

그림 말고 수칙만 이해

6. 네트워크 구조 이해하기

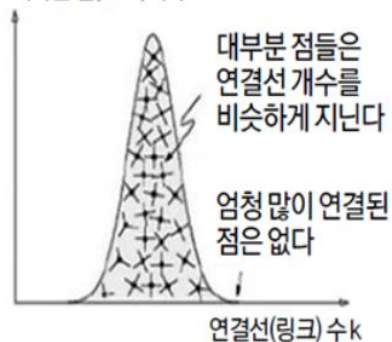
- 어떤 차이가 보이나요?

몰라도 됨

가우시안 분포

멱함수 분포

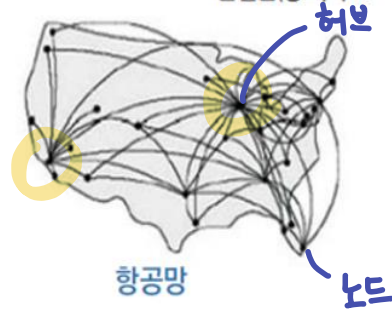
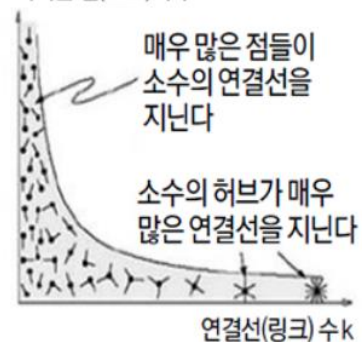
연결선 k 개를
지나는 점(노드)의 수



고속도로망

허브 안보임

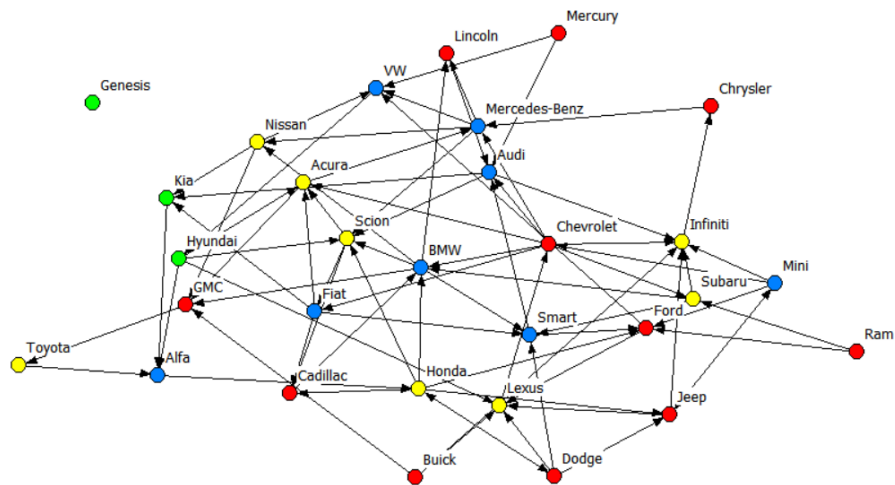
연결선 k 개를
지나는 점(노드)의 수



항공망

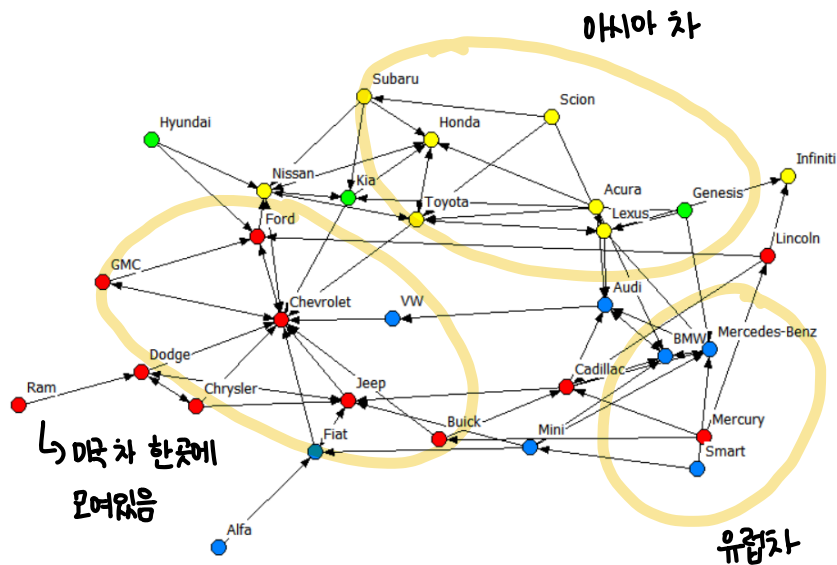
노드

- 어떤 차이가 보이나요?



무작위 네트워크

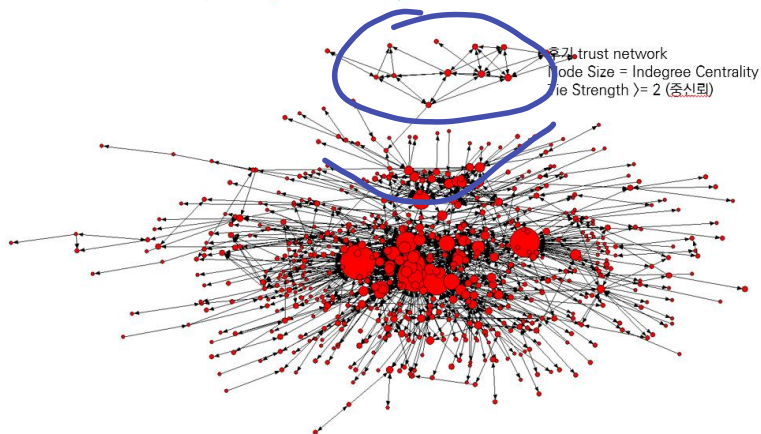
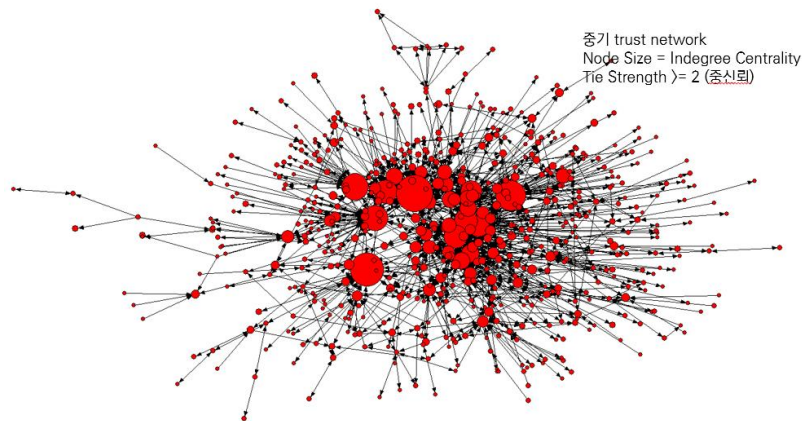
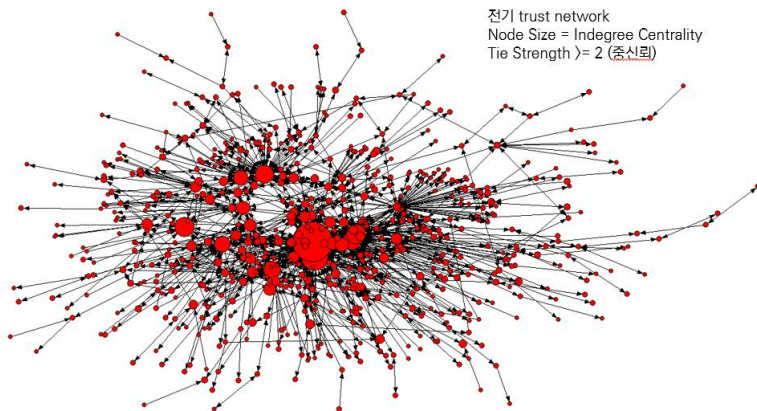
(사람들이 아무런 기준 없이 무작위로 연관검색을 한다고 가정했을 때 나타날 수 있는 자동차 연관검색 네트워크)



실제 네트워크

(실제 관찰된 자동차 연관검색 네트워크: 사람들이 어떤 기준에 따라 연관검색을 하는 것을 알 수 있음)

