

9. Ego-Network: 이웃네트워크와 구조적 공백

- 지금까지 네트워크 전체 수준을 분석하는 방법
 - 시각화: 네트워크 구조를 한 눈에 알아보는 방법
 - 네트워크 구조 분석: 응집성, 호혜성, 이행성, 집중도, 유유상종 등
- 네트워크를 구성하는 하위집단에 대한 분석 방법
 - 하위집단 분석: 네트워크는 어떻게 하위집단으로 나누어지는가
- 노드의 네트워크 특성과 관련된 분석 방법
 - 중심성: 연결중심성, 매개중심성, 근접중심성 등
 - Ego-network 분석: 국지 네트워크 (혹은 이웃네트워크)의 연결 특성 ⇒ 이번 주제

네트워크 수준

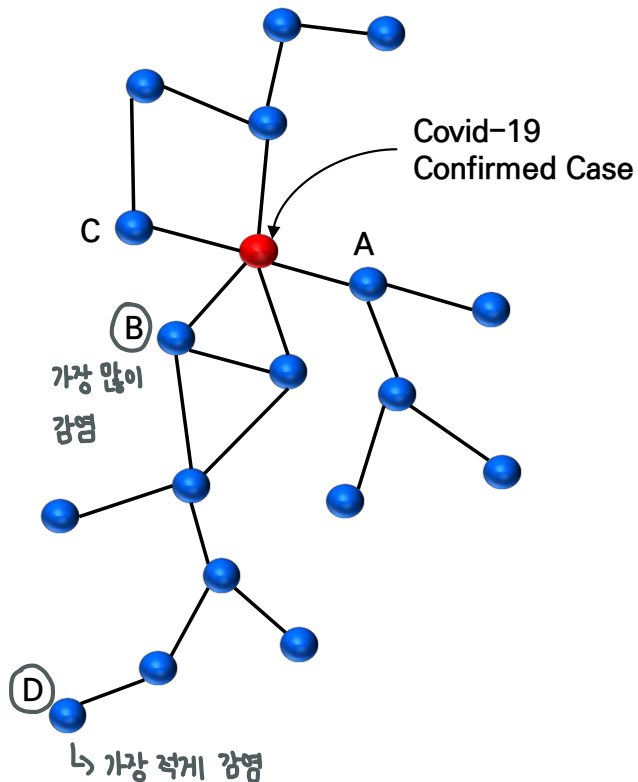
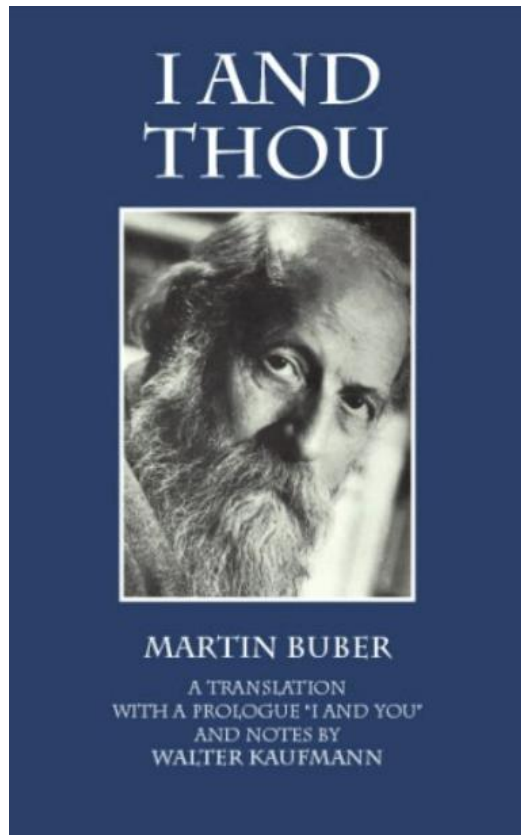
- 시각화
- 구조분석 (응집성 등)

(하위)집단 수준

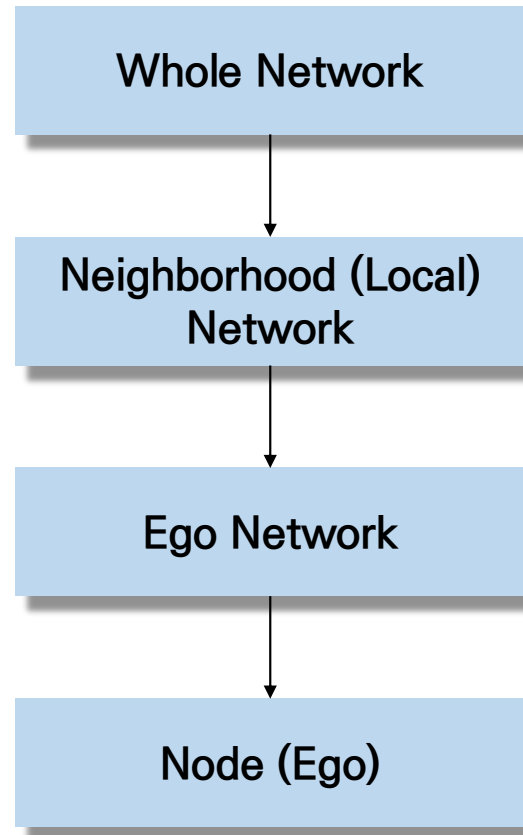
- 하위집단 분석
(Clique, Faction)

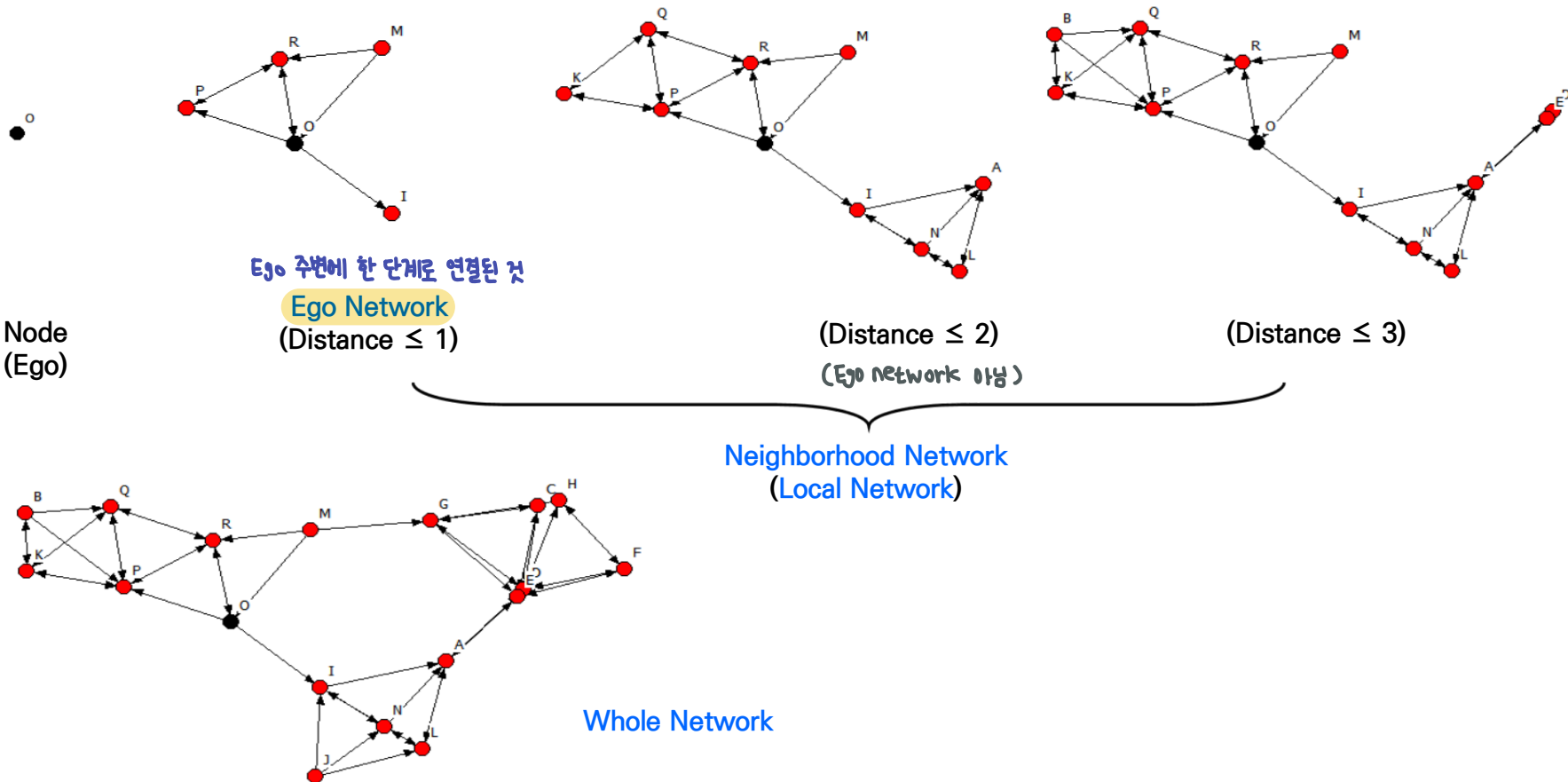
노드 수준

- 중심성 분석
- 에고네트워크 분석



Who is most (least)
susceptible to Covid-19?





- 기본 용어

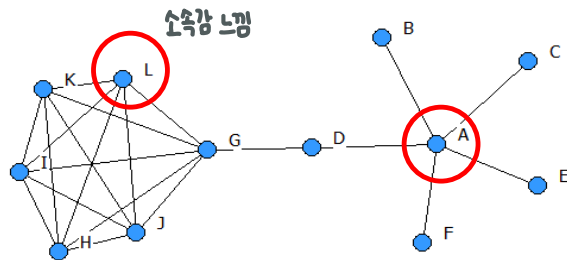
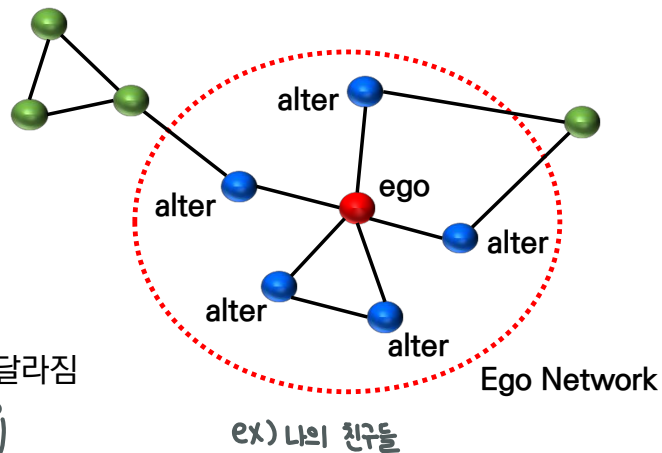
- **에고(ego) & 알터(alter)** ⇒ **에고 네트워크**: ego와 alter들로만 형성된 네트워크

- 왜 ego network가 중요한가?

