

# 네트워크 분석

소프트웨어융합캡스톤디자인

산업경영공학과 박혜원

## 1 네트워크 분석 공부

1-1. 그래프 이론

## 2 Bootnet 관련 논문 읽기

2-1. Bootstrap  
2-2. Edge Weight CI 추정  
2-3. 중심성 지수 안정성 평가

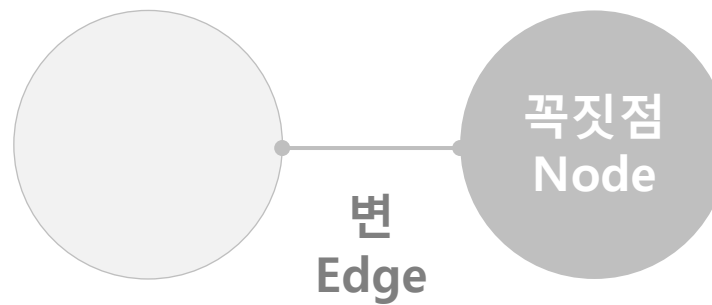
## 3 R Bootnet 사용

3-1. R package 사용

## 그래프란?

점과 그 점을 연결하는 곡선의 집합

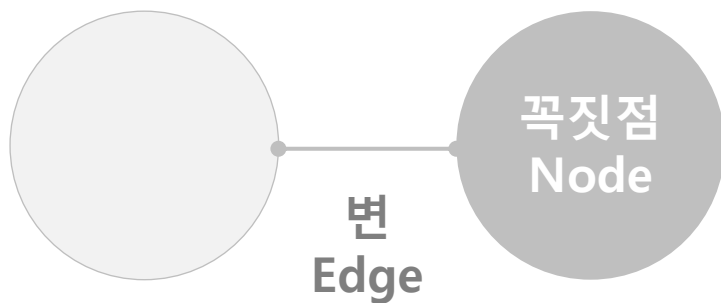
곡선 모양, 길이, 점의 위치에는 의미가 없으며,  
두 점의 직접적 연결 여부에만 관심을 가짐



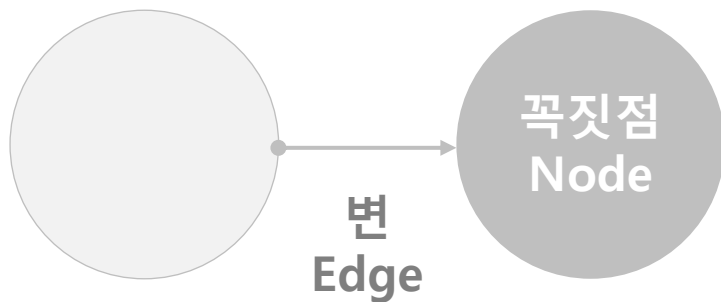
Node = Vertex

Edge = Link

자주 혼용되어 쓰인다!



두 점을  $A_i, A_j$  로 나타내면  
무향 그래프에서는  $A_i A_j$ 와  $A_j A_i$ 가 같지만



방향이 있는 유향 그래프에서는  
 $A_i A_j$ 와  $A_j A_i$ 가 달라진다.

인적 네트워크를 분석할 때 사람을 점으로, 관계를 변으로 나타내면  
무향 그래프에서는 방향이 없으므로 서로 아는 사이만 나타낼 수 있으나  
유향 그래프를 사용하면 한 쪽만 아는 사이 등 더욱 풍부한 관계를 표현할 수 있다.



가중 네트워크는 Edge에 가중치가 주어진 네트워크를 뜻한다.

단순히 '안다', '모른다'의 개념이 아닌,

'얼마나 친한가?' 를 연구할 때는 가중치를 사용할 수 있고,

Node간의 거리는 가중치의 역수로 구할 수 있다.

 A -----  B -----  C

 A ---15---  B -----2-----  C

차수(Degree): 점  $A_i$ 의 차수는  $A_i$ 를 끝점으로 가지는 변의 개수로 계산할 수 있다.

내차수(indegree): 점으로부터 시작하는 변의 수로 계산할 수 있다.

외차수(outdegree): 내차수와 반대로 점에서 끝나는 변의 수로 계산할 수 있다.

그룹화: 네트워크에서 동일한 경로 갖거나 인접한 Node들끼리 그룹화하여 분석할 수 있음

비슷한 특성 지닌 그룹 구분할 때 유용하게 사용 가능한 기법

## 중심성 (Centrality)

Node의 중요도를 평가하는 여러 지표를 통틀어 지칭하는 개념

### 연결 중심성

Degree의 개념으로 설명 가능  
단순히 하나의 Node에 얼마나 많은 Edge가  
연결 되었는지 보는 것

Hub: Node 중 가장 높은 연결중심성 가지는 것

### 근접 중심성

한 Node와 다른 Node들 사이의 거리 계산함으  
로써 가장 최단 거리 갖는 Node 찾아내는 것

거리: 출발점에서 도착점까지의 지난 점의 개수  
를 통해 알 수 있음

### 매개 중심성

소수의 Edge만으로도 다른 Node들 사이를 매  
개하는 중재자를 알아내기 위해 사용

Node와 Node 사이를 최단 경로로 연결해서  
다리 역할을 하는 Node 찾는 것



<https://velog.io/@mymelody/그래프-이론과-네트워크-분석>

심리 네트워크의 사용은 다양한 연구 분야에서 인기를 얻고 있다.

