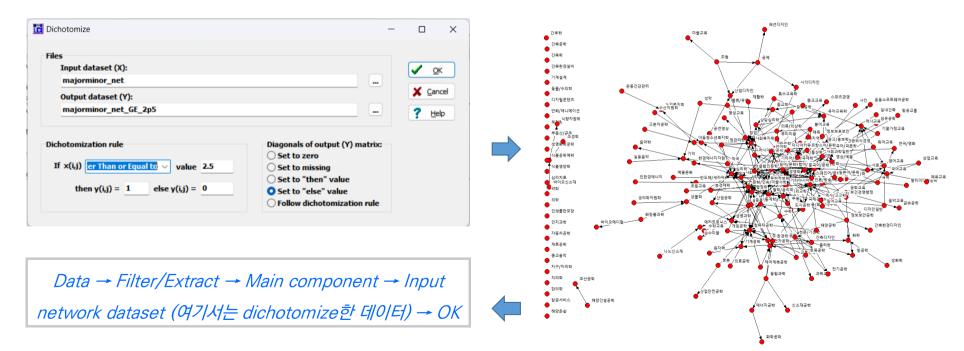
id	college	college_n	size	outdeg	indeg	coreperi
e-비즈니스	사회계열	3	14	4	0	2
가정관리학	자연계열	7	71	3	0	2
간호학	의약계열	5	143	0	0	2
건축공학	공학계열	1	122	0	0	2
건축디자인	공학계열	1	4	2	1	2
건축학	공학계열	1	185	0	0	2
건축환경디자인	예체능계열	4	42	0	1	2
건축환경설비	공학계열	1	21	0	0	2
게임공학	공학계열	1	34	3	0	2
경영학	사회계열	3	460	0	51	1
경제학	사회계열	3	278	1	18	1
경찰법행정	사회계열	3	83	0	2	2
고분자공학	공학계열	1	24	1	0	2
공연영상	예체능계열	4	12	2	0	2
공예	예체능계열	4	56	3	1	2
공학교육	교육계열	2	37	5	0	2
과학교육	교육계열	2	90	0	1	2
관광외식경영	사회계열	3	116	0	7	2
관광학	사회계열	3	26	8	0	1
광고/홍보학	사회계열	3	112	1	5	2
광학공학	공학계열	1	13	2	0	2
교육학	교육계열	2	96	6	4	2
コミコお majorminor_ne	ੁਨ⊦ਅਰ et majorminor a	tt descri	20	(+)		2

자료출처	GOMS (대졸자 직업이동조사) 설문조사에서 주, 복수, 부, 연계전공과 관련된 문항 사용 (4년제)
네트워크자료	GOMS에 제시된 전공코드는 120여개였으나 182개로 확대 182개 전공 간의 주전공, 복수(부)전공 흐름 row 전공 학생 중 column 전공을 복수전공한 학생의
속성자료	college_n: 계열 (공학 = 1, 교육 = 2, 사회= 3, 예체 능 = 4, 의약 = 5, 인문 = 6, 자연 = 7) size: 전공 규모 (설문에 포함된 학생의 수)

## 5

### 실습 (Faction)

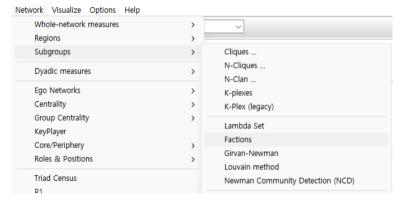
- 분석을 위한 준비
  - 먼저 네트워크 데이터를 불러들인 뒤, 2.5% cut-off 를 적용하여 dichotomize해 보자
  - 그 다음 dichotomize한 데이터를 시각화해 보자
  - Isolate와 small component가 많음 ⇒ 삭제하고 가장 큰 콤포넌트(main component)만 분석에 사용해 보자 (149×149 matrix)

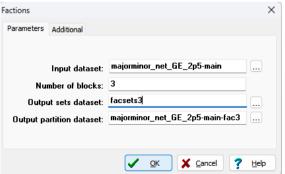


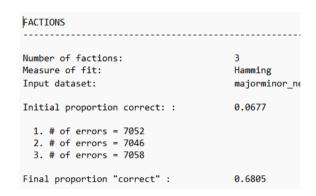
# 5 실습 (Faction)

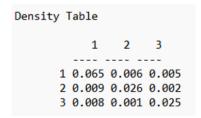
• 먼저 Faction을 이용하여 3개 집단으로 나누어보자 ⇒ FPC (final proportion correct)와 density table을 확인

#### Network → Subgroups → Factions → Input dataset → Number of blocks → OK









# 5 실습 (Faction)

• 같은 방식으로 4개, 5개, 6개 하위집단으로 나누어 보자 ⇒ FPC (final proportion correct)와 density table을 확인

#### Network → Subgroups → Factions → Input dataset → Number of blocks → OK

FACTIONS	
Number of factions: Measure of fit: Input dataset:	4 Hamming majorminor
Initial proportion correct: :	0.5055
1. # of errors = 5282 2. # of errors = 5264 3. # of errors = 5284	
Final proportion "correct" :	0.7613
Group Assignments:	