- 멀리 떨어진 노드보다는 가까이 있는 노드에 직접적인 영향을 받음
- 주변의 노드들이 어떻게 구조화되어 있는가에 따라 ego의 행동, 성과, 특성 등이 달라짐

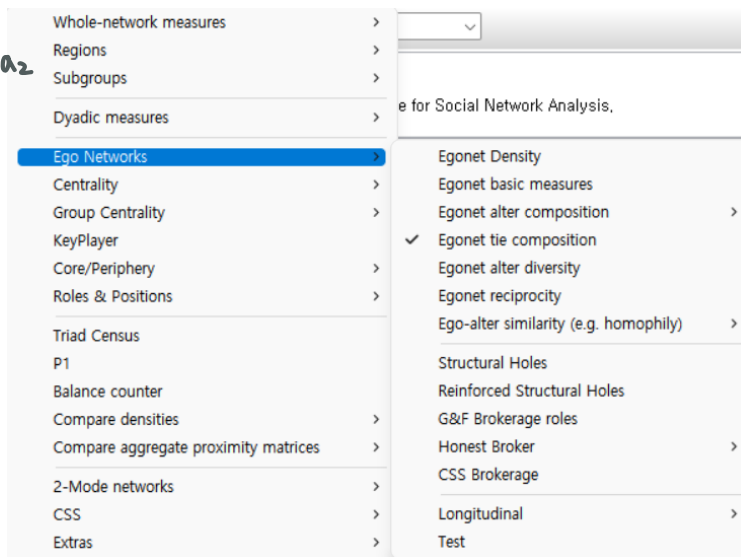
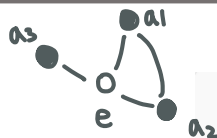
- 그림에서 A와 L의 ego network를 비교해 보자

- 행동의 자율성이 높은 노드는? 행동의 제약이나 구속을 많이 받는 노드는?
- Alter들을 잘 통제할 수 있는 노드는? Alter들로부터 감시와 압력을 많이 받는 노드는?
- alter의 정보를 혼자 통제하거나 자원을 독점할 수 있는 노드는?
- A~L 중에서 가장 적은 비용으로 가장 많은 이득을 누릴 수 있는 노드는?



• Ego-Network 분석

- ① Ego-network의 기본구조: density : ego network 어떻게 생겼나?
- ② Ego-Network에서 alter의 특성: composition, diversity
: 친구들 3총사? or 따로 따로 되어있나?
- ③ Ego-Network에서 Ego-Alter 관계의 특성: ego-alter
: 너하고 너 친구는 언제?
reciprocity, ego-alter similarity (homophily)
- ④ Structural Holes: effective size, efficiency, constraints,
ego-betweenness, N. of structural holes : 구조적 공백
- ⑤ Brokerage Roles ⇒ 이 수업에서는 다루지 않음



• Ego-Network 분석의 활용

- ① 중요한 노드의 발견 + 노드 특성 기술
- ② 노드의 ego-network 속성으로 다른 무엇을 설명하는 데 활용 (성과, 성적, 평판, diffusion of innovation 등)



어떤 사람이 더 정보획득에 유리?

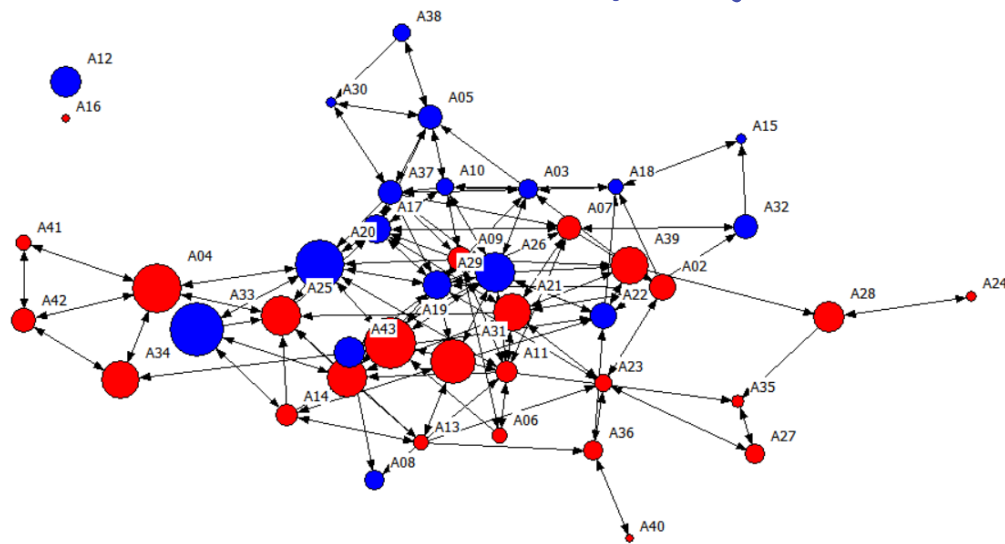
- Data: [study.xlsx](#)

- 어느 대학 통계학 전공필수 과목을 수강한 1학년 학생 43명 학생의 중간/기말고사 학습 네트워크 + 중간/기말고사 성적
- “중간(기말)고사를 준비하는 동안 누구하고 공부를 했는지 혹은 모르는 내용을 누구에게 물어보았는지 이름을 모두 적어주십시오.”

→ Matrix

- 여기서는 기말고사 자료를 활용 ⇒ study_final과 study_att를 UCINET으로 불러들이자

↳ Edgelist 1 ego alter



기말고사 이전 학습 네트워크

- 파란색 = 여성
- 빨간색 = 남성
- 노드크기 = ~~기말고사 성적~~

중간

① Size: ego와 연결된 alter의 수

- Degree centrality와 같은 개념 \Rightarrow 에고네트워크가 얼마나 큰가? 얼마나 많은 alter와 연결되어 있는가?
(연결 중심성)
- incoming, outgoing, both로 나누어 살펴볼 수 있음
- A1의 경우, incoming = 5, outgoing = 4, both = 7

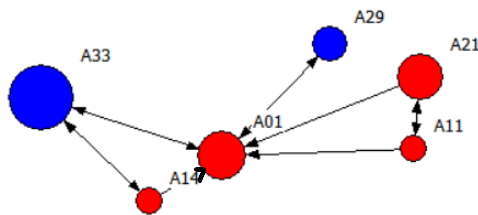
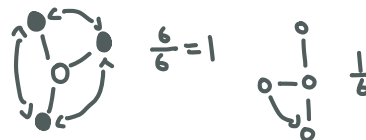
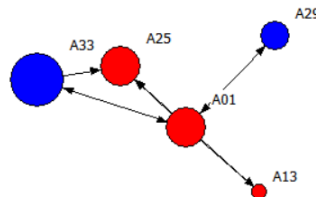
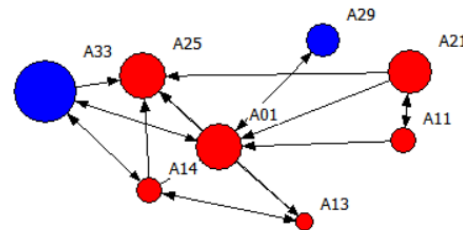
② Average Degree

- Ego가 가지고 있는 alter들 간에 형성된 연결의 평균 degree
- A1의 경우, incoming = 0.800 (4/5), outgoing = 0.75 (3/4), both = 1.571 $((1+1+3+1+3+0+2)/7)$

③ Density: alter간의 연결 밀도

- Ego의 alter들이 얼마나 긴밀히 연결되어 있는지 (degree of bonding)

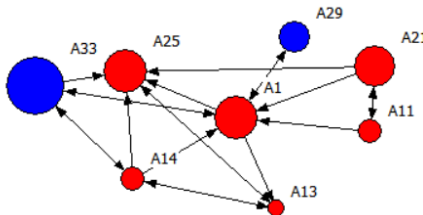
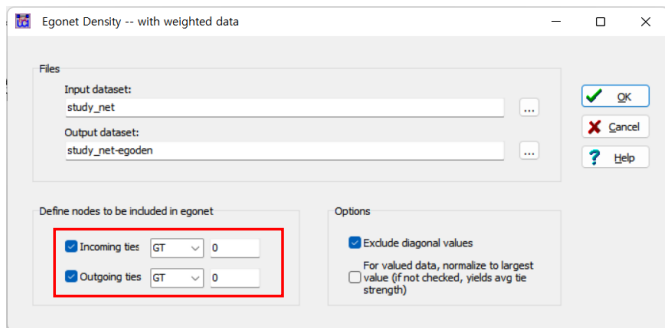
- $\text{density} = \frac{\text{Ties}}{\text{Pairs}} = \frac{n \times \text{Avg.Deg}}{n^2}$ (A1의 경우, both = $\frac{11}{42} = .262$)

A01의 에고네트워크
(incoming)A01의 에고네트워크
(outgoing)A01의 에고네트워크
(both)

- Incoming, outgoing, both로 나누어 구할 수 있음
 - Incoming을 구하는 경우: Incoming ties에만 체크
 - Outgoing을 구하는 경우: outgoing ties에만 체크
 - Incoming과 outgoing을 모두 고려하는 경우: 모두 체크

*Network → Ego Networks → Egonet Density →
Input dataset → OK*

이 주변에 있는 애들 사이에 생길 수 있는 링크의 26.2% 있음



		1 Size	Avg De	2 Density	3 Density
1	A1	7.000	1.571	0.262	
2	A2	5.000	1.400	0.350	
3	A3	7.000	2.143	0.357	
4	A4	5.000	1.000	0.250	
5	A5	6.000	1.833	0.367	
6	A6	3.000	1.333	0.667	
7	A7	7.000	1.714	0.286	
8	A8	2.000	0.000	0.000	
9	A9	9.000	2.778	0.347	
10	A10	8.000	2.875	0.411	
11	A11	10.000	1.100	0.122	
12	A12				
13	A13	6.000	1.167	0.233	
14	A14	5.000	1.800	0.450	
15	A15	2.000	0.000	0.000	
16	A16				
17	A17	9.000	2.667	0.333	
18	A18	5.000	0.400	0.100	
19	A19	5.000	2.000	0.500	
20	A20	10.000	2.400	0.267	
21	A21	8.000	1.250	0.179	
22	A22	8.000	1.000	0.143	
23	A23	6.000	0.667	0.133	
24	A24	1.000	0.000	0.000	
25	A25	8.000	2.125	0.304	
26	A26	12.000	3.250	0.295	
27	A27	2.000	0.000	0.000	
28	A28	3.000	0.000	0.000	
29	A29	13.000	2.308	0.192	
30	A30	3.000	1.333	0.667	
31	A31	5.000	1.200	0.300	
32	A32	3.000	0.000	0.000	