일반적인 논문에서는 심리 네트워크를 추정 및 해석하는 주제를 다루지만,  
**얼마나 정확한 네트워크가 추정되고,**  
**또 얼마나 안정적인지** 확인하고자 하는 작업은 거의 수행되지 않는다.

-> 심리 네트워크의 정확성, 안정성, 견고성을 평가하는 방법을 제안하는  
**R package bootnet**

Bootnet에서는  
샘플링 변동을 통해 네트워크 정확도 문제를 해결하는 것을 목표로 한다.

통계학에서의 Bootstrap은 현재 내가 가진 데이터들 중  
데이터를 무작위로 재추출함으로써 자료를 생성하고,  
그로부터 통계량을 구해내는 것을 뜻한다.  
중복을 허용한다.

즉, 데이터들 중에서 **n**개를 복원추출하고 그들의 평균을 구한다.  
이 과정을 수 천 번 반복하면 평균의 분포를 알 수 있게 되며  
이를 통해 정확도가 높은 통계량들을 추론해낼 수 있다.

매우 많은 수의 매개 변수를 필요로 하기 때문

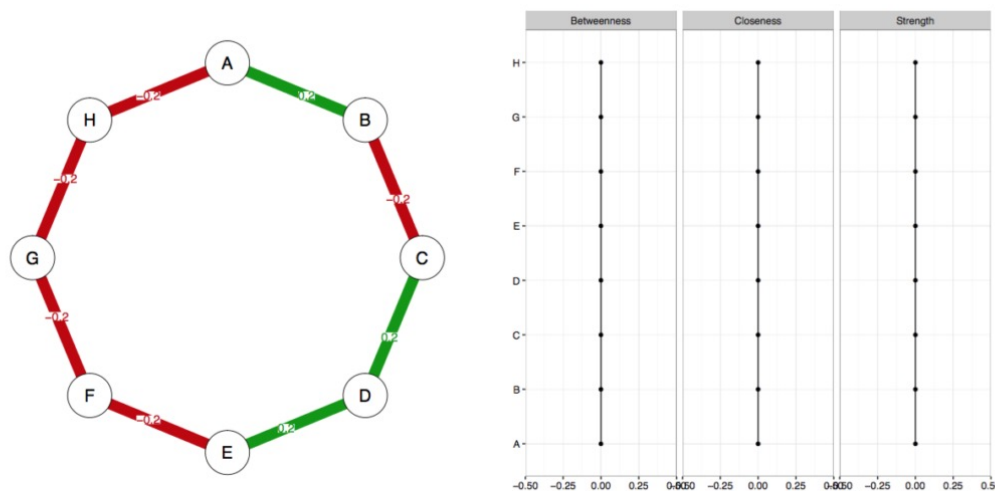


Figure 1. Simulated network structure (left panel) and the importance of each node quantified in *centrality indices* (right panel). The simulated network is a chain network in which each edge has the same absolute strength. The network model used was a Gaussian graphical model in which each edge represents partial correlation coefficients between two variables after conditioning on all other variables.

실제 네트워크 구조

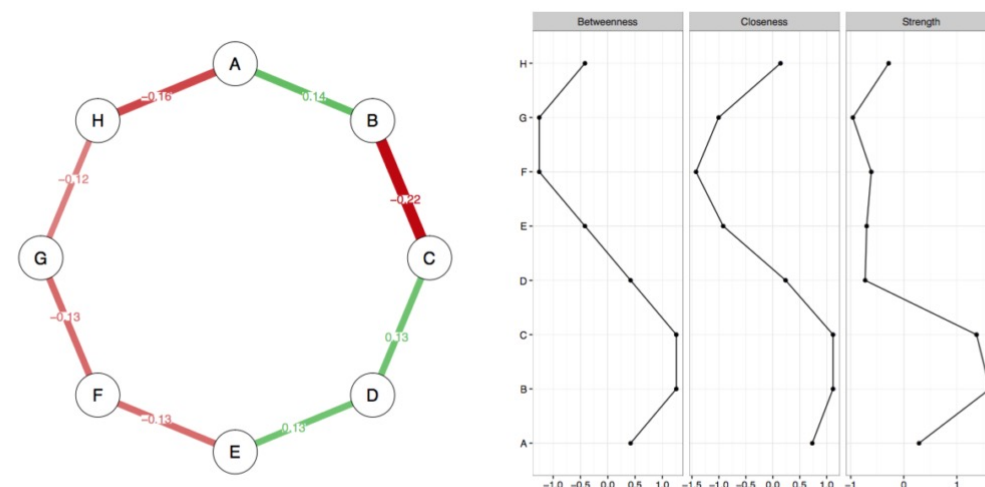


Figure 2. Estimated network structure based on a sample of 500 people simulated using the true model shown in Figure 1 (left panel) and computed centrality indices (right panel). Centrality indices are shown as standardized *z*-scores. Centrality indices show that nodes B and C are the most important nodes, even though the true model does not differentiate in importance between nodes.