- 비트코인 신뢰 네트워크의 변화

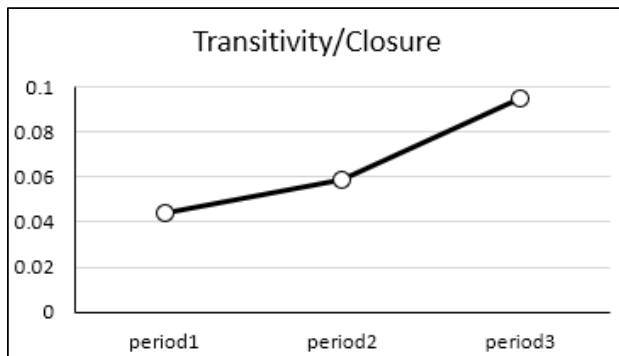
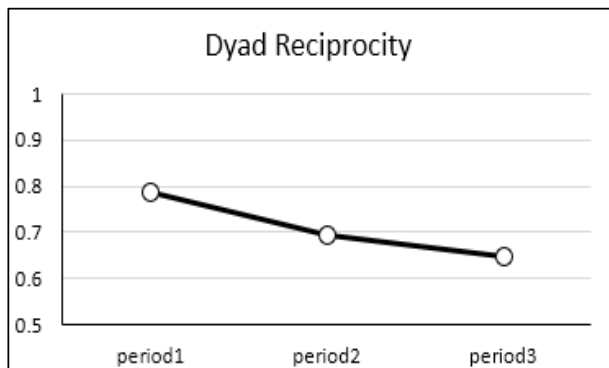
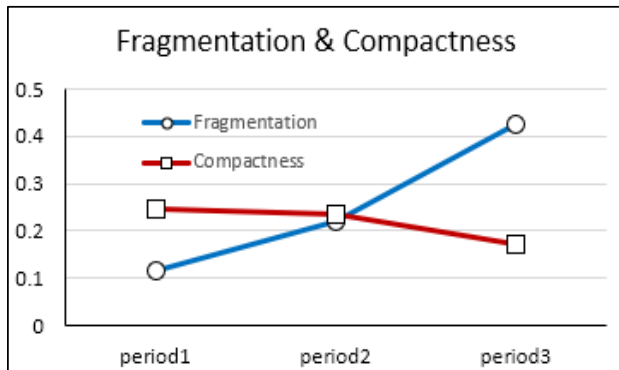
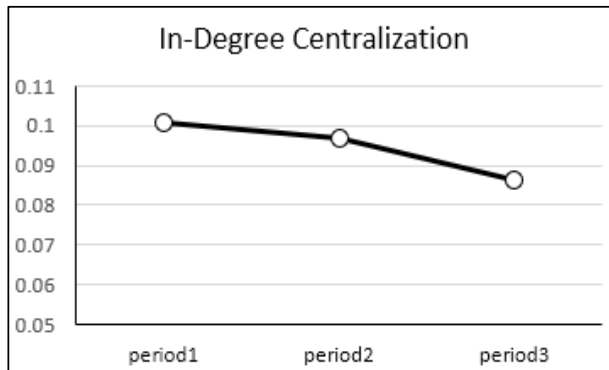


bitcoin_trust_network가 어떻게 달라지는지 추측해 보자.

1. 신뢰가 소수의 사람에게 집중되는 정도는 어떻게 변화하는가?
2. 서로 신뢰하는 사람들의 집단이 분화되는가, 하나로 모이는가?
3. 신뢰를 주고받는 사람들이 증가하는가 감소하는가?
4. 서로 신뢰하는 사람들끼리 모이는 양상이 증가하는가?

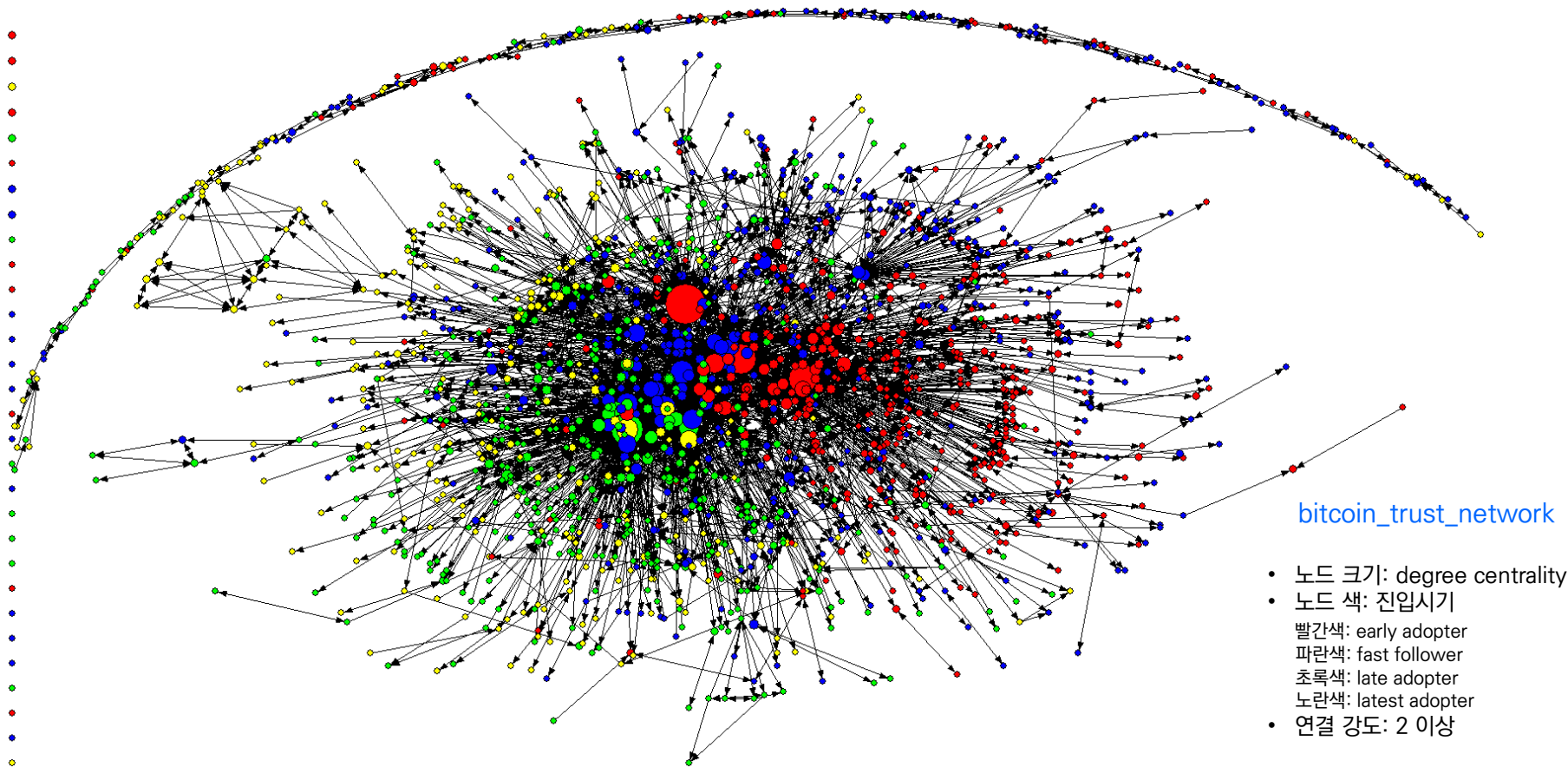
↳ 눈으로 보는 것 모겠음ㅠㅠ

- 오늘 배울 측정치를 가지고 측정해 보면...



- 평판/신뢰도가 높은 노드들이 점점 증가하여 그림 중앙으로 모임 (그림 1)
- 신뢰 네트워크는 지속적으로 분화. 아마도 신뢰도가 낮은 사람들이 MC에서 떨어져 나가는 것으로 해석 (그림 2)
- 신뢰도 평가는 대칭적이지 않으나 (그림 3)
- 평판/신뢰도가 높은 사람끼리 mutually respect (rewards)하는 것으로 해석 (그림 4)

- 비트코인 신뢰 네트워크 (전체 시기): 진입 시기에 따른 네트워크 구조: 무엇이 보이나요?

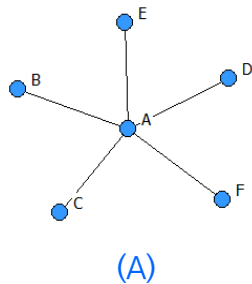


- 네트워크 구조 = 네트워크의 특징적 모양

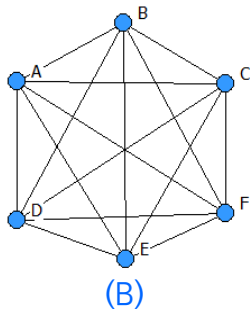
위계 : 적어남
수평 : 합이

- 창발(emergence): 네트워크 구조에 따라 시스템 특성(효율성, 혁신, 갈등, 권력행사방식, 정보확산 속도 등) 차이
- 아래 여러 네트워크 그림을 보고 각 네트워크가 어떤 특성을 가지고 있을지 생각해 보자.

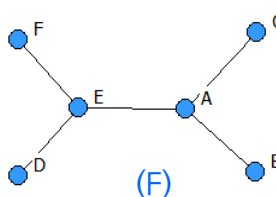
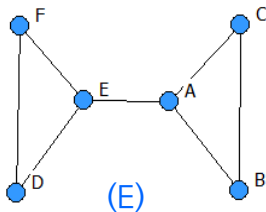
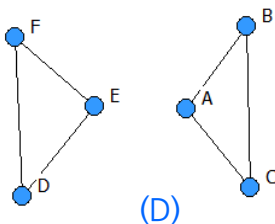
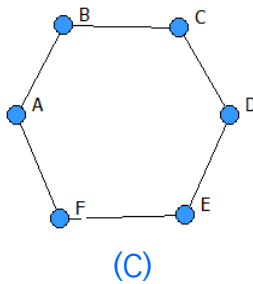
위계적