Both incoming and outgoing

		1 Size	Avg De	2 Density	3 Density
1	A1	5.000	0.800	0.200	
2	A2	3.000	1.333	0.667	
3	A3	6.000	1.833	0.367	
4	A4	5.000	1.000	0.250	
5	A5	6.000	1.833	0.367	
6	A6	2.000	1.000	1.000	
7	A7	6.000	1.500	0.300	
8	A8	2.000	0.000	0.000	
9	A9	7.000	2.286	0.381	
10	A10	7.000	1.857	0.310	
11	A11	8.000	1.250	0.179	
12	A12				
13	A13	4.000	1.250	0.417	
14	A14	3.000	0.667	0.333	
15	A15	2.000	0.000	0.000	
16	A16				
17	A17	8.000	1.875	0.268	
18	A18	4.000	0.000	0.000	
19	A19	5.000	2.000	0.500	
20	A20	9.000	2.222	0.278	
21	A21	6.000	1.167	0.233	
22	A22	7.000	1.143	0.190	
23	A23	6.000	0.667	0.133	
24	A24	1.000	0.000	0.000	
25	A25	7.000	1.857	0.310	
26	A26	10.000	2.400	0.267	
27	A27	2.000	0.000	0.000	
28	A28	2.000	0.000	0.000	
29	A29	11.000	2.182	0.218	
30	A30	3.000	1.333	0.667	
31	A31	4.000	1.000	0.333	
32	A32	2.000	0.000	0.000	

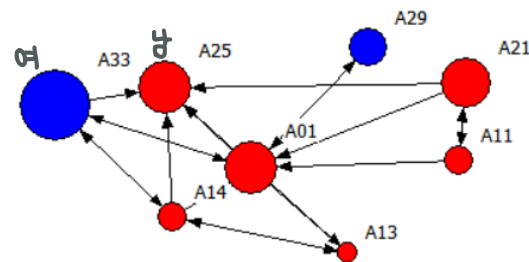
incoming

		1 Size	Avg De	2 Density	3 Density
1	A1	4.000	0.750	0.250	
2	A2	4.000	1.000	0.333	
3	A3	6.000	2.333	0.467	
4	A4	5.000	1.000	0.250	
5	A5	5.000	1.400	0.350	
6	A6	2.000	1.000	1.000	
7	A7	5.000	1.200	0.300	
8	A8	2.000	0.000	0.000	
9	A9	8.000	1.750	0.250	
10	A10	7.000	3.000	0.500	
11	A11	10.000	1.100	0.122	
12	A12				
13	A13	5.000	1.000	0.250	
14	A14	5.000	1.800	0.450	
15	A15	1.000	0.000	0.000	
16	A16				
17	A17	8.000	2.750	0.393	
18	A18	3.000	0.667	0.333	
19	A19	4.000	2.000	0.667	
20	A20	8.000	1.500	0.214	
21	A21	8.000	1.250	0.179	
22	A22	6.000	1.000	0.200	
23	A23	5.000	0.600	0.150	
24	A24	1.000	0.000	0.000	
25	A25	3.000	0.000	0.000	
26	A26	12.000	3.250	0.295	
27	A27	2.000	0.000	0.000	
28	A28	2.000	0.000	0.000	
29	A29	12.000	2.250	0.205	
30	A30	2.000	1.000	1.000	
31	A31	4.000	1.000	0.333	
32	A32	3.000	0.000	0.000	

outgoing

- ego를 둘러싼 alter들은 어떤 특성을 가지고 있는가?
- **범주형 속성**인 경우
 - 특정 범주의 비율, 범주의 이질성 (다양한 범주가 섞여 있는 정도)
 - 예: 여성 alter의 비율은? 남성 alter의 비율은? 성에 따른 이질성 (다양성)은?
 - A01의 경우 7명 중 여성(파란색)은 2명, 남성(붉은색)은 5명 \Rightarrow 여성의 비율은 0.286, 남성의 비율은 0.714
- **연속형 속성**인 경우
 - Alter들이 가지고 있는 특정 속성의 평균, 표준편차, 다양성
 - 예: alter들의 평균 성적은? alter들 성적의 이질성 (다양성)은?
 - A01의 경우 기말고사 성적이 높은 사람과 낮은 사람이 골고루 섞여 있음

ex) 내 주변에 있는 사람들의 성적의 분포

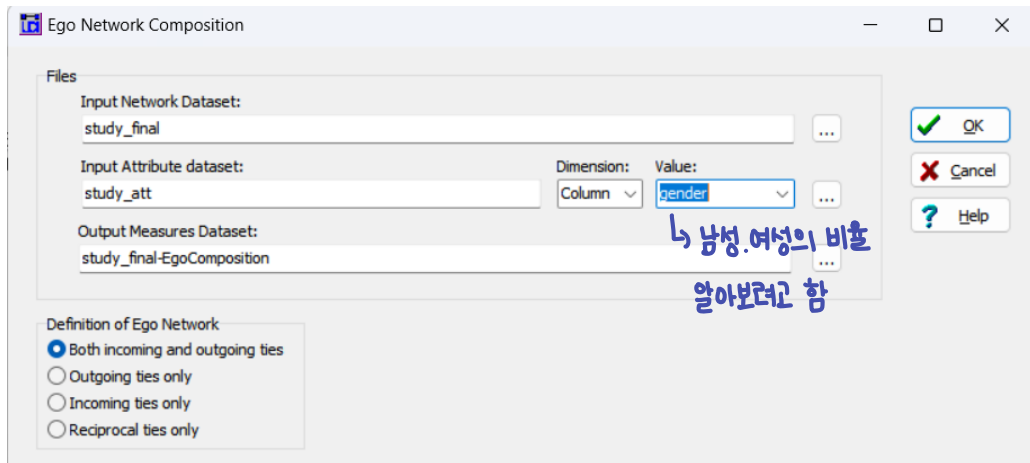


A01 주변 남성 비율 = $\frac{5}{7}$
 " 여성 " = $\frac{2}{7}$

- 범주형 속성인 경우

- 여기서는 alter들의 남성/여성의 비율이 각각 어느 정도인지 알아보자.
- 참고: Heterogeneity = $1 - \sum_{i=1}^n p_i^2$ (A01의 경우 $1 - 0.286^2 - 0.714^2 = 0.408$)
- 참고: IQV (Index of Qualitative Variation) = $\frac{K(1 - \sum p_i^2)}{(K-1)}$ ($K = \text{number of categories}$) (A01의 경우 $\frac{2 \times 0.408}{1} = 0.816$)

Network → Ego Networks → Egonet alter composition → Categorical alter attributes → OK

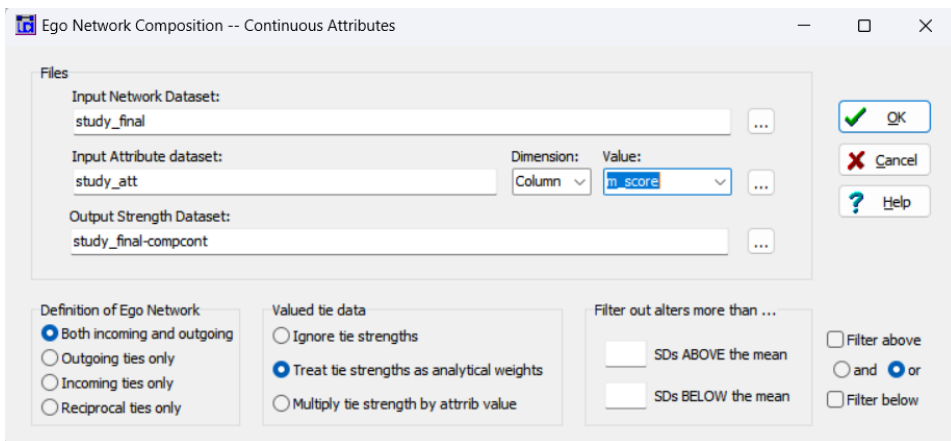


Ego Net Composition

		1	2	3	4	5	6	7
		gende	f1	f2	p1	p2	Heter	IQV
1	A01	2.000	2.000	5.000	0.286	0.714	0.408	0.816
2	A02	2.000	2.000	3.000	0.400	0.600	0.480	0.960
3	A03	1.000	6.000	1.000	0.857	0.143	0.245	0.490
4	A04	2.000	1.000	4.000	0.200	0.800	0.320	0.640
5	A05	1.000	6.000	0.000	1.000	0.000	0.000	0.000
6	A06	2.000	0.000	3.000	0.000	1.000	0.000	0.000
7	A07	2.000	4.000	3.000	0.571	0.429	0.490	0.980
8	A08	1.000	1.000	1.000	0.500	0.500	0.500	1.000
9	A09	2.000	6.000	3.000	0.667	0.333	0.444	0.889
10	A10	1.000	7.000	1.000	0.875	0.125	0.219	0.438
11	A11	2.000	3.000	7.000	0.300	0.700	0.420	0.840
12	A12	1.000	0.000	0.000				
13	A13	2.000	0.000	6.000	0.000	1.000	0.000	0.000
14	A14	2.000	1.000	4.000	0.200	0.800	0.320	0.640
15	A15	1.000	2.000	0.000	1.000	0.000	0.000	0.000
16	A16	2.000	0.000	0.000				

- 연속형 속성인 경우
 - 여기서는 alter의 중간고사 성적의 특성을 알아보자
 - Avg: alter의 평균 (여기서는 alter들의 중간고사 성적 평균)
 - StdDev: alter의 표준편차 (여기서는 alter들의 중간고사 성적 표준편차)
 - 참고: CV (Coefficient of Variation) = $\frac{\sigma}{\mu} = \frac{Std}{Avg} \Rightarrow$ 클수록 이질적

Network → Ego Networks → Egonet alter composition → Continuous alter attributes → OK



A 주변에 있는 7개의 노드들의 중간고사 평균

Ego Net Composition - Continuous Attribute measures

	1	2	3	4	5	6	7	8	9
	Avg	Sum	Min	Max	StdDev	EstSD	CV	Num	WtdNum
1 A01	63.571	445	36	94	18.391	19.865	0.289	7	7
2 A02	52.600	263	36	70	13.185	14.741	0.251	5	5
3 A03	55.286	387	41	70	11.132	12.024	0.201	7	7
4 A04	61.800	309	33	84	19.813	22.152	0.321	5	5
5 A05	45.500	273	37	56	7.205	7.893	0.158	6	6
6 A06	60	180	50	80	14.142	17.321	0.236	3	3
7 A07	55.571	389	46	70	8.192	8.848	0.147	7	7
8 A08	58	116	50	66	8	11.314	0.138	2	2
9 A09	60	540	43	84	12.347	13.096	0.206	9	9
10 A10	55	440	41	84	13.379	14.303	0.243	8	8
11 A11	58.300	583	34	84	16.273	17.153	0.279	10	10
12 A12								0	0

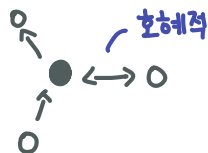
↳ 참고

Std Dev ÷ Avg

크기질성을 나타내는 대표적인 지표

① Reciprocity: ego와 alter가 얼마나 호혜적(쌍방향 관계)인가

$$O \longleftrightarrow O$$



- “나는 내 주변 사람들에게 주기만 한다” vs. “내 주변 사람들은 나에게 받는 만큼 나에게 준다”

- Ego-alter의 관계 쌍 중 호혜적인 쌍이 얼마나 되는가?