n=500 시뮬레이션 구조

(a) 참여자 (데이터) 가 적거나  
(b) Node가 많거나  
(c) a, b 둘 다일 때

**1단계**

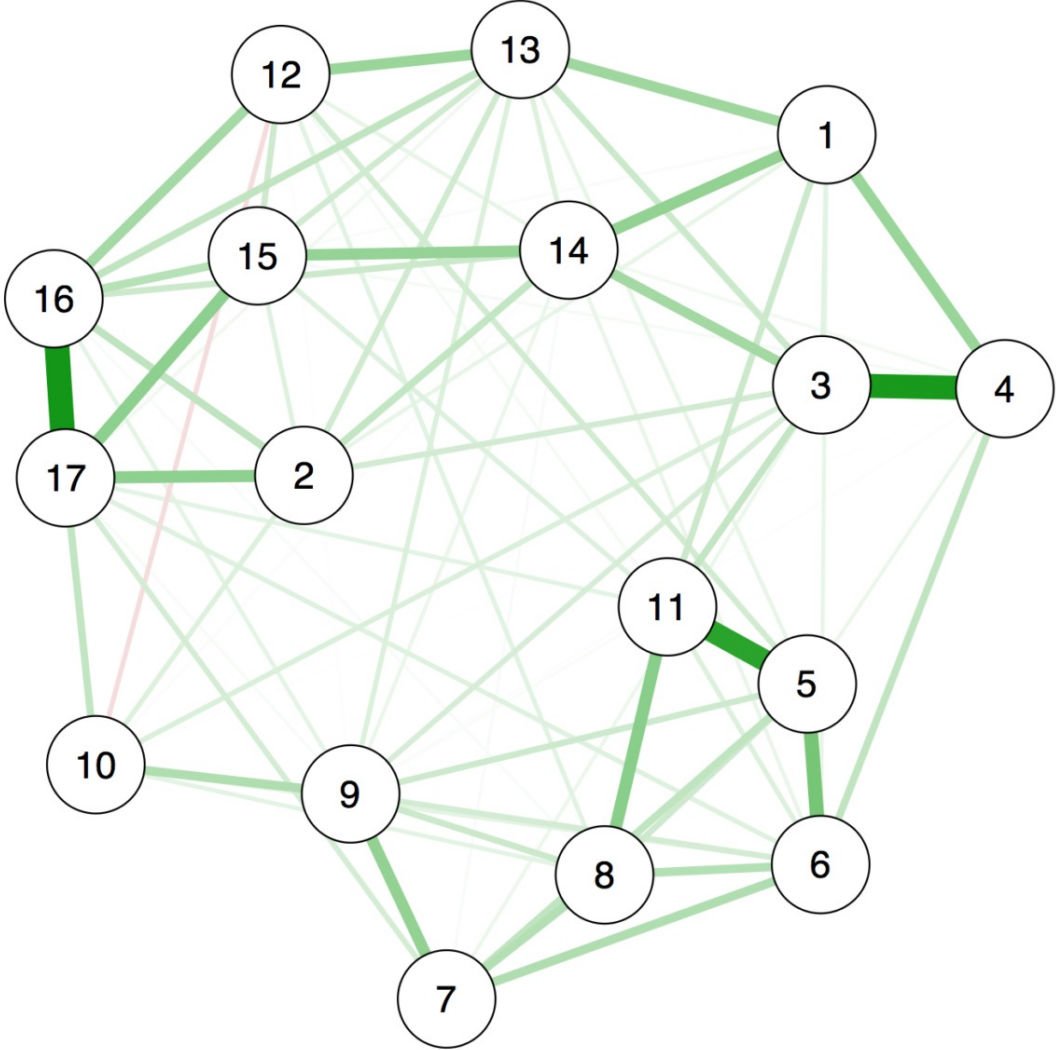
Edge Weight에 대한  
신뢰구간 추정

**2단계**

중심성 지수 안정성 평가

**3단계**

차이 검정

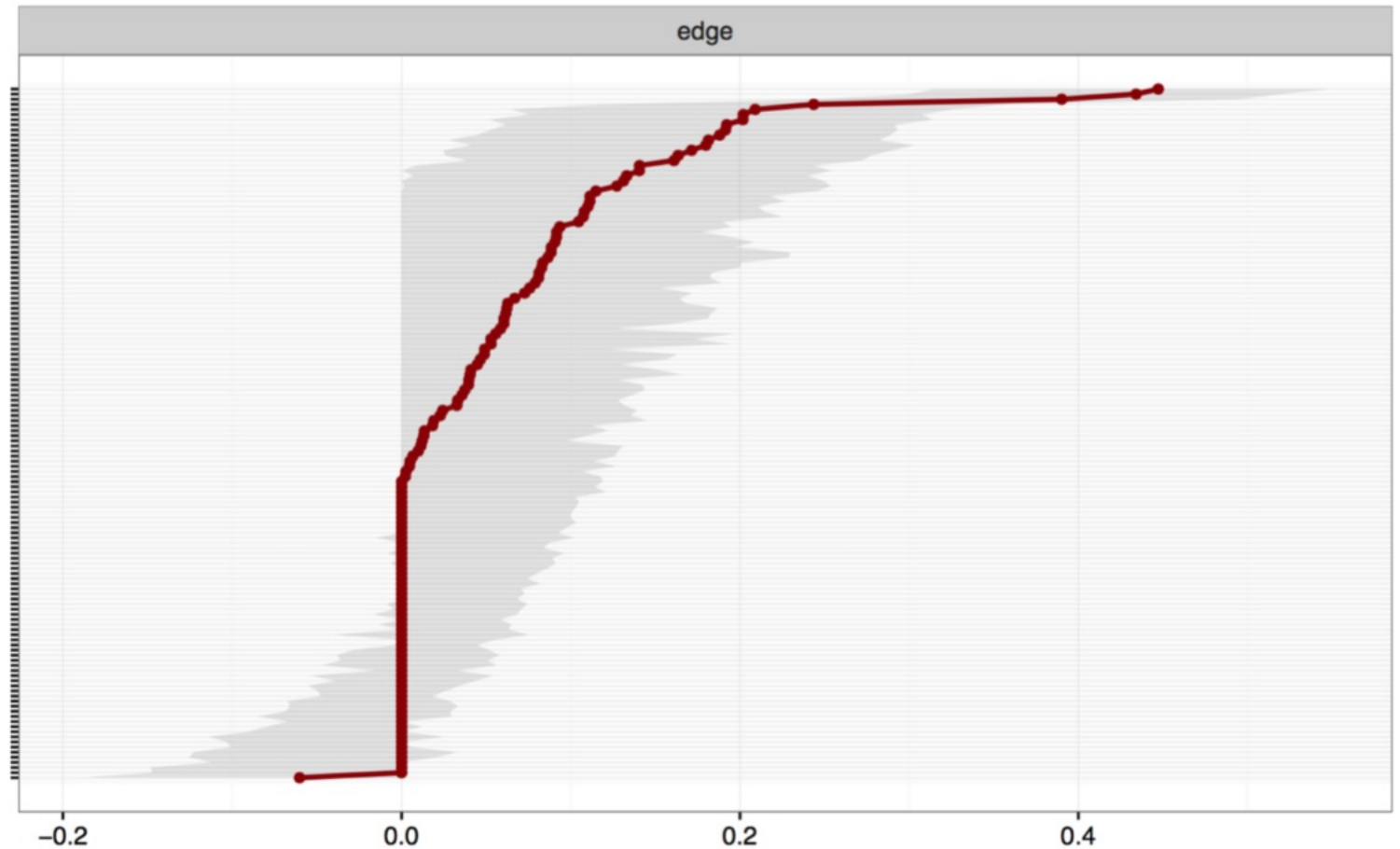


## Edge Weight의 정확도 추정

빨간 점: 네트워크 Edge Weight  
회색 영역: Edge Weight 주변의  
95% 신뢰구간

대부분의 Edge 신뢰구간이 겹치는 것은  
그룹간의 차이가 유의하지 않음을 뜻함