FACTIONS	
Number of factions: Measure of fit: Input dataset:	5 Hamming majorminor
Initial proportion correct: :	0.5103
1. # of errors = 4176 2. # of errors = 4178 3. # of errors = 4170	
Final proportion "correct" :	0.8109

FACTIONS	
Number of factions: Measure of fit:	6 Hamming
Input dataset: 하위집단\majorminor_net_GE_2p5-main	majormino
Initial proportion correct: :	0.5530
1. # of errors = 3490	
2. # of errors = 3452	
3. # of errors = 3464	
Final proportion "correct" :	0.8435

Density Table											
	1	2	3	4							
1	0.025	0.015	0.002	0.003							
2	0.008	0.101	0.004	0.008							
3	0.004	0.016	0.024	0.002							
4	0.004	0.009	0.002	0.029							

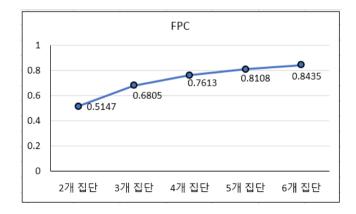
Density Table											
1	2	3	4	5							
1 0.126	0.013	0.004	0.004	0.004							
2 0.014	0.042	0.001	0.001	0.003							
3 0.021	0.004	0.027	0.003	0							
4 0.019	0.003	0.007	0.041	0.002							
5 0.016	0.003	0.001	0.003	0.038							

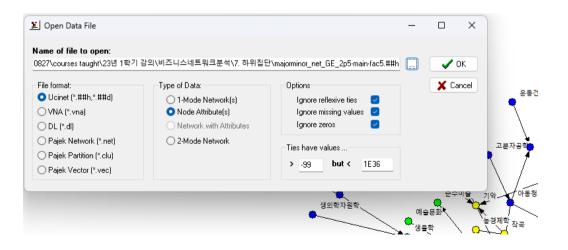
Density Table												
	1	2	3	4	5	6						
1	0.031	0	0.020	0.003	0.002	0						
2	0.003	0.027	0.028	0.007	0.007	0						
3	0.009	0.008	0.152	0.013	0.007	0.008						
4	0	0	0.013	0.047	0.002	0.005						
5	0.005	0.002	0.021	0.005	0.060	0.003						
6	0.002	0	0.022	0.002	0.002	0.054						

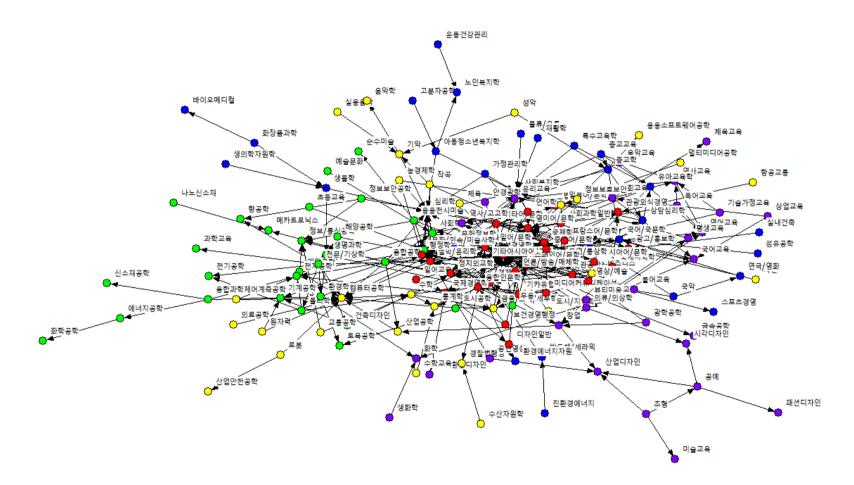
## 5

## 실습 (Faction)

- FPC를 확인해 보고, 몇 개의 집단으로 나누는 것이 좋을지 결정해 보자.
- 저는 5개 집단이 좋아 보입니다 (4개도 괜찮아 보입니다)
- 5개 집단으로 나눈 것을 시각화해 보자
- NetDraw로 네트워크 데이터(이름을 바꾸지 않았다면 majorminor\_net\_GE\_2p5-main.##h)를 열고, 하위집단을 나누어 놓은 데이터(이름을 바꾸지 않았다면 majorminor\_net\_GE\_2p5-main-fac5.##h)를 속성데이터로 불러들인 뒤 그림에 표시