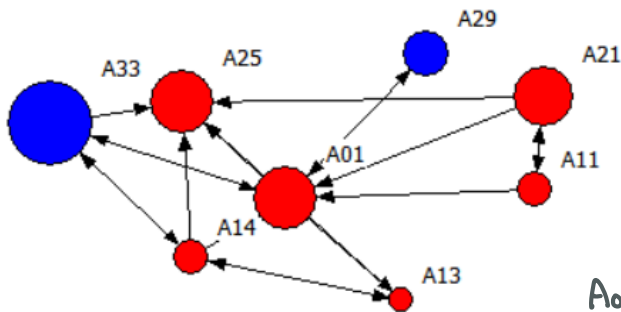
- Dyad Model ($= \frac{R}{R+U}$) or Arc Model ($= \frac{2R}{2R+U}$) 총 3개의 tie
R=1 U=2

$$\longrightarrow \frac{1}{3}$$

$$\longleftrightarrow \frac{2}{4}$$

- A01의 Dyad-based Reciprocity = $\frac{2}{7} = 0.286$ ← 7개의 ego-alter 쌍(dyad) 중에서 2개는 쌍방향이다

- A01의 Arc-based Reciprocity = $\frac{4}{9} = 0.444$ ← ego-alter 사이에 존재하는 9개의 화살표 중 4개는 쌍방향 연결에 속한다



$$A_{01} \text{ Dyad } \frac{2}{7}$$

$$\text{Arch } \frac{4}{9}$$

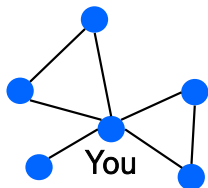
: ego와 alter 같은 집단에 속해 있는지 ex)성 : 0이 같은 성 5명 다른 성 2명 \Rightarrow 나의 alter 중에는 나와 같은 성 많음

② ego-alter similarity: ego의 속성과 alter의 속성이 얼마나 유사한가?

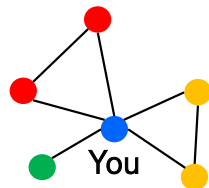
- 알터가 에고와 같은 집단에 속한 경우 vs. 다른 집단에 속한 경우 \Rightarrow 에고가 획득하는 정보나 자원, 행동, 정체성 등이 달라질 수 있음
- 예: 나는 마케팅부서 직원이다. 나는 주로 마케팅 부서 사람들 하고만 친하게 지낸다
- 예: 나는 마케팅부서 직원이다. 나는 부서를 가리지 않고 두루두루 친하게 지낸다
- ego-alter similarity가 높으면 높은 집단 소속감으로 필요한 지원 획득 가능하지만 다양한 자원과 정보를 획득하기는 어려움
- Ego-alter similarity가 낮으면 집단 소속감은 약화될 수 있지만 새로운 정보·자원 획득 가능



HIGH bonding social capital
(trust-based collaboration)
LOW bridging social capital



HIGH Ego-alter similarity

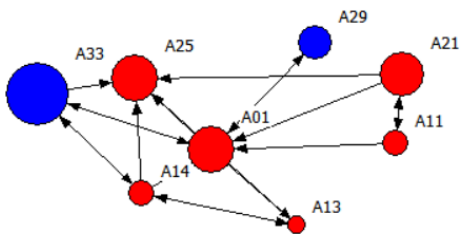
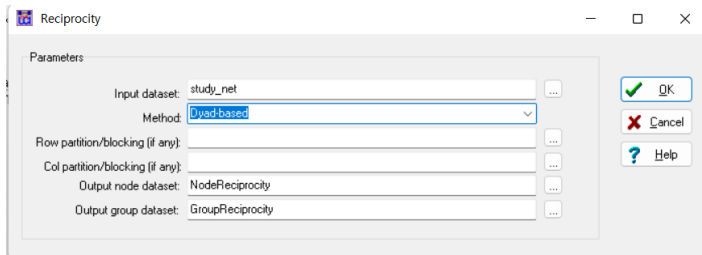


LOW Ego-alter similarity

LOW bonding social capital
HIGH bridging social capital
(information non-redundancy, resource diversity)

- Reciprocity
- dyad-based, arc-based, hybrid 중 선택
- 여기서는 dyad-based와 arc-based만 살펴봄

Network → Ego Networks → Egonet reciprocity →
Input dataset → OK



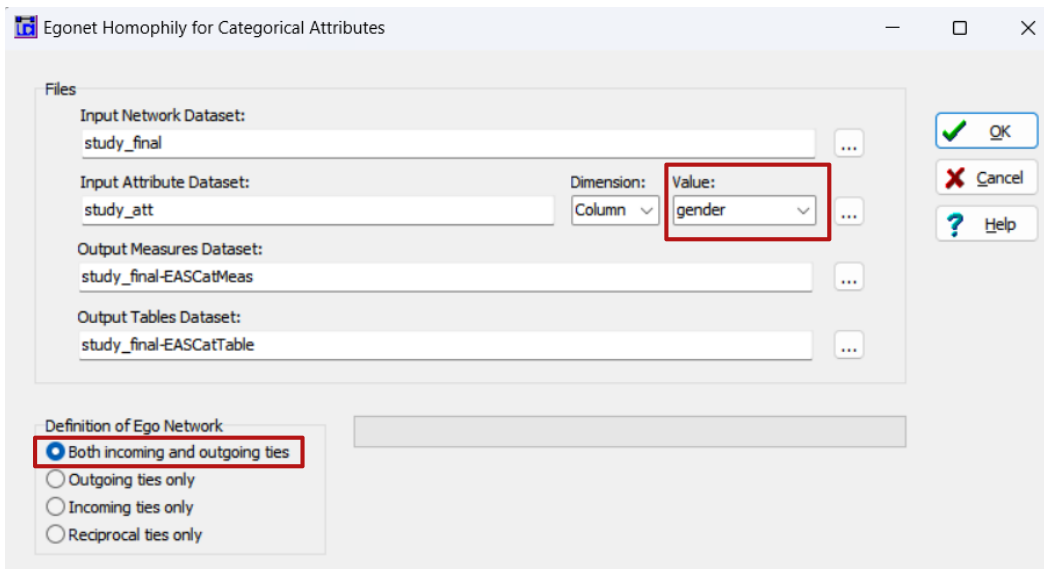
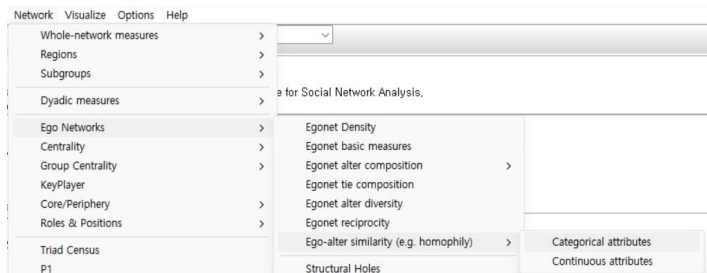
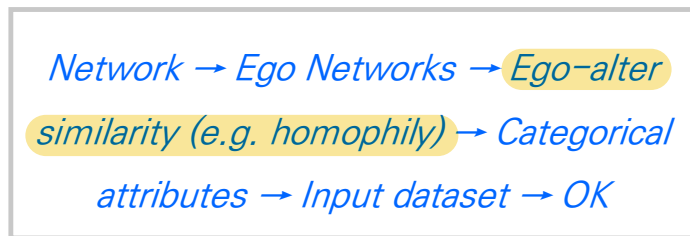
Node-level Reciprocity Statistics -- All values are Proportions

		1 Symmetric	2 Non-Symme	3 Out/NonSym	4 In/NonSym	5 Sym/Out	6 Sym/In
1	A1	0.286	0.714	0.400	0.600	0.500	0.400
2	A2	0.400	0.600	0.667	0.333	0.500	0.667
3	A3	0.714	0.286	0.500	0.500	0.833	0.833
4	A4	1.000	0.000	0.000	1.000	1.000	1.000
5	A5	0.833	0.167	0.000	1.000	1.000	0.833
6	A6	0.333	0.667	0.500	0.500	0.500	0.500
7	A7	0.571	0.429	0.333	0.667	0.800	0.667
8	A8	1.000	0.000	0.000	1.000	1.000	1.000
9	A9	0.667	0.333	0.667	0.333	0.750	0.857
10	A10	0.750	0.250	0.500	0.857	0.857	0.857
11	A11	0.800	0.200	1.000	0.000	0.800	1.000
12	A12						
13	A13	0.500	0.500	0.667	0.333	0.600	0.750
14	A14	0.600	0.400	1.000	0.000	0.600	1.000
15	A15	0.500	0.500	0.000	1.000	1.000	0.500
16	A16						
17	A17	0.778	0.222	0.500	0.500	0.875	0.875
18	A18	0.400	0.600	0.333	0.667	0.667	0.500
19	A19	0.800	0.200	0.000	1.000	1.000	0.800
20	A20	0.700	0.300	0.333	0.667	0.875	0.778
21	A21	0.750	0.250	1.000	0.000	0.750	1.000
22	A22	0.625	0.375	0.333	0.667	0.833	0.714
23	A23	0.833	0.167	0.000	1.000	1.000	0.833
24	A24	1.000	0.000	0.000	1.000	1.000	1.000
25	A25	0.250	0.750	0.167	0.833	0.667	0.286
26	A26	0.833	0.167	1.000	0.000	0.833	1.000
27	A27	1.000	0.000	0.000	1.000	1.000	1.000
28	A28	0.333	0.667	0.500	0.500	0.500	0.500
29	A29	0.769	0.231	0.667	0.333	0.833	0.909
30	A30	0.667	0.333	0.000	1.000	1.000	0.667
31	A31	0.600	0.400	0.500	0.500	0.750	0.750
32	A32	0.667	0.333	1.000	0.000	0.667	1.000

Node-level Reciprocity Statistics (Arc Model) -- All values are Proportions

		1 Reciproca	2 UnRecipro	3 Out/UnRec	4 In/UnRec	5 UnRecip/O	6 UnRecip/I
1	A1	0.444	0.556	0.400	0.600	0.500	0.600
2	A2	0.571	0.429	0.667	0.333	0.500	0.333
3	A3	0.833	0.167	0.500	0.500	0.167	0.167
4	A4	1.000	0.000	0.000	1.000	0.000	0.000
5	A5	0.909	0.091	0.000	1.000	0.000	0.167
6	A6	0.500	0.500	0.500	0.500	0.500	0.500
7	A7	0.727	0.273	0.333	0.667	0.200	0.333
8	A8	1.000	0.000	0.000	1.000	0.000	0.000
9	A9	0.800	0.200	0.667	0.333	0.250	0.143
10	A10	0.857	0.143	0.500	0.500	0.143	0.143
11	A11	0.889	0.111	1.000	0.000	0.200	0.000
12	A12						
13	A13	0.667	0.333	0.667	0.333	0.400	0.250
14	A14	0.750	0.250	1.000	0.000	0.400	0.000
15	A15	0.667	0.333	0.000	1.000	0.000	0.500
16	A16						
17	A17	0.875	0.125	0.500	0.500	0.125	0.125
18	A18	0.571	0.429	0.333	0.667	0.333	0.500
19	A19	0.889	0.111	0.000	1.000	0.000	0.200
20	A20	0.824	0.176	0.000	1.000	0.125	0.222
21	A21	0.857	0.143	1.000	0.000	0.250	0.000
22	A22	0.769	0.231	0.333	0.667	0.167	0.286
23	A23	0.909	0.091	0.000	1.000	0.000	0.167
24	A24	1.000	0.000	0.000	1.000	0.000	0.000
25	A25	0.400	0.600	0.167	0.833	0.333	0.714
26	A26	0.909	0.091	1.000	0.000	0.167	0.000
27	A27	1.000	0.000	0.000	1.000	0.000	0.000
28	A28	0.500	0.500	0.500	0.500	0.500	0.500
29	A29	0.870	0.130	0.667	0.333	0.167	0.091
30	A30	0.800	0.200	0.000	1.000	0.000	0.333
31	A31	0.750	0.250	0.500	0.500	0.250	0.250
32	A32	0.800	0.200	1.000	0.000	0.333	0.000

- Ego-alter similarity (유사성을 측정하려는 속성이 범주형 속성인 경우) ⇒ **Categorical attribute**
- 같은 성과 주로 공부하는 사람은? 다른 성과 주로 공부하는 사람은?