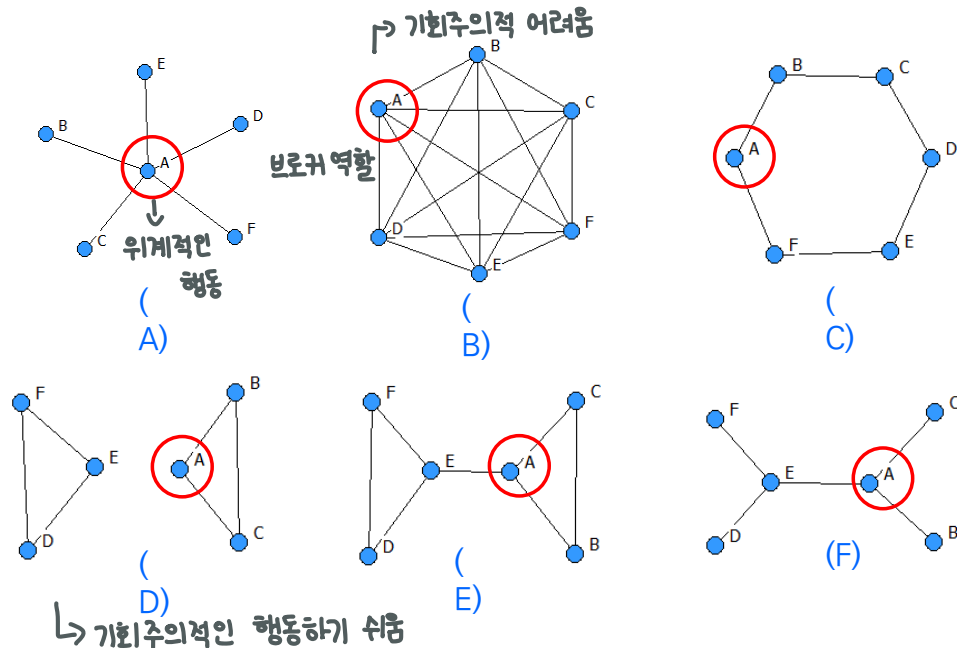
수평적



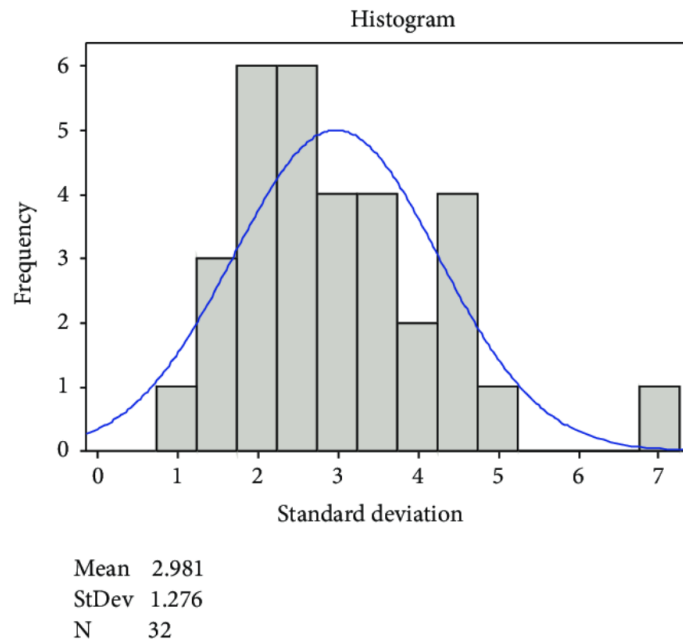
수평적



- **배태성(embeddedness):** 네트워크 구조에 따라 노드의 특성이 달라짐 ↳ 내가 어느 네트워크에 있느냐에 따라
 - 노드의 행동, 태도, 사고방식은 전체 네트워크 구조에서 개별 노드가 차지하는 위치에 따라 달라진다
 - 예: 응집된 구조에서의 집단 순응, 분화된 구조에서의 기회주의적 행동, 수평적 구조에서의 협력지향성
- 각 네트워크에서 노드 A는 어떤 특성을 갖고 있을지 말해 보자.



- 시각화 vs 수치화
 - 그림으로 판을 읽고, 수치(측정치)로 정확히 확인하자
- 네트워크 구조를 요약/수치화하는 다양한 방법
 - 응집성(cohesion)
 - 집중화(centralization)
 - 호혜성(reciprocity)
 - 이행성(transitivity)
 - 군집화계수(clustering coefficient)
 - 동질성(homophily)

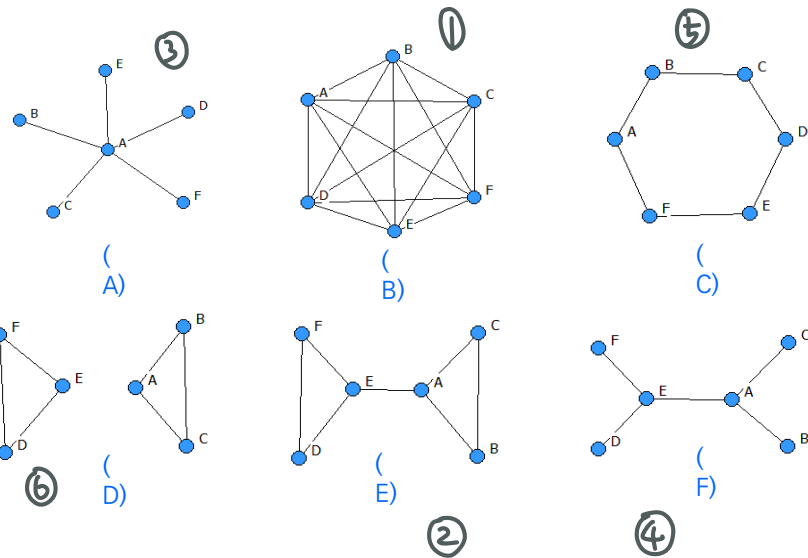


• 응집성

- 네트워크 내의 노드들이 얼마나 긴밀하게 연결되어 있는지 측정
- Cohesive network: 노드 간 연결이 많고 단단히 뭉쳐 있음 : B
- Sparse network: 노드 간 연결이 적고 서로 분리되어 있음 : D

• 응집성 측정의 다양한 방법

- 비슷 (
- 연결밀도(density) : 링크가 얼마나 있느냐
 - 평균 연결정도(average degree) : 한 노드가 평균적으로 몇개의 링크?
 - 연결도(connectedness) \Leftrightarrow 분절도(Fragmentation)



- 연결밀도 & 평균 연결정도: 노드들이 얼마나 연결되어 있는가?
- 연결밀도 (density): 가능한 모든 연결 중 실제 존재하는 연결 수 (binary data 인 경우) 혹은 강도 (valued data인 경우)

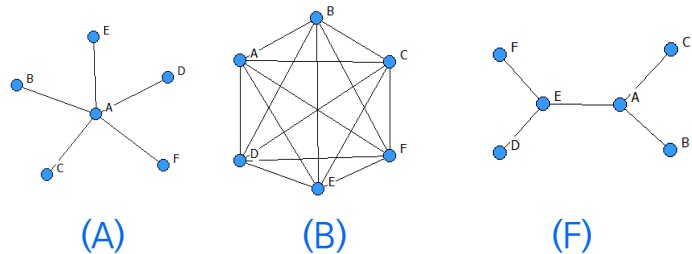
- (A) $5/15 = 0.33$
- (B) $15/15 = 1.00$
- (F) $5/15 = 0.33$
- (G) $9/6 = 1.50$
- (H) $6/6 = 1.00$

→ 노드 하나 당 몇개의 연결이 있느냐

- 평균 연결정도 (average degree): 노드 하나 당 연결의 수 (binary data인 경우) 혹은 강도 (valued data인 경우) ⇒ 크기가 다른 네트워크를 비교할 때

- (A) $10/6 = 1.67$
- (B) $30/6 = 5$
- (F) $10/6 = 1.67$
- (G) $9/3 = 3$
- (H) $6/3 = 2$

$$C_2 = \frac{6 \times 5}{2 \times 1} = 15$$



	A	B	C
A		0	3
B		5	0
C		1	0

(G)

⇒ 방향이 있는
네트워크 (강도)
valued data

	A	B	C
A		0	2
B		3	0
C		1	0

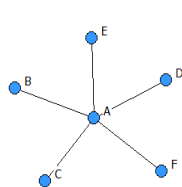
(H)

- 연결도 (connectedness): 전체 노드쌍 중 어떤 경로를 통해서든 연결될 수 있는 쌍의 비율 (\leftrightarrow 분리도 (fragmentation))

- 모든 노드쌍이 연결되어 있다면 $\text{connectedness} = 1$ ($\text{fragmentation} = 0$)
- 모든 노드쌍이 전혀 연결되어 있지 않다면 $\text{connectedness} = 0$ ($\text{fragmentation} = 1$)
- 주로 방향 데이터에 많이 활용 (비방향 데이터인 경우에는 (D)와 같은 경우를 제외하고 대부분 1)

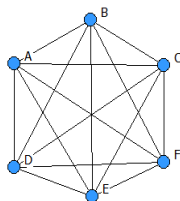


- 네트워크 (A), (B), (F): $\text{connectedness} = 15/15 = 1$ ($\text{fragmentation} = 1 - 1 = 0$)
- 네트워크 (D): $\text{connectedness} = 6/15 = 0.40$ ($\text{fragmentation} = 1 - 0.4 = 0.6$)
- 네트워크 (I): $\text{connectedness} = 4/6 = 0.67$ ($\text{fragmentation} = 1 - 0.67 = 0.33$)

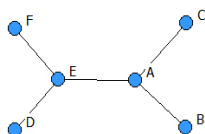


(A)

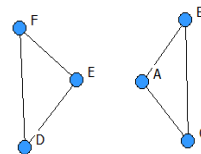
E와 F 연결



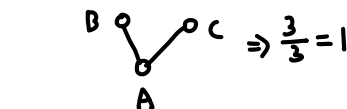
(B)



(F)



(D)

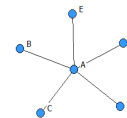


(I)

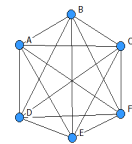
$$\frac{4}{6} = 66\%$$

$A \rightarrow B$	○
$A \rightarrow C$	○
$B \rightarrow A$	×
$B \rightarrow C$	×
$C \rightarrow A$	○
$C \rightarrow B$	○

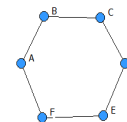
- **집중도**: 연결이 소수의 사람에게 집중된 정도
 - 집중도가 높으면 소수의 사람에게 권한, 정보, 자원이 집중
 - 집중도가 낮으면 모든 사람에게 권한, 정보, 자원이 균일하게 분포
 - 옆의 그림에서 가장 집중도가 높은 네트워크와 집중도가 가장 낮은 네트워크를 찾아봅시다
 - Directed data의 경우: in-centralization & out-centralization
- **집중도의 측정방식**
 - 완전히 집중화된 네트워크(star-network)와 얼마나 유사한가?
 - 집중도 = 네트워크 연결수 차이의 합 / Start network 연결수 차이의 합
 - 네트워크 연결수 차이의 합 = \sum (가장 연결이 많은 노드의 연결수 - 다른 노드의 연결수)
 - Star network 연결수 차이의 합 = $(n-1)(n-2)$