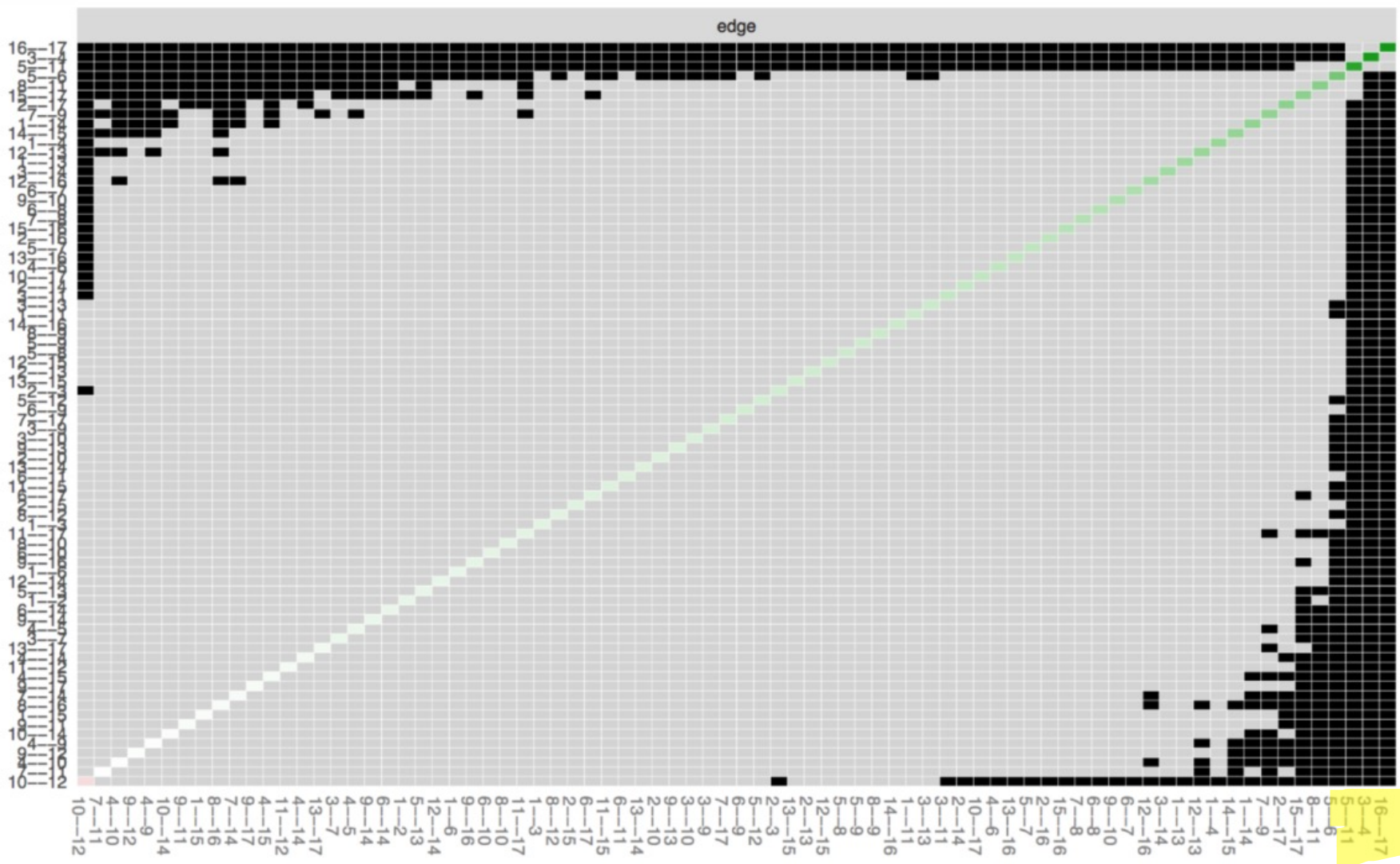
Node수가 고정되어 있다면  
359명 보다 더 많은 참여자  
확보해야 됨을 의미



### <부트스트랩으로 신뢰구간 구하기>

1. 앞서 말한 부트스트랩 방법대로 내 손에 있는 데이터들 중에서  $n$ 개를 복원추출한다.
2. 추출한 표본으로부터 알고 싶은 통계량(평균 등)을 구하는 과정을  $R(>1000)$ 번 반복한다.
3.  $x\%$  신뢰구간을 구하기 위해  $R$ 개의 결과 분포의 양쪽 끝에서  $(100-x)/2\%$  만큼 잘라낸다.
4. 절단한 양 끝점들이  $x\%$  부트스트랩 신뢰구간 양 끝점이 되는 것이다.