- Ego-alter similarity (유사성을 측정하려는 속성이 범주형 속성인 경우) ⇒ **Categorical attribute**

I. $\text{Pct Same} = \frac{a}{a+b} = \frac{IL}{\text{Ego-Network Size}}$ (0 = perfect heterophily, 1 = perfect homophily)

↳ alter 중 나와 같은 애들이 얼마나 있는가 $\frac{5}{7} = 0.714$

II. $\text{EI-Index} = \frac{b-a}{b+a} = \frac{EL-IL}{EL+IL}$ (-1 = perfect homophily, 1 = perfect heterophily)

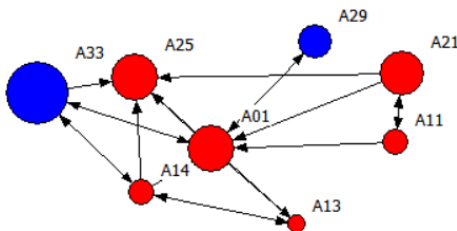
EL=2 IL=5 $\frac{2-5}{2+5} =$

	Within Group	Between Group
Tie	a	b
No Tie	c	d

- Ego-alter similarity (유사성을 측정하려는 속성이 연속형 속성인 경우) ⇒ Continuous attribute
 - 이 수업에서는 다루지 않음

1
Node-level alter-ego similarity

		1	2	3	4	5	6	7	8
		Pct Sa	EL Ind	Jaccar	Yules	Cohen	Corr/P	fInGro	fOutGr
		me	ex	d	Q	Kappa	hi	up	oup
1	A01	0.714	-0.429	0.192	0.356	0.087	0.129	5	2
2	A02	0.600	-0.200	0.115	0.067	0.012	0.021	3	2
3	A03	0.857	-0.714	0.333	0.858	0.345	0.412	6	1
4	A04	0.800	-0.600	0.160	0.545	0.098	0.170	4	1
5	A05	1	-1	0.353	1	0.394	0.495	6	0
6	A06	1	-1	0.125	1	0.109	0.240	3	0
7	A07	0.429	0.143	0.107	-0.333	-0.087	-0.129	3	4
8	A08	0.500	0	0.056	0.200	0.022	0.043	1	1
9	A09	0.333	0.333	0.100	-0.556	-0.189	-0.251	3	6
10	A10	0.875	-0.750	0.389	0.888	0.406	0.465	7	1
11	A11	0.700	-0.400	0.259	0.346	0.114	0.145	7	3
12	A12							0	0
13	A13	1	-1	0.250	1	0.222	0.354	6	0
14	A14	0.800	-0.600	0.160	0.545	0.098	0.170	4	1
15	A15	1	-1	0.118	1	0.137	0.271	2	0
16	A16							0	0
17	A17	0.667	-0.333	0.300	0.600	0.252	0.279	6	3
18	A18	0.800	-0.600	0.222	0.761	0.220	0.296	4	1
19	A19	0.400	0.200	0.074	0.275	0.074	0.127	2	2



- A01의 경우,

- $\text{Pct Same} = \frac{5}{7} = 0.714$

- $\text{EI-Index} = \frac{2-5}{2+5} = -0.429$

- 구조적 공백의 개념

- Structural Holes: Empty space in social structure (R. Burt)

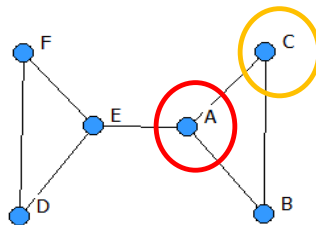
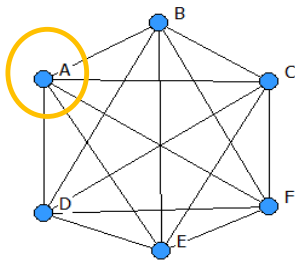
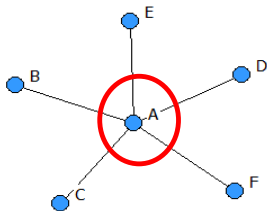
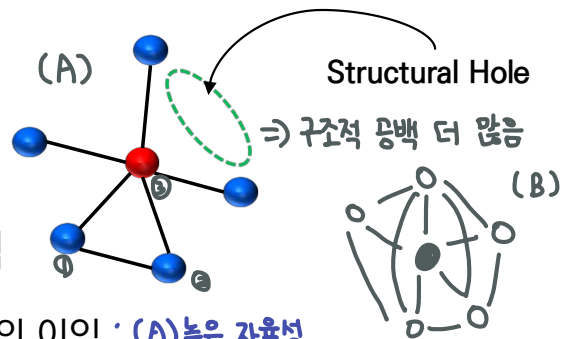
- 왜 구조적 공백이 중요할까? 구조적 공백이 많은 노드가 갖는 두 가지 이점

- ① 높은 구조적 자율성(autonomy): 알터 간의 관계 자율적으로 조정 가능 \Rightarrow 제3자의 이익 : (A) 높은 자율성
 \hookrightarrow ego \Rightarrow ①, ② 사이를 ①이 조절

※ 구조적 자율성이 낮으면 제약성(constraint) 증가: 알터들이 직접 연결되면 행동 제약 증가, 기회 감소
 \hookrightarrow 구조적 자율성 반대개념

- ② 이질적 정보 획득 기회(non-redundancy): 구조적 공백이 많으면 적은 비용으로 비중복적 정보 획득 가능
 \Rightarrow 중복되지 않은 정보

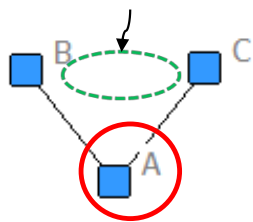
- 구조적 공백과 관련된 값을 해석할 때에는 항상 무슨 네트워크인지 맥락을 감안해야 함



① 구조적 제약성(constraint) \leftrightarrow 구조적 자율성(autonomy)

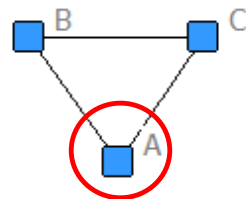
- 일반적으로 네트워크 분석에서 구조적 자율성 대신 **제약성(constraint)**을 활용
- 따라서 제약성은 구조적 자율성이 얼마나 제한되는지를 의미함
- 구조적 제약성은 **구조적 공백이 거의 없는 상태** \Rightarrow alter들을 마음대로 조정하기 힘든 구조적 조건

Structural Hole



A는 높은 자율성

A는 높은 제약성



- A가 없어지면 B와 C는 외톨이
- B와 C는 A에 대한 의존도 높음
- 따라서 A는 자율성이 높고 제약성 낮음

- A가 없어지더라도 B-C 관계 유지
- B와 C는 A에 대한 의존도 낮음
- 따라서 A는 자율성 낮고 제약성 높음

• 구조적 공백 (Ronald Burt)

- 두 alter간에 직접적 연결이 없는 상태
- 즉, 직접 연결이 없는 두 노드를 **예고가 간접적으로 연결**(spanning the structural holes)해 주는 상태

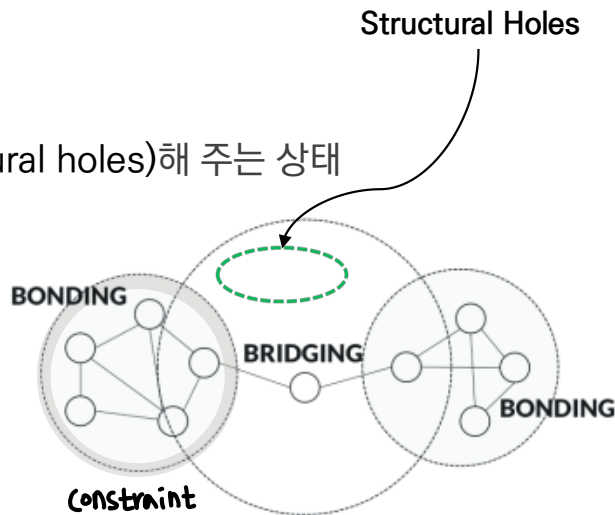
• 사회자본 (Social Capital)

- Benefits from network positions (social relations)
- 내가 누구냐와 상관없이 내가 누구와 연결되어 있느냐에 따라 얻게 되는 이득

• 두 가지 종류의 사회자본

ex) 가족 : 소속감, 정체성

- **Bonding social capital**: 신뢰에 기반한 밀도 있는 관계가 주는 이득 (Structural Holes가 적고 Constraint가 높음)
- **Bridging social capital**: '구조적 공백'을 포착/활용함으로써 얻을 수 있는 이득 (Structural Holes가 많고 Constraint가 낮음)
 ↳ 두 개의 집단을 이어줌으로써 ex) 정보 사서 반대쪽에 팜



- 전자 대기업의 supply chain managers discussion network
 - Supply chain issue에 대한 아이디어가 있다면 적어주세요.
 - 이 supply chain issue를 다른 사람과 토의한 적이 있다면 그 사람 이름을 적어주세요.
- Structural holes argument가 맞다면...
 - 9번 매니저가 205번 매니저보다 bridging social capital을 더 많이 가지고 있음 ⇒ 평판, 연봉, 승진에서 유리

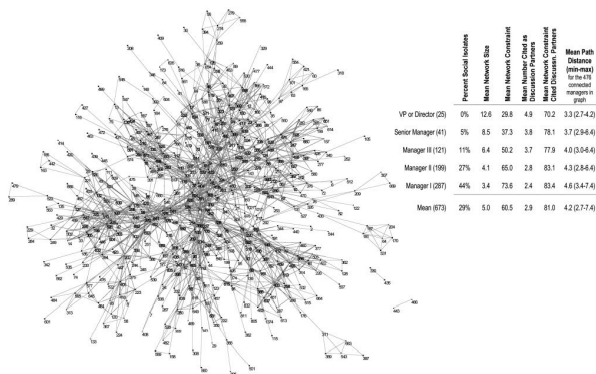


FIG. 2.—Supply-chain discussion network (excludes 193 social isolates)