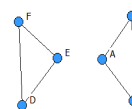
$$(0+4+4+4+4+4)/20 = 1$$



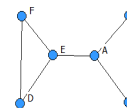
$$(0+0+0+0+0+0)/20 = 0$$



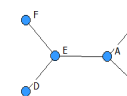
$$(0+0+0+0+0+0)/20 = 0$$



$$(0+0+0+0+0+0)/20 = 0$$

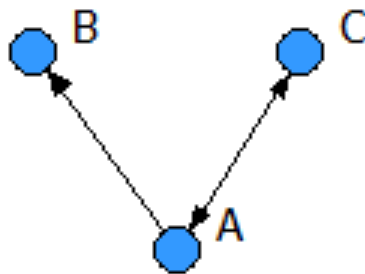


$$(0+1+1+1+0+1)/20 = 0.20$$

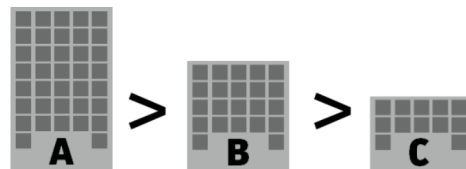


$$(0+2+2+2+0+2)/20 = 0.40$$

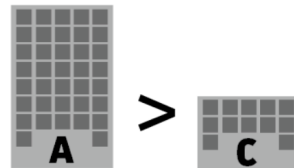
- 네트워크에서 일방적 관계가 많은가, 아니면 서로 주고 받는 양방향 관계가 많은가?
- 호혜성: 정보, 자원, 신호, 감정을 서로 주고 받는 정도
 - 네트워크에 속한 노드들이 대부분 양방향으로 연결된 경우 → 대체로 수평적이고 평등한 관계가 지배적 (예: 가족, 친구)
 - 네트워크에 속한 노드들이 대부분 일방향으로 연결된 경우 → 대체로 수직적이고 위계적 관계가 지배적 (예: 공식 보고체계)
 - 예: 공급사슬네트워크에서는 일방향 관계가 지배적 vs. 기술협력 네트워크에서는 양방향 관계가 지배적
 - 예: 업무 네트워크에서는 일방향 관계가 지배적 vs. 동아리(친구, 비공식 모임)에서는 쌍방향 관계가 지배적
- Dyad reciprocity: 전체 노드쌍 중 양방향 노드쌍의 비율
 - 옆의 그림: 2개의 노드쌍 중 양방향 노드쌍은 1개 → $1/2$
- Arc reciprocity: 전체 연결 중 양방향 연결에 속하는 연결의 비율
 - 옆의 그림: 전체 연결 3개 중 양방향 연결에 속하는 연결은 2개 → $2/3$



- 질문: 잘 기능하는 안정적인 묶음 구조가 존재하는가?
- 이행성의 수학적 정의
 - $(A \rightarrow B) \& (B \rightarrow C) \Rightarrow (A \rightarrow C)$ or $(A < B) \& (B < C) \Rightarrow (A < C)$



So:



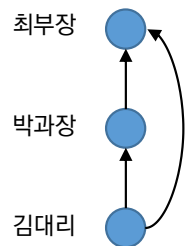
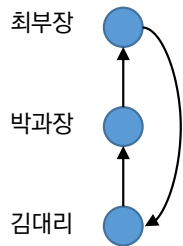
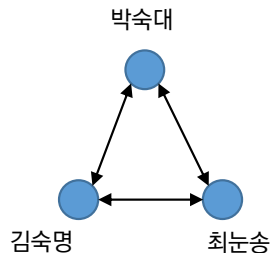
- 이행성은 네트워크 구조에 대해 무엇을 알려주는가?

① 네트워크 구조가 안정적이다

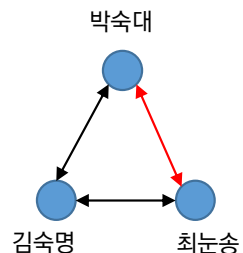
- 3개 노드 간의 관계(연결의 패턴)가 안정을 갖춘 상태 (다음 장 위쪽 그림 참고)
- 이행성이 높으면 \Rightarrow 삼자 이상의 관계가 안정적으로 유지되는 구조
- 이행성이 낮으면 \Rightarrow 불안정하고 변화의 가능성이 높은 구조

② 덩어리(묶음)이 존재하는 구조이다

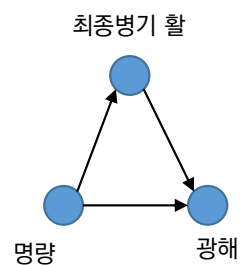
- 덩어리 구조: 이행형 삼자관계가 많이 분포하는 네트워크는 덩어리 구조를 갖는 경향 (다음 장 아래쪽 그림 참고)
- 이행형 삼자관계가 많으면 \Rightarrow 관계가 중첩되면서 점점 커지고 덩어리 구조(clumpy structure) 출현 \Rightarrow 클러스터링
- 공통의 관심사, 신뢰, 특성에 근거하여 노드들이 서로 묶여 있는 구조
- 예: 팀원들끼리 강하게 결속; 친구들끼리 study 'group'; 매운맛 음식끼리 강하게 연결; 같은 장르의 영화끼리 강하게 연결

안정적 보고/자
문 네트워크불안정한 보고/자문
네트워크

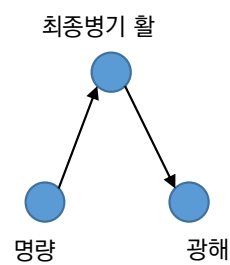
안정적 친구 네트워크



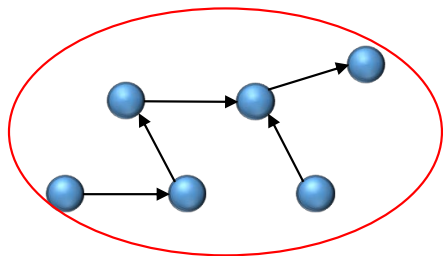
불안정한 친구 네트워크



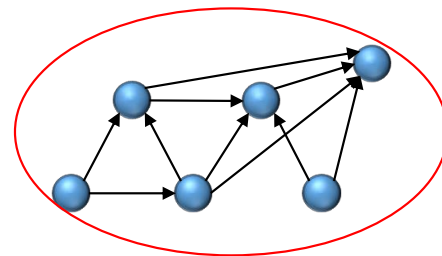
안정적 공동시청 네트워크



불안정한 공동시청 네트워크

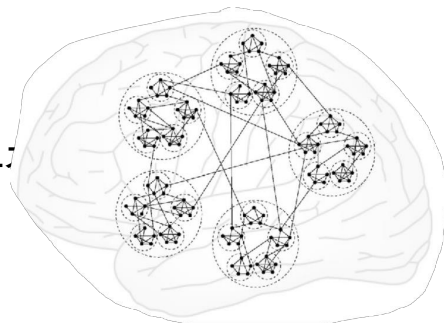


이행형 삼자 관계가 없음 (내 친구의 친구는 내 친구가 아님) ⇒ 흩어진 구조

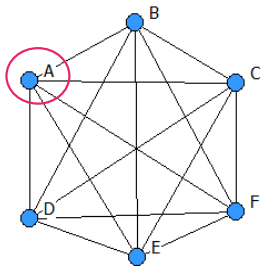


이행형 삼자관계가 지배적 (내친구의 친구는 내 친구임) ⇒ 덩어리 구조

- **군집화**: 네트워크에 있는 노드들이 서로 무리지어 있는 정도
 - Transitivity (이행형 삼자관계)가 중첩적으로 결합되면 덩어리로 군
 - 따라서 transitivity가 높으면 군집화도 높아질 가능성 높음

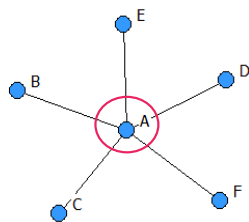


Brain Network as a Small World



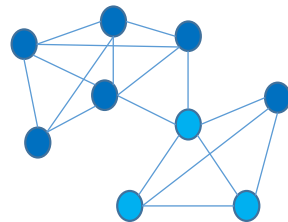
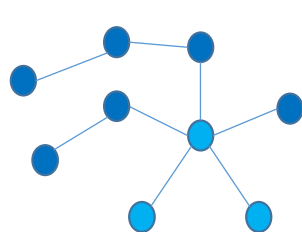
내가 알고 있는 사람들이 모두 서로 다 잘 알고 있는 사람들이라면...