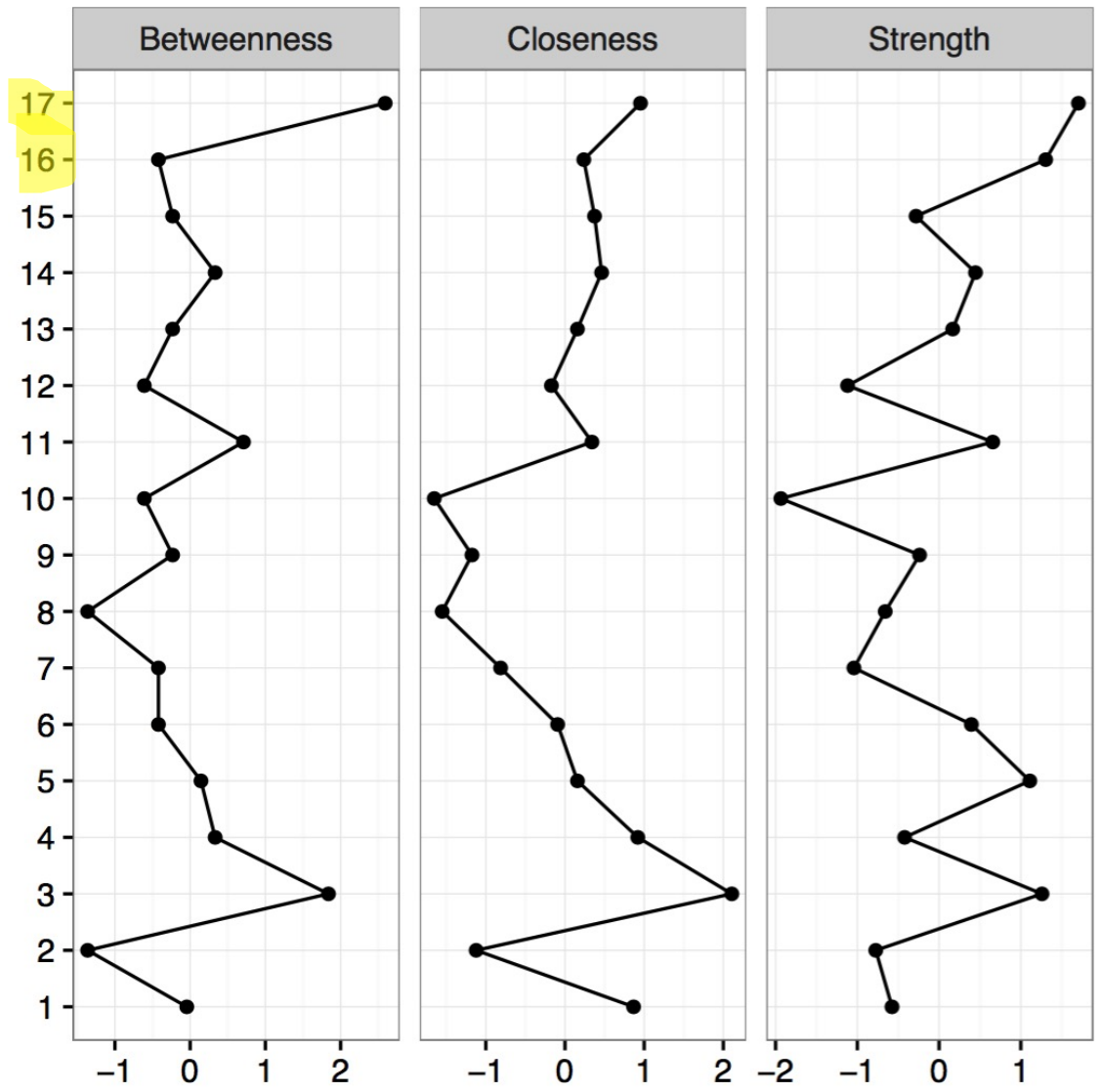




Edge A-B가  
Edge A-C 보다 유의하게 큰지  
알고 싶을 때 사용한다

검은 부분: 두 Edge 사이에  