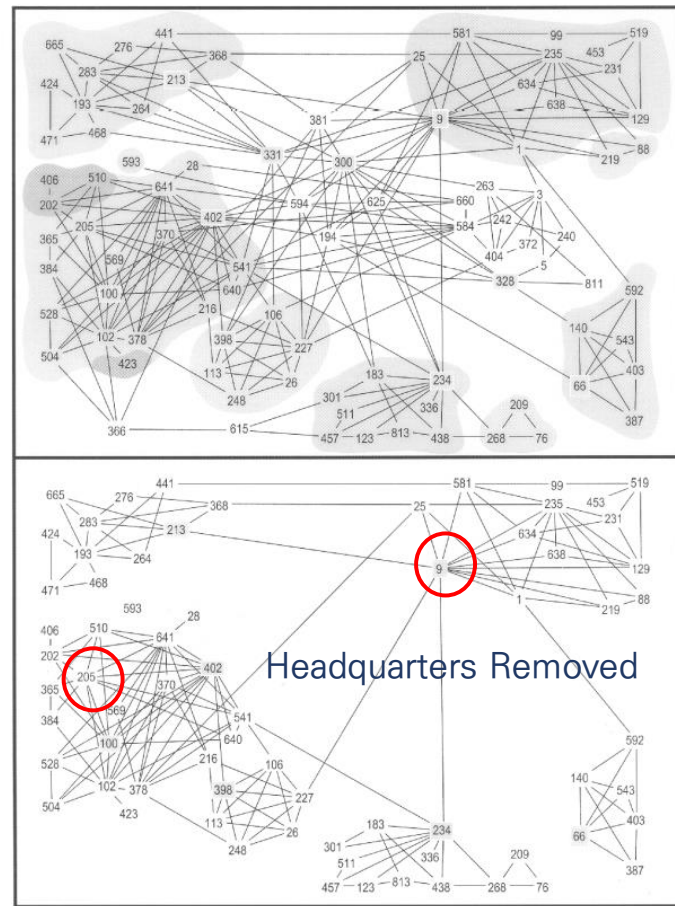


FIG. 3.—Core network in the supply chain

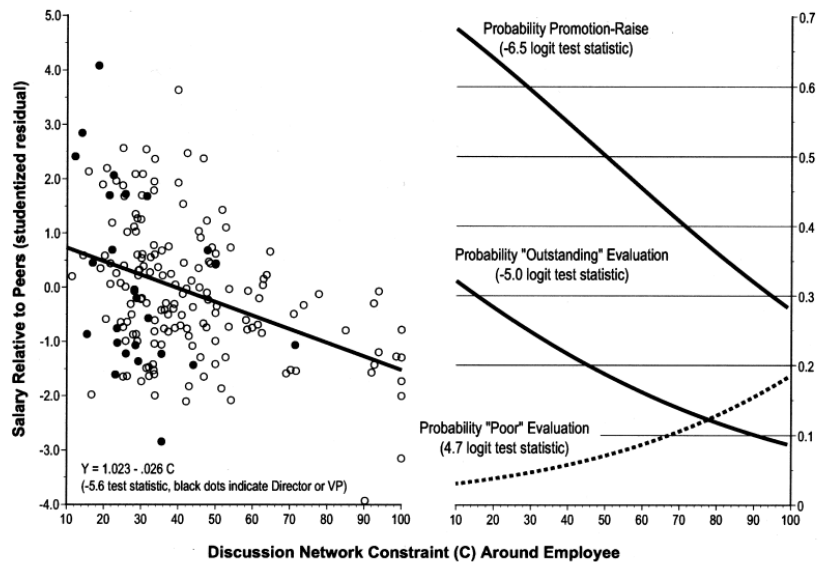


FIG. 4.—Brokerage and employee performance

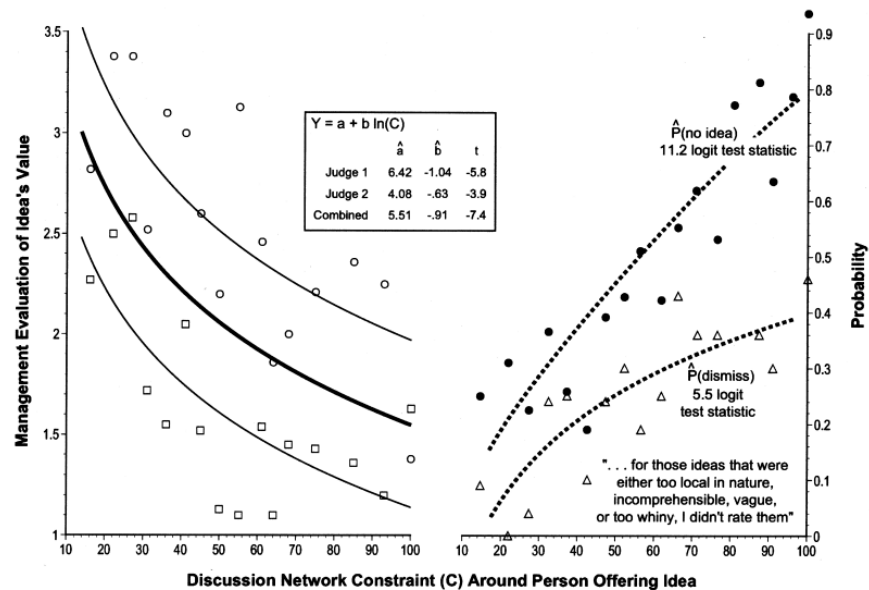
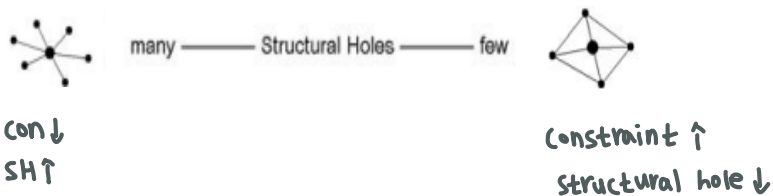


FIG. 5.—Brokerage and employee best idea

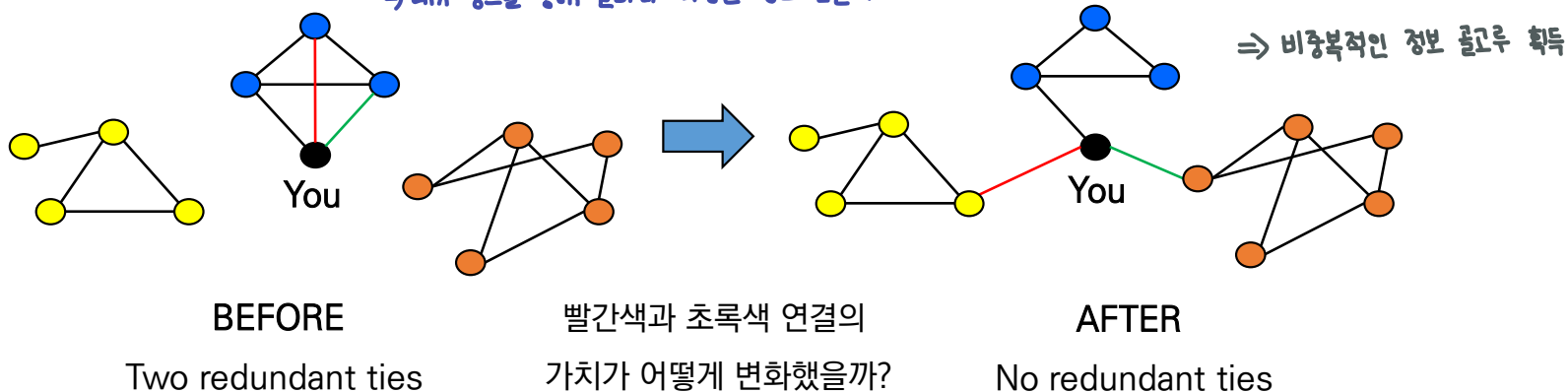


② 이질적 정보획득의 기회

- 구조적 공백을 많이 가지고 있는 노드는 **비중복적(이질적이고 다양한) 정보를 획득**할 수 있는 구조적 위치에 있음
- 예: 다섯명이 항상 몰려다닌다면? Bonding social capital은 높지만 새롭고 신선한 정보를 획득할 수 있는 기회는 많지 않음 (Bridging social capital 낮음)
- 예: 반대로 여러 사람들을 두루두루 만나는 경우는? Bonding social capital은 약하지만 새롭고 신선한 정보를 획득할 수 있는 기회 많음 (Bridging social capital 높음)
- 아래 그림에서 'You'가 가지고 있는 빨간색과 초록색 연결의 가치가 어떻게 변화할지 생각해 보자
- 두 가지 측정 방식: **effective size** & **efficiency**

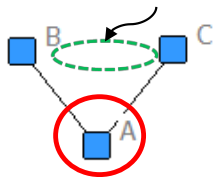
↳ 연결 하나 당 얼마나 다양한 정보 얻을 수 있는가

↳ 내가 링크를 통해 얼마나 다양한 정보 얻을 수 있는가



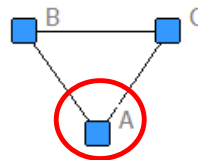
- **Effective Size**: 비중복적 정보를 얼마나 효과적으로 (많이) 획득할 수 있는지 측정
 - 비중복적 정보를 많이 획득하려고 alter를 무한정 늘릴 수는 없다
 - 연결을 유지하는 데에는 비용 (시간, 돈 등)이 수반되기 때문
- **Efficiency**: 비중복적 정보를 얼마나 효율적으로 (비용 대비) 획득할 수 있는지 측정
 - 적은 연결로 많은 비중복적 정보를 획득하는 것이 많은 연결로 많은 비중복적 정보를 획득하는 것보다 좋다
 - $$\text{Efficiency} = \frac{\text{Effective Size}}{\text{Ego Network Size}}$$
 (연결 1개 당 비용 대비 정보 획득 효과 - ego network size로 표준화)

Structural Hole



가정

예고 (A)가 하나의 연결을 유지하는 데 드는 비용 = 1
 예고 (A)가 각 알터로부터 획득하는 정보의 가치 = 2



- 순이익 = 2 = 4(정보이익) - 2(연결비용)
- B와 C가 가진 정보를 획득할 수 있는 대안이 없으므로 A-B, A-C 모두 효과적 연결
- 따라서 Effective Size = 2 (2개 연결 모두 효과적)
- Efficiency = $2/2 = 1$

- 순이익 = 2 = 4(정보이익) - 2(연결비용)
- B와 C가 가지고 있는 정보를 획득하는 데 A-B, A-C 연결이 모두 필요한가?
- A-C 혹은 A-B 하나만으로도 B와 C가 가지고 있는 정보(가치 = 4) 획득 가능
- 따라서 Effective Size = 1 (A-B 혹은 A-C 중 하나는 redundant tie)
- Efficiency = $1/2 = 0.5$