- 이행형 삼자관계들이 완벽히 중첩
- clustering coefficient = 1

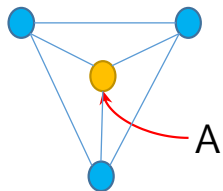


내가 알고 있는 사람들이 서로 아무도 알지 못한다면

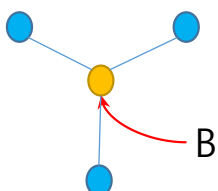
- 이행형 삼자관계가 하나도 없음
- clustering coefficient = 0



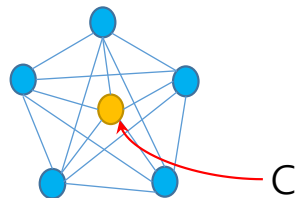
- Overall clustering coefficient (OCC)
 - 특정 노드(ego)와 연결된 노드들(alter) 간의 연결밀도의 평균값
 - 아래 예에서 A는 B보다 OCC가 높다 (노드 수준) & Network1은 Network2보다 OCC가 높다 (네트워크 수준)
- Weighted overall clustering coefficient (WOCC)
 - 연결된 노드(alter)가 많은 경우 그에 비례해 가중치 부여 \Rightarrow 네트워크 밀도를 감안했으므로 WOCC 사용하는 것이 좋음
 - 예: A는 alter가 3명인데 모두 연결, C는 alter가 5명인데 모두 연결 \rightarrow A, C 중 군집화 정도가 높은 노드는?
 - 가중치 = 어떤 노드에 연결된 alter들 간에 형성될 수 있는 모든 쌍(pair)의 수 $(= g(g-1)/2)$
- 군집화 계수는 이행성과 마찬가지로 어떤 속성에 따라 노드들이 묶여 있는 구조를 나타냄



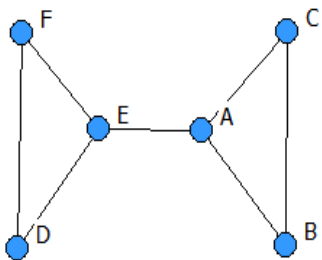
Network 1



Network 2



Network 3



노드	연결 노드 (g)	가능한 노드쌍 = $g(g-1)/2$	연결된 쌍	연결밀도
A	3 (B, C, E)	3 (B-C, B-E, C-E)	1 (B-C)	0.33
B	2 (A, C)	1 (A-C)	1 (A-C)	1.00
C	2 (A, B)	1 (A-B)	1 (A-B)	1.00
D	2 (E, F)	1 (E-F)	1 (E-F)	1.00
E	3 (A, D, F)	3 (A-D, A-F, D-F)	1 (D-F)	0.33
F	2 (D, E)	1 (D-E)	1 (D-E)	1.00

Overall graph clustering coefficient: 0.778

Network density: 0.467

$(cc - density)/cc$: 0.400

Weighted Overall graph clustering coefficient: 0.600

Small world index: 1.381

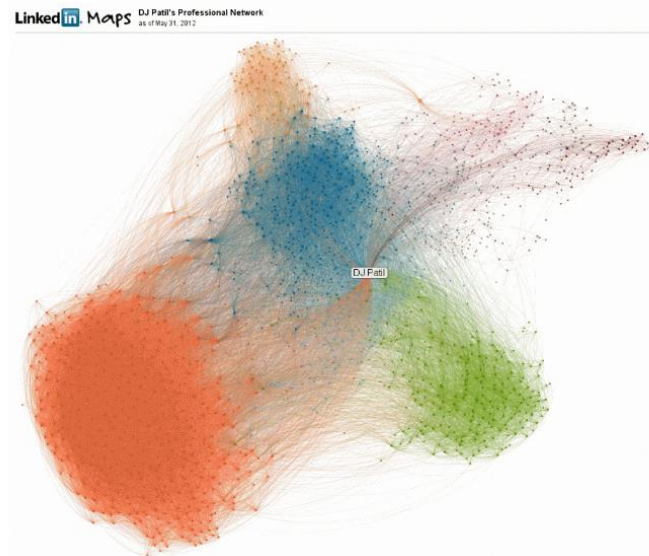
Node Clustering Coefficients

	1	2
	Clus Coef	nPairs
1 A	0.333	3.000
2 B	1.000	1.000
3 C	1.000	1.000
4 D	1.000	1.000
5 E	0.333	3.000
6 F	1.000	1.000

$$OCC = (0.33 + 1 + 1 + 1 + 0.33 + 1)/6 = 0.778$$

$$WOOC = (0.33*3 + 1.00*1 + 1.00*1 + 1.00*1 + 1.00*1 + 0.33*3) / (3 + 1 + 1 + 1 + 3 + 1) = 0.60$$

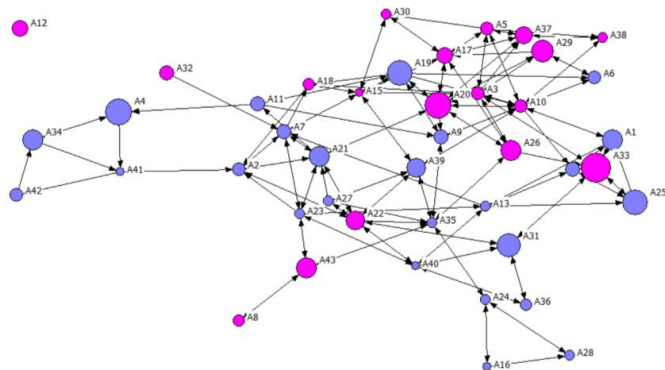
- 군집화 계수가 높은 네트워크는 어떤 특성을 가지고 있을까?
- **Small World** (Watts and Strogatz)
 - 규모가 큰 네트워크가 ‘작은 세상’인 이유는?
 - 3억 미국인은 6단계 만에 만날 수 있다
 - 수십억 facebook 이용자는 3~4단계 만에 친구가 될 수 있다
 - 5천만 대한민국 국민들 사이에서 ‘가짜 뉴스’는 삼시간에 퍼진다
 - 수십억 세계 인구 사이에서 ‘오징어게임’ 명성이 며칠 만에 확산된다
 - 케빈 베이컨은 어떻게 수많은 할리우드 스타들과 3단계 만에 연결된다...
 - 왜?
 - 규모가 큰 네트워크 (예: SNS)에서는 fully connected network은 불가능할 뿐만 아니라 비효율적
 - 대신에 네트워크가 여러 개의 군집으로 나뉘고, 이 군집들을 이어주는 ‘hub’ 노드가 존재 (앞의 brain network 참조) ⇒ 노드 간 거리를 극적으로 줄여주고 정보확산 가속화하는 매우 효율적 네트워크
 - 작은 세상 네트워크: 높은 clustering coefficient + 상대적으로 소수의 broker



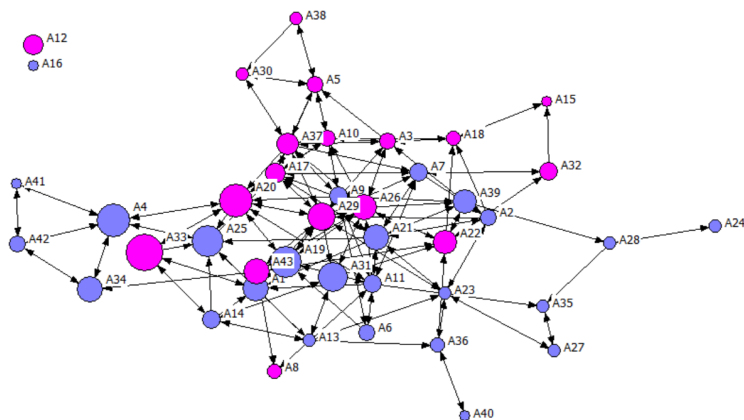
LinkedIn Professional Network

<https://blogs.unimelb.edu.au/sciencecommunication/2012/09/25/its-a-small-world-after-all/>

- OO대학 OO학과 통계수업을 수강한 학생들의 학습네트워크 자료
 - 중간고사 직전과 기말고사 직전에 설문조사 실시 → 과제수행, 시험준비 등을 위해 공부를 같이 했거나 도움을 받은 사람 복수 선택
- 자료파일 설명
 - study_mid: 학기 시작 ~ 중간고사 직전까지의 학습 네트워크
 - study_final: 중간고사 ~ 기말고사 직전까지의 학습 네트워크
 - study_att: score (중간고사 성적), gender (1 = 남성, 2 = 여성), score_rec2 (1 = 중간고사 성적 61점 이상, 2 = 60점 이하), score_rec3 (1 = 중간고사 성적 61점 이상, 2 = 41~60점, 3 = 40점 이하)



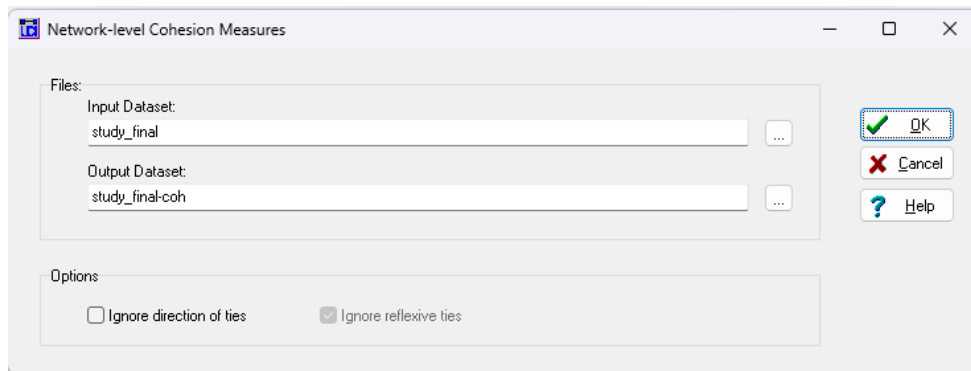
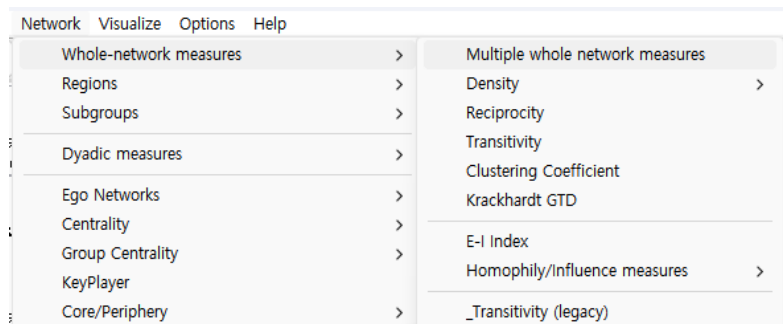
중간고사 이전 학습네트워크



중간고사 이후 학습네트워크

- OO대학 OO학과 통계수업을 수강한 학생들의 학습네트워크 자료
 - 아래를 참고하여 study_mid와 study_final 각각의 whole network measure를 구해보자

Network → Whole-network measures → Multiple whole network measures → Input Dataset → OK



Whole network measures

		1
		study_mid

1	# of nodes	43
2	# of ties	169
3	Avg Degree	3.930
4	Indeg H-Index	6
5	K-core index	4
6	Deg Centralization	0.107
7	Out-Centralization	0.124
8	In-Centralization	0.099
9	Indeg Corr	-0.068
10	Outdeg Corr	-0.065
11	Density	0.094
12	Components	3
13	Component Ratio	0.048
14	Connectedness	0.931
15	Fragmentation	0.069
16	Transitivity/Closure	0.253
17	Avg Distance	3.107
18	Prop within 3	0.605
19	SD Distance	1.304
20	Diameter	7
21	Wiener Index	5223
22	Dependency Sum	3542
23	Breadth	0.625
24	Compactness	0.375
25	Small Worldness	
26	Mutuals	0.075
27	Asymmetrics	0.037
28	Nulls	0.888
29	Arc Reciprocity	0.805
30	Dyad Reciprocity	0.673

30 rows, 1 columns, 1 levels.