상당한 차이가 있음을 나타냄

대각선: Edge Weight 값



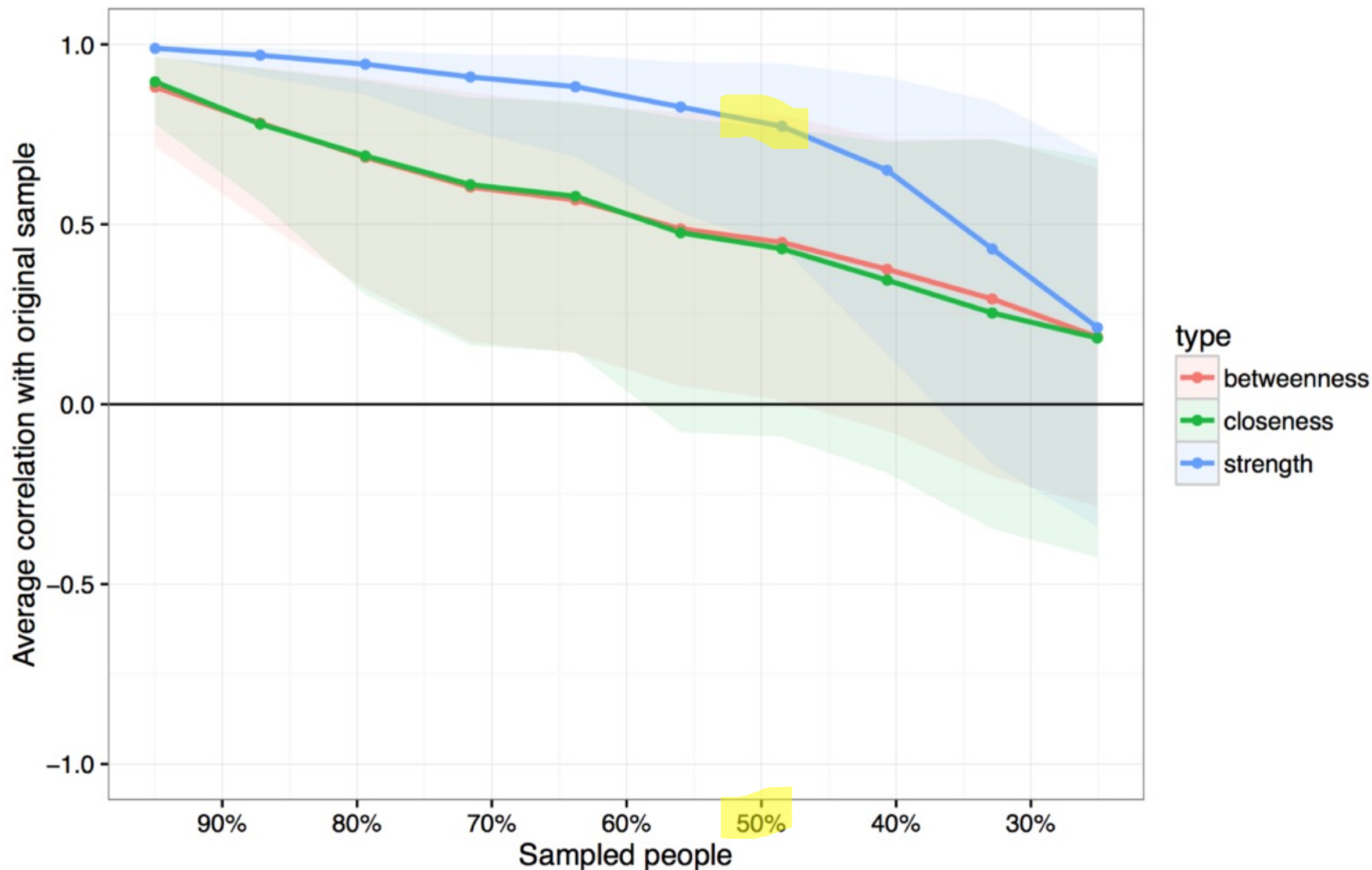
중심성의 정확성을 파악하기 위해서는  
중심성 지수 순서의 안정성을 조사하는 것이 좋다