Network → Ego Networks → Structural Holes → Input dataset → Method → OK

7개의 사람들을 통해 유니크한 정보 3.61개 얻을 수 있음

Structural Hole Measures

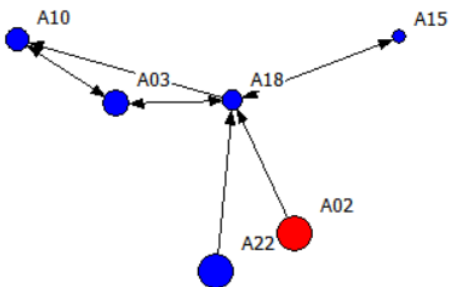
연결 하나 당 자홀

	1	2	3	4	5	6	7	8	9	10	11
	Degree	EffSize	Efficiency	Constraint	Hierarchy	EgoBet	Ln(Constraint)	Indirects	Density	AvgDeg	Open Pairs
1 A01	7	5.611	0.802	0.360	0.028	16	-1.022	0.565	0.262	1.571	31
2 A02	5	3.429	0.686	0.580	0.141	6	-0.545	0.595	0.350	1.400	13
3 A03	7	4.625	0.661	0.442	0.105	16.000	-0.816	0.659	0.357	2.143	27
4 A04	5	4	0.800	0.444	0.043	14	-0.811	0.467	0.250	1	15
5 A05	6	4.182	0.697	0.455	0.045	13.667	-0.788	0.618	0.367	1.833	19
6 A06	3	1.500	0.500	1.042	0.210	0.500	0.041	0.667	0.667	1.333	2
7 A07	7	5.318	0.760	0.389	0.104	21	-0.943	0.569	0.286	1.714	30
8 A08	2	2	1	0.500	0	2	-0.693	0	0	0	2
9 A09	9	6.333	0.704	0.348	0.040	23.950	-1.056	0.728	0.347	2.778	47
10 A10	8	5.143	0.643	0.399	0.044	16.333	-0.920	0.738	0.411	2.875	33
11 A11	10	8.750	0.875	0.246	0.080	56.667	-1.403	0.485	0.122	1.100	79
12 A12	0	0				0		0		0	0
13 A13	6	4.556	0.759	0.467	0.110	13.500	-0.761	0.593	0.233	1.167	23
14 A14	5	3.250	0.650	0.567	0.046	7.833	-0.568	0.654	0.450	1.800	11
15 A15	2	2	1	0.556	0.278	1	-0.588	0	0	0	2
16 A16	0	0				0		0		0	0
17 A17	9	6.219	0.691	0.341	0.029	31.233	-1.075	0.722	0.333	2.667	48
18 A18	5										18
19 A19	5										10
20 A20	10	7.706	0.771	0.286	0.021	47.583	-1.251	0.671	0.267	2.400	66
21 A21	8	6.500	0.813	0.314	0.069	37.500	-1.159	0.526	0.179	1.250	46

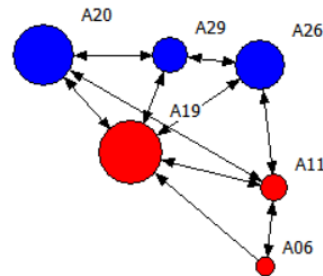
Structural holes

- 각 노드의 degree (ego-network size), Effsize, Efficiency, Constraints, Open Pairs (구조적 공백의 수)를 확인해 보자
 - A18의 A19는 모두 degree가 5이지만, Open Pairs는 18개, 10개이다
 - A18과 A19 중 Effective Size, Efficiency가 큰 노드는 무엇일지 추정해 보자. 마찬가지로 Constraint가 큰 노드는 무엇일까?

- 아래 제시된 A18과 A19의 Ego-Network을 확인해 보고 앞에서 답한 것이 맞는지 비교해 보자
 - Open Pairs (Structural Holes) 가 많을수록 Effective Size와 Efficiency는 증가하고, Constraint는 감소 (Autonomy 증가)
 - A18은 학습 스케줄이나 내용을 자율적으로 조정할 수 있고, 시험과 관련된 다양한 (비중복적) 정보를 획득할 수 있는 구조적 위치에 있다 (놀라운 점: 그런데 성적은 낮다!!)



A18의 Ego-Network

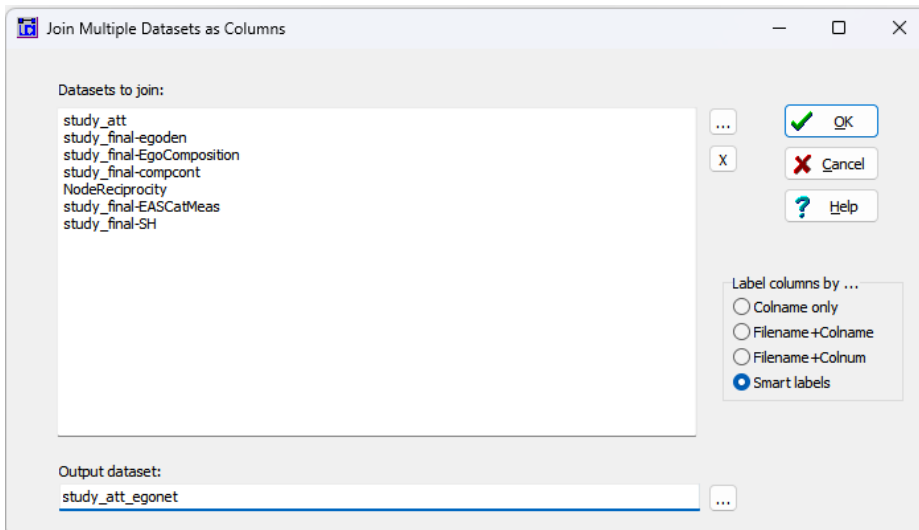


A19의 Ego-Network

constraints ↑
effective size ↓

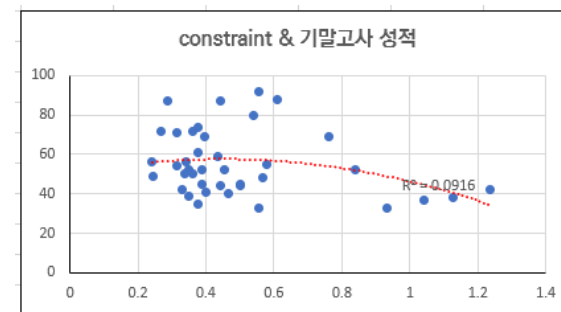
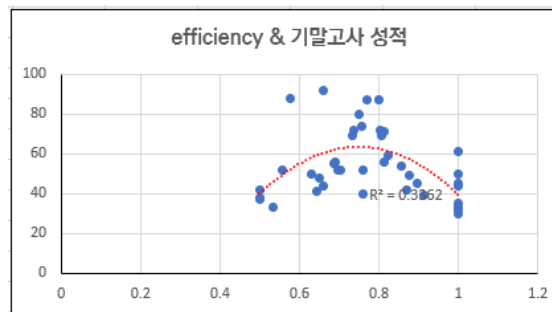
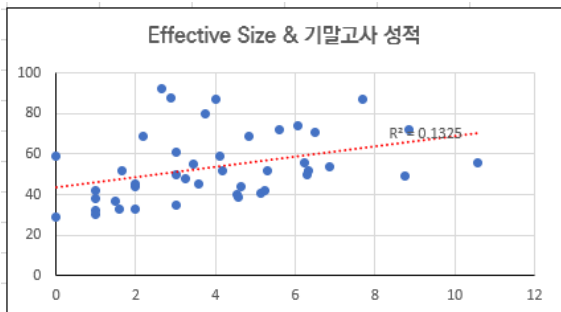
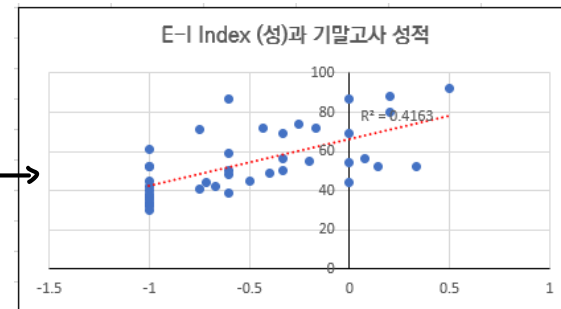
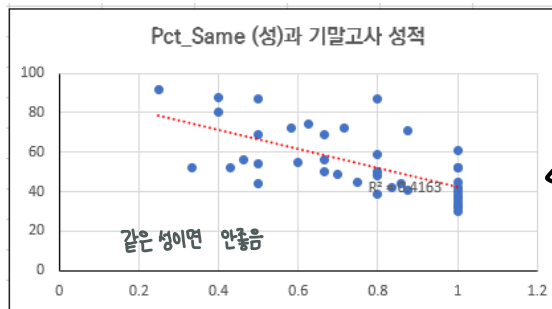
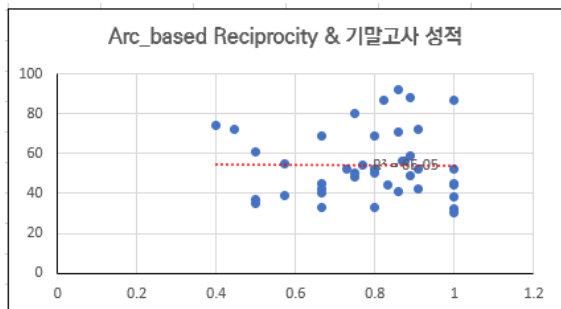
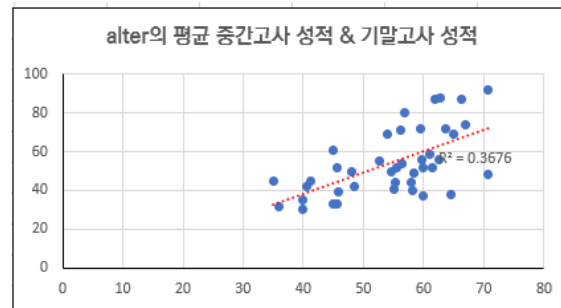
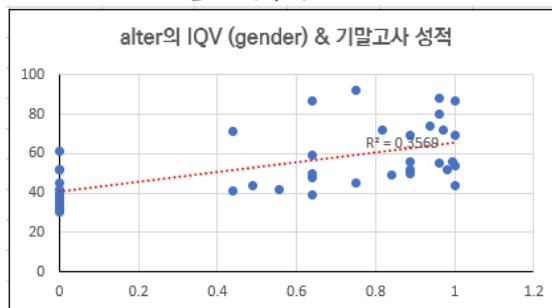
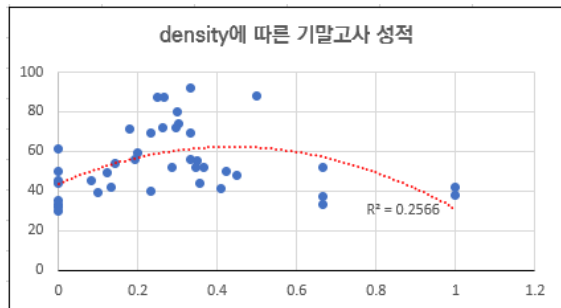
		1	2	3	4	5	6	7	8	9	10	11
		Degree	EffSize	Efficiency	Constraint	Hierarchy	EgoBet	Ln(Constraint)	Indirects	Density	AvgDeg	Open Pairs
18	A18	5	4.571	0.914	0.349	0.145	10	-1.052	0.238	0.100	0.400	18
19	A19	5	2.889	0.578	0.610	0.075	3.667	-0.494	0.685	0.500	2	10

- 중심성과 마찬가지로 ego-network 분석을 통해 나온 값은 노드의 네트워크 속성
- 따라서 이를 활용하여 노드의 고유 속성 (성적, 성과, 행동방식 등)과 연결하여 분석할 수 있음
- 중심성에서 배운 것처럼 각 ego-network 분석에서 나온 결과 데이터를 속성 데이터와 결합해 보자 (UCINET에서 join 명령어 활용) **Data → join → columns**
- 저장할 파일 이름은 지정할 수 있으나, 수업에서는 study_att_egonet으로 저장해 보자. 아래와 같이 study_att_egonet.##h 파일 (43×48 matrix)이 생성되었는지 확인하고, 파일을 엑셀로 export해 보자
- 중심성분석에서 했던 것과 마찬가지로 ego-network 분석에서 나온 값을 활용하여 기말고사 성적이 어떤 네트워크 속성에 영향을 받는지 알아보자.