Whole network measures

		1
		study_fin
		al

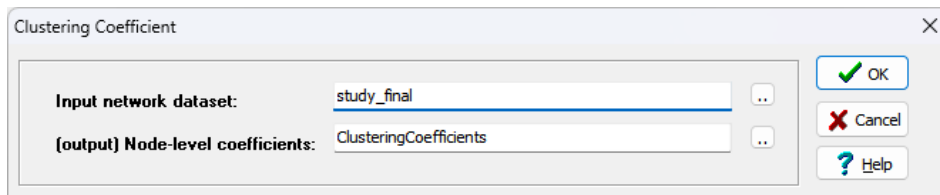
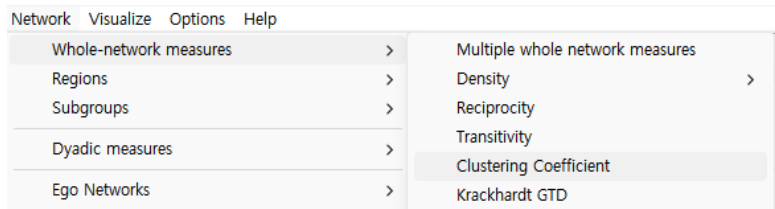
1	# of nodes	43
2	# of ties	188
3	Avg Degree	4.372
4	Indeg H-Index	7
5	K-core index	5
6	Deg Centralization	0.193
7	Out-Centralization	0.186
8	In-Centralization	0.162
9	Indeg Corr	-0.064
10	Outdeg Corr	-0.024
11	Density	0.104
12	Components	3
13	Component Ratio	0.048
14	Connectedness	0.908
15	Fragmentation	0.092
16	Transitivity/Closure	0.277
17	Avg Distance	2.952
18	Prop within 3	0.640
19	SD Distance	1.304
20	Diameter	8
21	Wiener Index	4842
22	Dependency Sum	3202
23	Breadth	0.614
24	Compactness	0.386
25	Small Worldness	
26	Mutuals	0.083
27	Asymmetrics	0.042
28	Nulls	0.875
29	Arc Reciprocity	0.798
30	Dyad Reciprocity	0.664

30 rows, 1 columns, 1 levels.

- 응집도: average degree와 density는 증가.
Connectedness는 감소 ⇒ 왜 이런 결과가 나왔을지 그림을 비교하며 잘 생각해 봅시다
- 집중도: in-centralization과 out-centralization이 모두 증가 ⇒ 소수의 사람이 매우 많이 공동학습하려고 함.
소수의 사람들에게 공동학습이 집중 ⇒ 어떤 의미일지 생각해 보자.
- 호혜성: 거의 변화 없음 ⇒ 서로 학습을 주고 받는 관계는 변함이 없음
- 이행성: 약간 증가 ⇒ 공동학습을 하는 무리(집단)이 약간 형성되어 가는 것으로 해석할 수 있음

- 군집화는 whole network 정도는 중간고사와 기말고사 네트워크 간에 큰 차이는 없다

Network → Whole-network measures → Clustering Coefficient → Input network dataset → OK



Input dataset: study_mid (C:\

Overall graph clustering coefficient: 0.355
 Network density: 0.094
 (cc - density)/cc: 0.736
 Weighted Overall graph clustering coefficient: 0.264
 Small world index: 2.458

Node Clustering Coefficients

		1	2
		Clus Coef	nPairs
1	A1	0.500	10.000
2	A2	0.300	15.000
3	A3	0.194	36.000
4	A4	0.167	3.000

study_mid.##h

Input dataset: study_final (

Overall graph clustering coefficient: 0.302
 Network density: 0.104
 (cc - density)/cc: 0.656
 Weighted Overall graph clustering coefficient: 0.275
 Small world index: 2.398

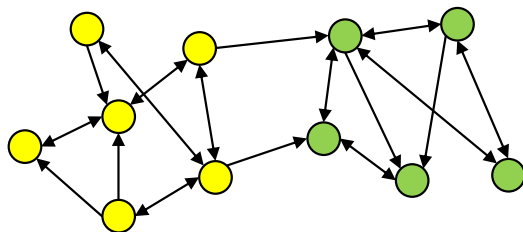
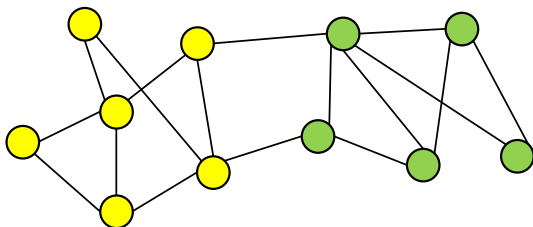
Node Clustering Coefficients

		1	2
		Clus Coef	nPairs
1	A1	0.262	21.000
2	A2	0.350	10.000
3	A3	0.357	21.000
4	A4	0.250	10.000

study_final.##h

- 군집화 정도는 중간고사와 기말고사 네트워크 간에 큰 차이는 없으나 WOCC가 약간 증가했음

- 지금까지는 전체 네트워크의 구조적 특성을 파악
- 전체 네트워크는 노드들이 가지고 있는 속성에 따라 구조화될 수도 있음
 - 예: 주로 같은 성끼리만 상호작용하면, 혹은 성과 무관하게 상호작용하면 전체 네트워크 구조는 어떤 모습일까?
 - 예: 대학생끼리는 조언을 주고 받는 호혜적 관계, 교수끼리는 조언을 주고 받는 호혜적 관계, 대학생이 교수에게 조언을 주는 일방적 관계라면, 전체 네트워크 구조는 어떤 모습일까?
- 집단에 따른 네트워크 구조를 살펴보는 방법
 - ① 유유상종 (동류성)
 - ② 집단 내(간) 호혜성

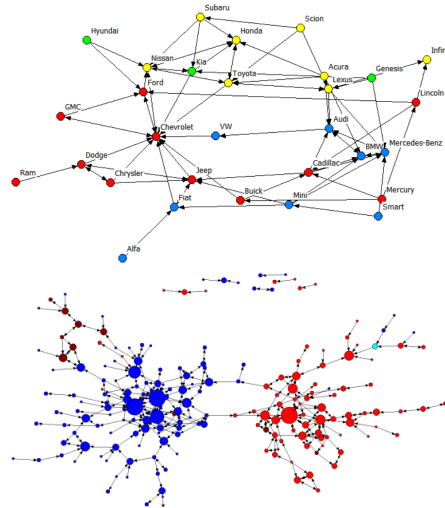


- 유유상종 (Birds of Feather Flock Together)?

- 네트워크는 여러 집단으로 구성 (성, 전공, 인종, 국적, 브랜드, 맛, 시장 세그먼트, 성적, 기업규모...)
- 집단 내 응집성이 높고 집단 간 응집성이 낮다면 집단 간 분리 → 유유상종 (동류성)
- 예: 우리 부서는 연령대가 비슷한 사람끼리 친하다 ⇒ 집단 간 갈등, 대립, 역기능
- 예: 화장품 브랜드 연관검색 네트워크를 보니 고객들은 백화점 입점 브랜드끼리 연관 검색하는 경향이 있다 ⇒ 우리 회사 화장품 브랜드의 마켓 포지션을 어떻게 할 것인가?
- 예: 사람들은 같은 장르의 음악을 듣는 경향이 줄어들고 혼합장르 음악을 듣는 경향이 증가하고 있다 ⇒ 다음에 제작할 음악의 장르 정체성을 어떻게 할 것인가?



People who have similar characters and similar interests will often choose to spend time together.



Social network(acquaintance network) of drug users in Hartford

- 아래의 세 개 네트워크를 비교해 보자.
- 집단 내/간 밀도를 요약해 보자. 무엇을 발견했는지 간단히 요약해 보자.

	A	B	C	D	E	F	G	H	I
A	0	1	1	1	0	0	0	0	0
B	1	0	1	1	0	0	0	0	0
C	1	1	0	1	0	0	0	0	0
D	1	1	1	0	0	0	0	0	0
E	0	0	0	0	0	1	1	0	0
F	0	0	0	0	1	0	1	0	0
G	0	0	0	0	1	1	0	0	0
H	0	0	0	0	0	0	0	0	1
I	0	0	0	0	0	0	0	1	0

Network 1: Perfect Homophily
(ex_homo.##h)

집단	1	2	3
1	1.00	0.00	0.00
2	0.00	1.00	0.00
3	0.00	0.00	1.00

	A	B	C	D	E	F	G	H	I
A	0	0	0	0	1	1	1	1	1
B	0	0	0	0	1	1	1	1	1
C	0	0	0	0	1	1	1	1	1
D	0	0	0	0	1	1	1	1	1
E	1	1	1	1	0	0	0	1	1
F	1	1	1	1	0	0	0	1	1
G	1	1	1	1	0	0	0	1	1
H	1	1	1	1	1	1	1	0	0
I	1	1	1	1	1	1	1	0	0

Network 2: Perfect Heterophily
(ex_hetero.##h)