17번이 16번보다 중심성이 확실히 높다?

특정 Node A가 다른 Node B의 중심성 지수보다  
**유의하게 큰지** 여부를 알고 싶을 때는  
부트스트랩 방법을 사용한다.

데이터를 샘플링 해서 새 네트워크를 구성하고  
중앙값을 다시 추정하는 과정을 수천 번을 반복 한다.

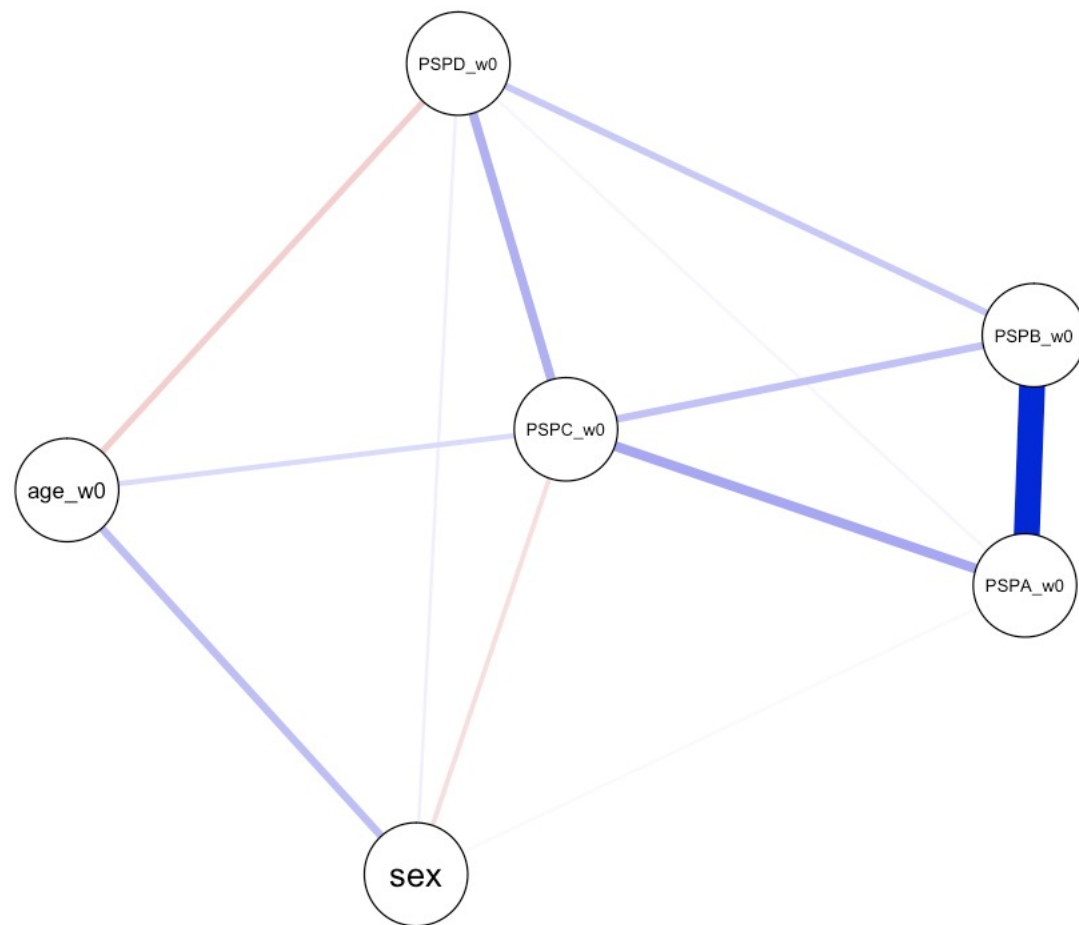
전체 데이터 집합의 중심성 수치의 순서와  
참가자의 50%를 떨어뜨린  
데이터 집합의 중심성 순서가 매우 유사하다면  
중심성의 순서가 안정적이라는 의미다.