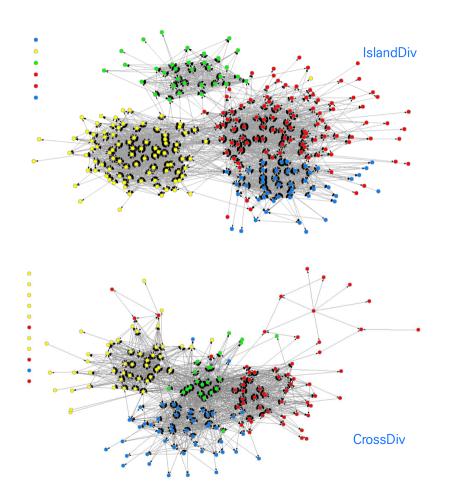


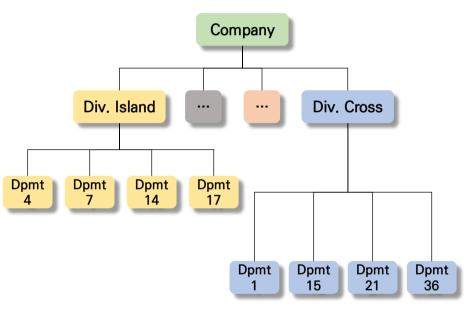




### 과제 2: 데이터 설명

- HR Analyst인 송씨는 직원들의 커뮤니케이션 네트워크 구조가 회사의 성과에 커다란 영향을 미친다고 생각하고 있습니다
- 그래서 어떤 회사에 있는 사업부 중 성과가 가장 높은 사업부와 성과가 가장 낮은 사업부를 뽑아 각 사업부 내 직원들의 이메일 커뮤니케이션 패턴을 분석해 보기로 했습니다. 사업부는 다음과 같습니다
  - Island Division: 4개 부서에 직원수 280명으로 구성; 성과가 가장 낮은 사업부
  - Cross Division: 4개 부서에 직원수 190명으로 구성; 성과가 가장 높은 사업부
- Company.xlsx: 직원들의 이메일 송수신 데이터이며 사업부별로 별도의 시트에 저장
  - IslandDiv: Island Division 직원들(직원수 = 280명)이 이메일을 주고받은 기록
  - CrossDiv: Cross Division 직원들(직원수 = 190명)이 이메일을 주고받은 기록
  - 두 데이터 시트 모두 edge list 형태로 기록 (첫번째 열 = 이메일을 보낸 사람의 ID, 두번째 열 = 이메일을 받은 사람의 ID)
  - IslandDiv\_att & CrossDiv\_att: 두 사업부 직원의 속성 데이터 (각 직원의 부서코드 사업부마다 4개 부서)





## 6 과제 2: 연습문제

- 1. 두 개 사업부의 직원 커뮤니케이션 네트워크 그림을 그려보세요. 속성자료(부서)도 활용해 보세요. 앞 쪽의 그림과 비슷한지 확인해 보십시오. 이 문제는 확인만 하고 제출하지 않아도 됩니다.
- 2. 두 사업부의 전체 커뮤니케이션 네트워크 구조를 비교해 보고자 합니다. 관련된 분석을 한 후 분석 결과를 토대로 다음 쪽에 있는 표를 채우십시오. 응집도, 집중도, 호혜성, 이행성 중에서 두 사업부 간 가장 커다란 차이를 보이는 것은 무엇일까요? 그 차이가 무엇을 의미하는 것인지 짧게 적어주십시오.
- 3. 각 사업부별로 부서 내/간 커뮤니케이션 패턴에 차이가 있는지를 알아보고자 합니다. UCINET을 활용하여 적절한 분석을 수행한 후 다음 에 답하십시오.
  - ① 두 사업부의 density matrix를 캡처해서 제시하십시오. Island 사업부에서 부서내 커뮤니케이션이 가장 활발한 부서는 어떤 부서이고 그 부서의 densit 는 얼마입니까? Cross 사업부에서 어느 부서와 어느 부서 간의 커뮤니케이션이 가장 활발합니까? 왜 그렇게 생각하는지 근거를 들어 설명하십시오.
  - ② 두 사업부의 E-I index를 구한 후 다음 쪽에 나와 있는 것처럼 전체, 그리고 부서별 E-I index를 정리하십시오. 두 사업부 중 어떤 사업부가 부서내 커뮤니케이션 패턴이 강할까요? E-I index를 가지고 답하십시오.
  - ③ Cross 사업부에서 부서간 커뮤니케이션을 많이 하는 부서는 어떤 부서인가요? 왜 그렇게 생각하는지 E-I index를 가지고 답하십시오.
- 4. Cross Division을 Faction 방법에 따라 4개의 집단으로 나누어 본 후 다음에 답하십시오.
  - ① Final Proportion Correct 값은 얼마입니까?
  - ② Cross 사업부의 직업 커뮤니케이션 네트워크 그림에 Faction 방법에 따라 나누어진 4개의 집단을 표시한 후 그림을 제시하십시오. 1번에서 그렸던 그림과 어떤 차이가 있는지 간단히 적어주십시오.

2번 표

#### 3번 표 (예시)

네트워크의 구조적 특성	IslandDiv	CrossDiv
Density		
Average degree		
Connectedness		
Out-Centralization		
In-Centralization		
Arc Reciprocity		
Transitivity		

Island Divisi	on	Cross Division	
Expected E-I Index		Expected E-I Index	
Whole E-I Index		Whole E-I Index	
G1 E-I Index		G1 E-I Index	
G2 E-I Index		G2 E-I Index	
G3 E-I Index		G3 E-I Index	
G4 E-I Index		G4 E-I Index	

ı