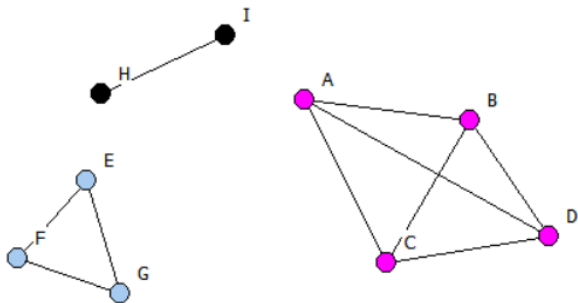
집단	1	2	3
1	0.00	1.00	1.00
2	1.00	0.00	1.00
3	1.00	1.00	0.00

	A	B	C	D	E	F	G	H	I
A	0	1	0	1	0	1	0	0	1
B	1	0	0	0	1	1	0	1	0
C	0	0	0	0	0	0	1	0	0
D	0	1	1	0	0	1	0	0	1
E	0	1	0	0	0	0	1	0	1
F	0	0	1	0	1	0	0	0	0
G	0	1	0	0	0	0	0	1	1
H	0	0	1	1	0	0	1	0	1
I	0	1	1	0	0	1	0	0	0

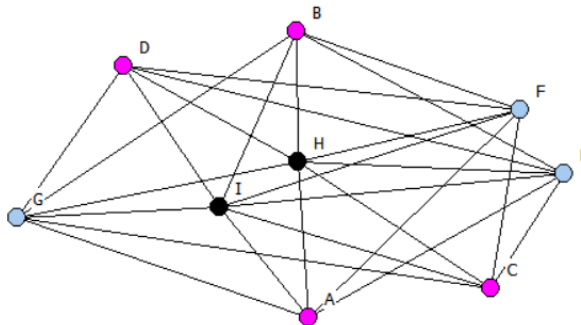
Network 3: random (no pattern)
(ex_random.##h)

집단	1	2	3
1	0.42	0.42	0.38
2	0.25	0.33	0.50
3	0.50	0.33	0.50

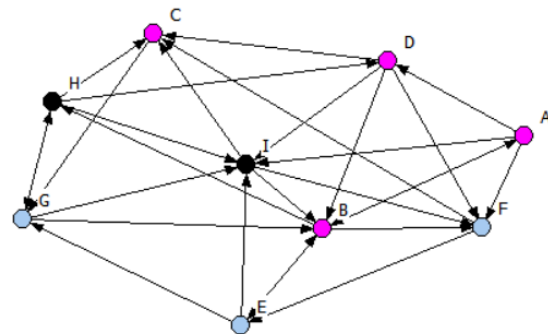
- 조금 어렵다면 그림을 그려서 확인해 보자



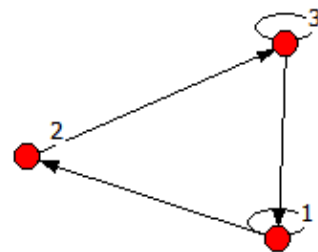
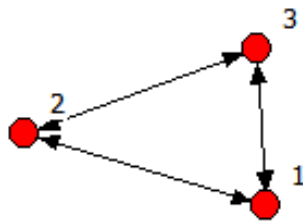
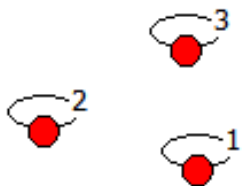
Network 1: Perfect Homophily
(`ex_homo.##h`)



Network 2: Perfect Heterophily
(`ex_hetero.##h`)



Network 3: Mixed (no pattern)
(`ex_mixed.##h`)



tie strength > 0.4

- 유유상종 (동류성)의 측정방법

- 집단 내에 존재하는 연결과 집단 간에 존재하는 연결의 분포를 비교

	Within Group	Between Group
Tie	a	b
No Tie	c	d

- E-I Index

- Internal Tie의 수와 External Tie의 수를 가지고 계산 $\Rightarrow E-I \text{ Index} = \frac{EL-IL}{EL+IL} = \frac{b-a}{b+a}$
 - $-1(\text{완벽한 유유상종}) \leq E-I \text{ Index} \leq 1(\text{완벽한 이질성})$
 - 단점: E-I index는 집단의 수나 규모에 민감 (예: 집단이 많아지면 EL가 자연스럽게 증가하므로 실제로 유유상종이 있다하더라도 E-I index는 0보다 커질 수 있음)
 - 이를 방지하기 위해 무작위 그래프에서 생길 수 있는 expected E-I Index와 실제 E-I Index를 비교하는 **permutation 방식을 사용**

	A	B	C	D	E	F	G	H	I
A	0	1	1	1	0	0	0	0	0
B	1	0	1	1	0	0	0	0	0
C	1	1	0	1	0	0	0	0	0
D	1	1	1	0	0	0	0	0	0
E	0	0	0	0	0	1	1	0	0
F	0	0	0	0	1	0	1	0	0
G	0	0	0	0	1	1	0	0	0
H	0	0	0	0	0	0	0	0	1
I	0	0	0	0	0	0	0	1	0

Network 1: Perfect Homophily
([ex_homo.##h](#))

	A	B	C	D	E	F	G	H	I
A	0	0	0	0	1	1	1	1	1
B	0	0	0	0	1	1	1	1	1
C	0	0	0	0	1	1	1	1	1
D	0	0	0	0	1	1	1	1	1
E	1	1	1	1	0	0	0	1	1
F	1	1	1	1	0	0	0	1	1
G	1	1	1	1	0	0	0	1	1
H	1	1	1	1	1	1	1	0	0
I	1	1	1	1	1	1	1	0	0

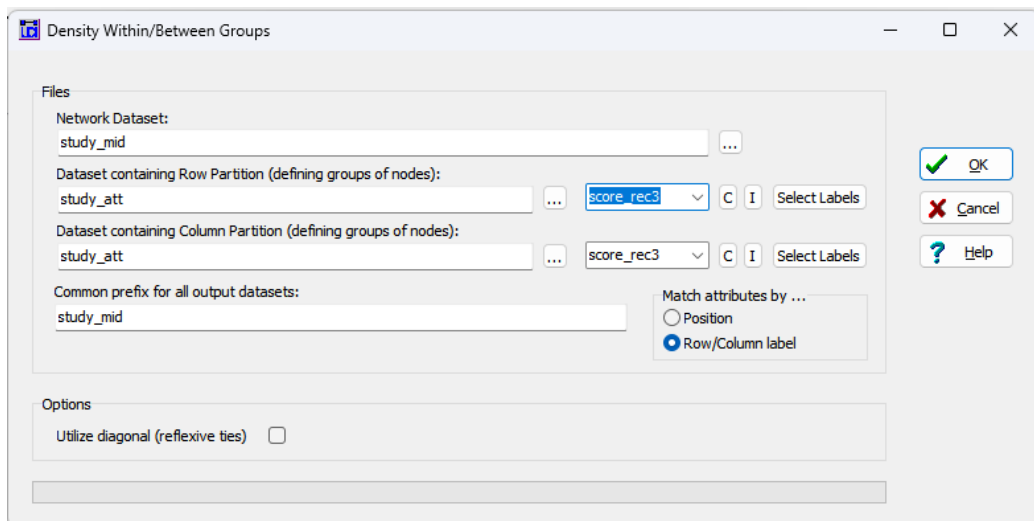
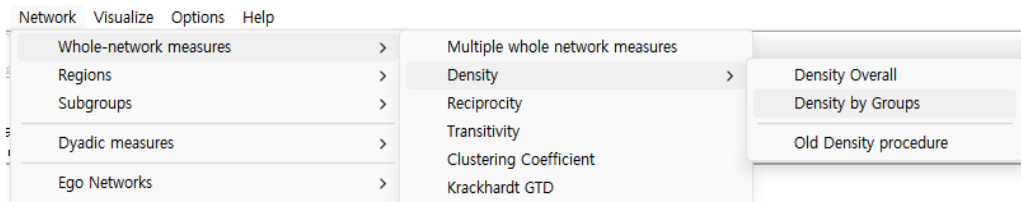
Network 2: Perfect Heterophily
([ex_hetero.##h](#))

	A	B	C	D	E	F	G	H	I
A	0	1	0	1	0	1	0	0	1
B	1	0	0	0	1	1	0	1	0
C	0	0	0	0	0	0	1	0	0
D	0	1	1	0	0	1	0	0	1
E	0	1	0	0	0	0	1	0	1
F	0	0	1	0	1	0	0	0	0
G	0	1	0	0	0	0	0	1	1
H	0	0	1	1	0	0	1	0	1
I	0	1	1	0	0	1	0	0	0

Network 3: random (no pattern)
([ex_random.##h](#))

- Network 1의 $E-I = \frac{0-20}{0+20} = -1$
- Network 2의 $E-I = \frac{52-0}{52+0} = 1$
- Network 3의 $E-I = \frac{20-8}{20+8} = 0.43$

- 중간고사 성적집단 (상, 중, 하 3개 집단: score_rec3) 에 따른 유유상종이 나타나는지 살펴보자
- 유유상종을 알아보기 전에 먼저 집단 내/간 밀도를 알아보자



*Network → Whole-network measures →
Density → Density by Groups → Network
Dataset에 네트워크 데이터를 넣고 → Dataset
containing...에 속성 데이터를 넣은 후 분할할
변수 선택 (여기서는 score_rec3) → OK*

- 중간고사 성적집단 (상, 중, 하 3개 집단: score_rec3) 에 따른 유유상종이 나타나는지 살펴보자
- 유유상종을 알아보기 전에 먼저 집단 내/간 밀도를 알아보자

Number of ties			
	1	2	3
1	23	24	17
2	21	25	12
3	15	14	18

3 rows, 3 columns, 1 levels.

Density of ties			
	1	2	3
1	0.126	0.114	0.087
2	0.100	0.119	0.057
3	0.077	0.067	0.099

3 rows, 3 columns, 1 levels.