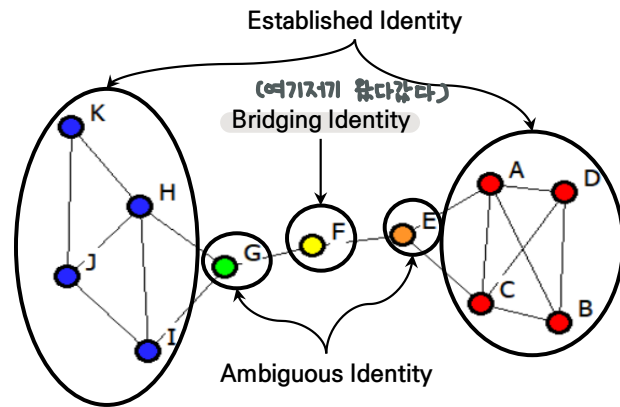
- study_att: 속성 데이터
- study_final-egoden: 기본구조
- study_final-Egocomposition: alter의 특성 (categorical)
- study_final-compcont: alter의 특성 (continuous)
- Nodereciprocity: ego-alter reciprocity
- Study_final-EASCatMeas: ego-alter similarity (성별 유사성)
- Study_final-SH: 구조적 공백

→ 성이 얼마나 이질적이나

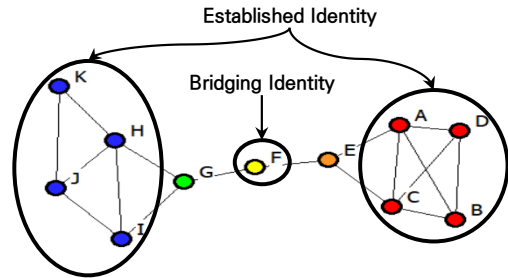


브랜드나 제품의 정체성이라는 관점으로 접근

- 자료 (6장에서 실습했던 데이터)
 - cosmetic_ingredient.xlsx ⇒ 화장품-성분 list ⇒ Cleanser 종류만 선택 (총 281개)
 - 사용하는 성분에 따라 cleanser 네트워크 구성 (one-mode network로 변환)
 - normalize 후 5%를 cut-off로 사용하여 dichotomize (상세한 네트워크 자료 구성 방법은 생략)
- 네트워크분석을 할 때에는 항상 ‘어떤’ 네트워크인가를 생각해야 합니다
 - 사회자본은 노드가 사람, 기업 등이고 연결은 정보나 자원인 경우에 적용 가능한 개념
 - 그러나 다른 유형의 네트워크라면?? ⇒ 상상력을 발휘하자!!
- 예: 노드가 브랜드나 제품이고 연결이 연관검색이나 성분 유사도인 경우는 브랜드/제품의 정체성 (identity)과 관련이 있을 수 있음
 - 사람들은 정체성이 비슷한 브랜드(메이커)를 검색하는 경향 ⇒ 람보르기니, 페라리, 포르쉐를 함께 검색하지만 람보르기니와 현대차를 함께 검색하지는 않음
 - 어떤 메이커의 예고네트워크가 유사한 노드로 구성되어 있고 densely closed network이라면 ⇒ 잘 정립된 정체성 (established identity)
 - 어떤 메이커의 예고네트워크가 이질적인 노드로 구성되어 있고 densely closed network라면 ⇒ 모호한 정체성 (ambiguous identity)
 - 어떤 메이커의 예고네트워크에 structural holes가 많다면 ⇒ 융합적 혹은 연결적 정체성 (bridging identity)



- 몇 가지 가설을 생각해 봤습니다.
- 화장품 성분 유사도 네트워크에서 어떤 화장품의 alter들이 유사한 성분인지 다른 성분인지에 따라 평판이 달라질 수 있을까?
 - 어떤 화장품이 성분이 비슷한 화장품들과 과한 밀도 높은 네트워크를 이루고 있다면... (alter 화장품의 성분이 유사하다면...) ⇒ 표준적인 성분을 사용하는 화장품 ⇒ 소비자에게 쉽게 각인이 되어 평판이 올라갈 수 있으나, 소비자에게 신선함을 주지 못하므로 평판이 떨어질 수도 있음
 - 성분이 상이한 화장품들을 이어주는 위치에 있다면... (alter 화장품의 성분이 서로 다르다면...) ⇒ 여러 성분들을 융합한 화장품 ⇒ 소비자에게 신선함을 주어 평판이 올라갈 수 있으나, 기존의 화장품 정체성과 다르므로 각인 효과가 떨어질 수 있음
 - 가설 1-1: 에고네트워크의 *constraint*가 높으면 화장품에 대한 평균 평점이 올라갈 것이다.
 - 가설 1-2: 에고네트워크의 *constraint*가 높으면 화장품에 대한 평균 평점이 떨어질 것이다.
- 성분이 비슷한 화장품이 상대적으로 가격이 높다면... (alter들의 평균 가격이 높다면...)
 - 비슷한 성분의 다른 화장품에 비해 가격이 낮으므로 가격 경쟁력이 있고, 가격이 높은 제품과 연결되어 고급 이미지를 구축할 수 있음 ⇒ 평판이 올라갈 수 있지 않을까?
 - 가설 2: 성분이 비슷한 다른 화장품의 평균 가격이 높으면 화장품에 대한 평균평점이 올라갈 것이다.



- 앞에서 배운 다양한 분석을 수행할 수 있으나 여기서는 가설과 관련된 분석만 수행합니다.
 - Alter의 평균가격을 구하기 위해 continuous alter attributes 분석을 수행하고
 - Constraint를 구하기 위해 structural holes 분석을 수행한 후
 - Attribute data와 join하여 데이터셋 구축 (아마도 281x27 행렬)하고 excel로 export
 - Isolate, 평점이 0으로 되어 있는 데이터 등 추가적인 데이터 정제가 필요하지만, 일단 있는 그대로 분석해 봅니다

id	Price	rank	Combinatio	Dry	Normal	Oily	Sensitive	Avg	Sum	Min	Max	StdDev	EstSD	CV	Num	WtdNum	Degree	EffSize	Efficiency	Constraint
3 Muslin	22	4.19999980	0	0	0	0	0								0	0	0	0		
3% Glycolic	38	4.40000009	1	1	1	1	1	37.2000007	372	17	80	15.9486675	16.8113727	0.42872762	10	10	10	2.59999999	0.25999999	0.35125184
7 Day Scrub	22	4.59999990	0	0	0	0	0	35.3846168	920	15	75	13.9147872	14.1903543	0.39324399	26	26	26	16.7692298	0.64497041	0.14457625
8% Glycolic	40	4.40000009	1	1	1	1	1	38.5757560	1273	8	95	20.4480571	20.7651004	0.53007531	33	33	33	21.8484840	0.66207528	0.11605840
A Glowing	34	5	1	1	1	1	1	35.0816345	5157	8	95	17.1973381	17.2561321	0.49020916	147	147	147	125.136054	0.85126566	0.02756656
A Perfect	25	4.5	0	0	0	0	0	32	416	9	79	17.1015071	17.7998123	0.53442209	13	13	13	3.92307686	0.30177515	0.27937540
Aburatori	12	4.59999990	1	1	1	1	1								0	0	0	0		
Acne	38	4.19999980	1	1	1	1	1	32.9285697	461	25	44	5.93373250	6.15772533	0.18020012	14	14	14	7	0.5	0.25342273
Acne	17	4.40000009	1	1	1	1	0	32.6428565	457	16	95	21.6683025	22.4862594	0.66379916	14	14	14	6.14285707	0.43877550	0.25566962
Acne	21	4.19999980	1	1	1	1	0	37.4166679	898	17	95	20.6355609	21.0793876	0.55150723	24	24	24	14.75	0.61458331	0.15364536
AHA/BHA	39	4.59999990	0	0	0	0	0	44.1764717	751	20	80	17.6343097	18.1770305	0.39917877	17	17	17	9.11764717	0.53633219	0.21078319
Alpha Beta	38	4.40000009	1	1	1	1	1	32.0701751	1828	8	80	12.2658557	12.3748874	0.38246073	57	57	57	43.3508758	0.76054167	0.06797057

- 엑셀 분산형 차트로 확인해 본 결과...
- Constraint가 증가할수록 평가가 낮아짐
 - 유사한 성분의 화장품들과 비슷하게 만드는 것보다...
 - 서로 다른 성분을 조합하여 만든 화장품이 평가가 좋다
- Alter 화장품들(성분이 유사한 화장품들)의 가격이 높을수록, 해당 화장품의 평가는 좋아진다
 - 그러나 관계는 미미하다

```
. reg rank price oily constraint mean_price zegers
```

Source	SS	df	MS	Number of obs	=	224
Model	4.82306938	5	.964613876	F(5, 218)	=	2.91
Residual	72.20193	218	.331201514	Prob > F	=	0.0144
				R-squared	=	0.0626
				Adj R-squared	=	0.0411
Total	77.0249994	223	.345403585	Root MSE	=	.5755

rank	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
price	.0015511	.0025904	0.60	0.550	-.0035544 .0066565
oily	.1216675	.0790476	1.54	0.125	-.0341278 .2774629
constraint	-.6061116	.2202189	-2.75	0.006	-1.040142 -.1720809
mean_price	.0072555	.0078082	0.93	0.354	-.0081337 .0226448
zegers	.1041764	.6626166	0.16	0.875	-1.201778 1.410131
_cons	4.132188	.304535	13.57	0.000	3.531978 4.732398