Autocorrelation: 0.078

Ties within each group

	1	2
Number		
Densit		
y		
1	23.000	0.126
2	25.000	0.119
3	18.000	0.099

3 rows, 2 columns, 1 levels.

study_mid

- 1(상), 2(중), 3(하) 집단 내/간에 존재하는 연결의 수
- 예: 중간고사 전에는 상-상 집단 간 연결수는 23, 중간고사 이후에는 45개로 증가

- 1(상), 2(중), 3(하) 집단 내/간 연결 밀도
- 예: 중간고사 전에는 상-상 집단 간 연결밀도는 0.126, 중간고사 이후에는 0.247로 증가

- 중간고사와 기말고사 사이에 어떤 변화가 있었는지 정리해 보자

Number of ties			
	1	2	3
1	45	29	8
2	31	36	10
3	7	8	14

3 rows, 3 columns, 1 levels.

Density of ties			
	1	2	3
1	0.247	0.138	0.041
2	0.148	0.171	0.048
3	0.036	0.038	0.077

3 rows, 3 columns, 1 levels.

Autocorrelation: 0.228

Ties within each group

	1	2
Number		
Densit		
y		
1	45.000	0.247
2	36.000	0.171
3	14.000	0.077

3 rows, 2 columns, 1 levels.

study_final

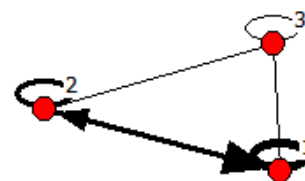
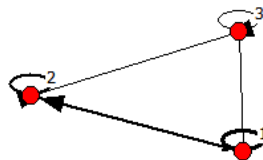
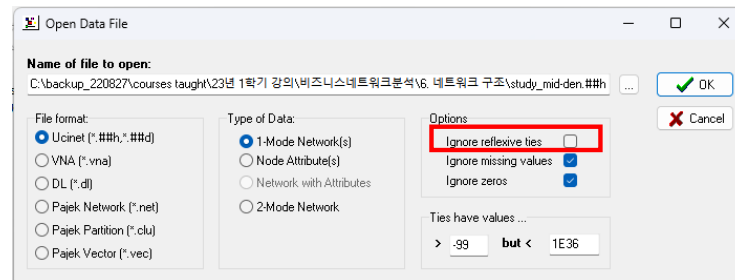
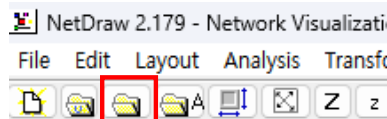
- 그림을 그려 확인해 보자
- 그림을 그리기 전에 먼저 저장된 집단내/간 연결밀도 데이터를 확인해 보자.
- 이 데이터를 NetDraw로 그려보자. (주의: 반드시 아래와 같이 세번째에 있는 아무 것도 쓰여지지 않은 서류함을 선택하고, 대화창에서 ignore reflexive ties를 해제해 주어야 함)

	1	2	3
1	0.12637363	0.11428571	0.08673469
2	0.10000000	0.11904761	0.05714285
3	0.07653061	0.06666667	0.09890110

study_mid-den.##h

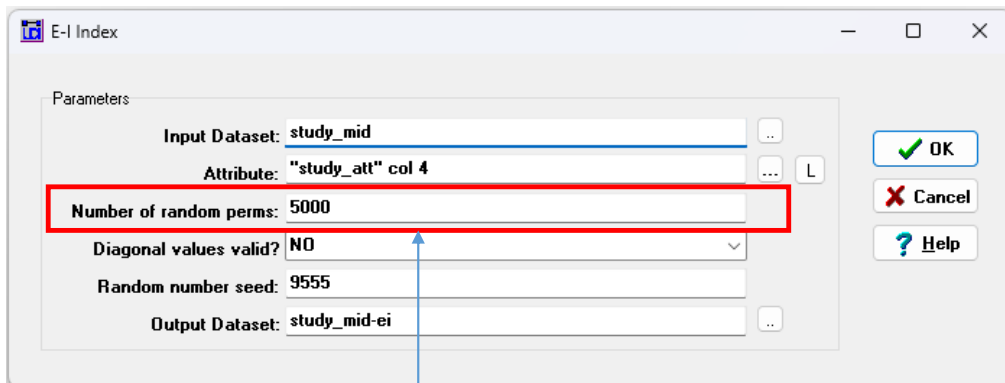
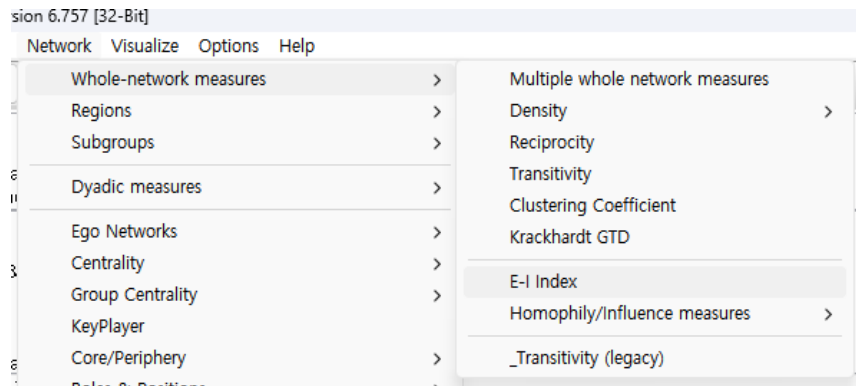
	1	2	3
1	0.24725274	0.13809524	0.04081632
2	0.14761905	0.17142857	0.04761904
3	0.03571428	0.03809523	0.07692307

study_final-den.##h



- 성적에 따른 유유상종이 증가한 것으로 보임
- 그렇다면 진짜 증가한 것인지 알아보기 위해 E-I Index를 구해서 비교해 보자

Network → Whole-network measures → E-I Index → Input Dataset에 네트워크 데이터를 넣고 → Attribute에 속성 데이터를 넣은 후 유유상종을 알아보기 위한 속성 변수의 column 위치를 넣어줌 (여기서는 네번째 컬럼에 score_rec30이 있음) → OK



무작위 네트워크 (random network)를 5000번 만들어 Expected E-I Index를 구한다는 의미임

E-I INDEX

```
-----
Adjacency dataset: :      study_mid (C:\ba
Attribute: :            "study_att" col
# of Permutations: 5000
Random seed: :          20913
Individual E-I scores: : study_mid-ei
```

Warning: This procedure ignores direction of ties.

Warning: Row Attribute vector has been recoded.
Here is a translation table:

Old Code		New Code
1	=>	1
2	=>	2
3	=>	3

Density matrix

		1	2	3
	1			
	2			
	3			

202 ties.

중간고사 E-I Index

Whole Network Results

		1 Freq	2 Pct	3 Possible	4 Density
1	Internal	78.000	0.386	574.000	0.136
2	External	124.000	0.614	1232.000	0.101
3	E-I	46.000	0.228	658.000	0.364

Max possible external ties: 1232.000

Max possible internal ties: 574.000

E-I Index: 0.228

Expected value for E-I index is: 0.364

Max possible E-I given density & group sizes: 1.000

Min possible E-I given density & group sizes: -1.000

Re-scaled E-I index: 0.228

Permutation Test

Number of iterations = 5000

		1 Obs	2 Min	3 Avg	4 Max	5 SD	6 P >= Ob	7 P <= Ob
1	Internal	0.386	0.188	0.317	0.515	0.044	0.079	0.948
2	External	0.614	0.485	0.683	0.812	0.044	0.948	0.079
3	E-I	0.228	-0.030	0.366	0.624	0.088	0.948	0.079

Group level E-I Index

		1 Intern	2 Extern	3 Total	4 E-I
1	1	26.000	46.000	72.000	0.278
2	2	32.000	43.000	75.000	0.147
3	3	20.000	35.000	55.000	0.273

Internal Tie 수와 External Tie 수 확인

Expected E-I Index = 0.364

실제 E-I Index = 0.228

실제 < Expected 이므로 homophily

어떤 집단의 homophily가 가장 높은지
확인 ⇒ 2번 (즉, 성적 중)이 가장 높음

E-I INDEX

```
Adjacency dataset: : study_final
Attribute: : "study_att"
# of Permutations: 5000
Random seed: : 4426
Individual E-I scores: : study_final
```

Warning: This procedure ignores direction of ties.

Warning: Row Attribute vector has been recoded.
Here is a translation table:

Old Code		New Code
1	=>	1
2	=>	2
3	=>	3

Density matrix

	1	2	3
1	0.297	0.171	0.046
2	0.171	0.210	0.048
3	0.046	0.048	0.099

226 ties.

중간고사 E-I Index

Whole Network Results

		1 Freq	2 Pct	3 Possible	4 Density
1	Internal	116.000	0.513	574.000	0.202
2	External	110.000	0.487	1232.000	0.089
3	E-I	-6.000	-0.027	658.000	0.364

Max possible external ties: 1232.000

Max possible internal ties: 574.000

E-I Index: -0.027

Expected value for E-I index is: 0.364

Max possible E-I given density & group sizes: 1.000

Min possible E-I given density & group sizes: -1.000

Re-scaled E-I index: -0.027

Permutation Test

Number of iterations = 5000

		1 Obs	2 Min	3 Avg	4 Max	5 SD	6 P >= Ob	7 P <= Ob
1	Internal	0.513	0.195	0.318	0.504	0.039	0.000	1.000
2	External	0.487	0.496	0.682	0.805	0.039	1.000	0.000
3	E-I	-0.027	-0.009	0.363	0.611	0.079	1.000	0.000

E-I Index is significant ($p < 0.05$)

Group level E-I Index

		1 Intern	2 Extern	3 Total	4 E-I
1	1	54.000	45.000	99.000	-0.091
2	2	44.000	46.000	90.000	0.022
3	3	18.000	19.000	37.000	0.027

Internal Tie 수와 External Tie 수 확인

Expected E-I Index = 0.364

실제 E-I Index = -0.027

실제 < Expected 이므로 homophily

유의미한 수준의 homophily

어떤 집단의 homophily가 가장 높은지
확인 ⇒ 1번 (즉, 성적 상)이 가장 높음