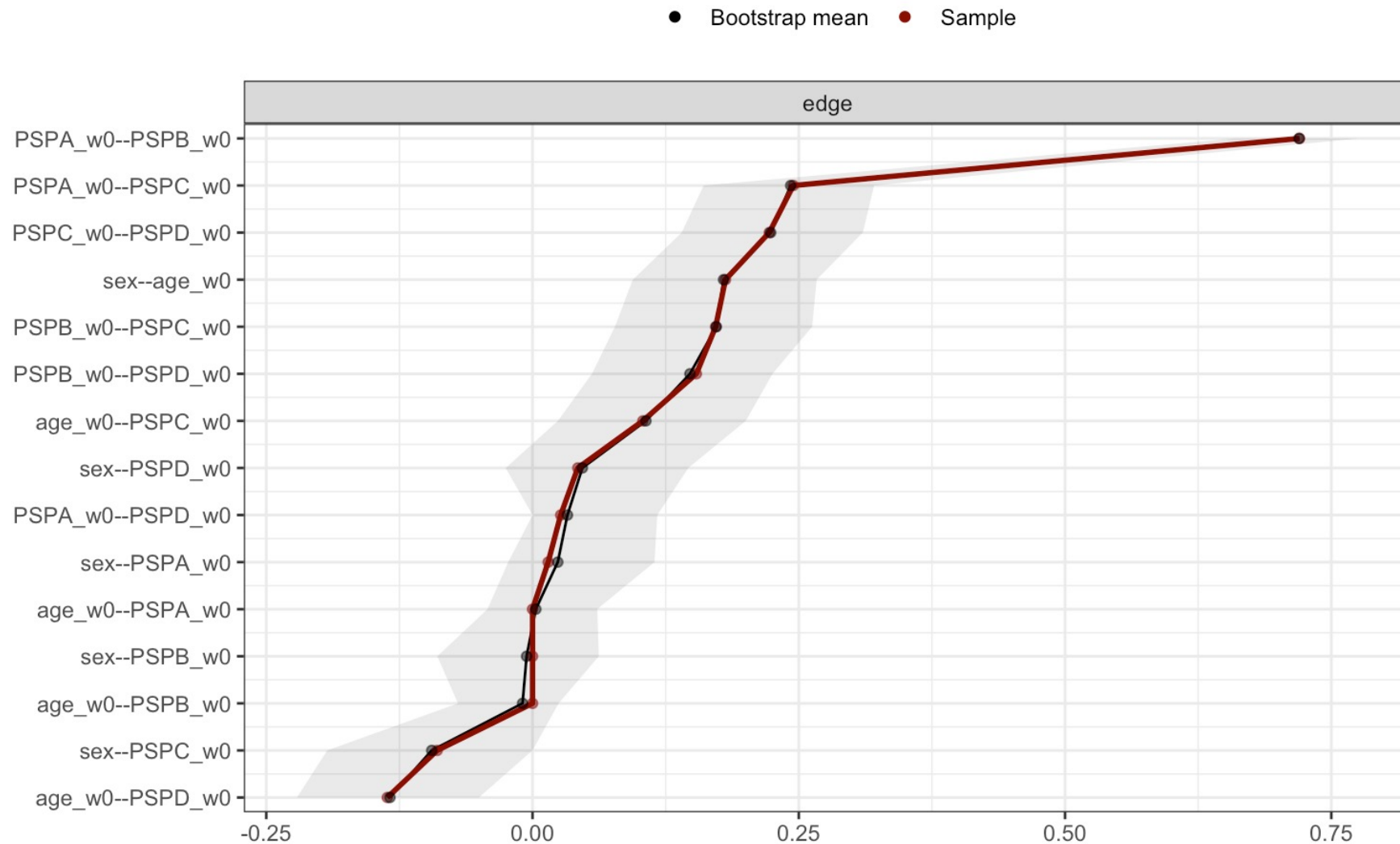


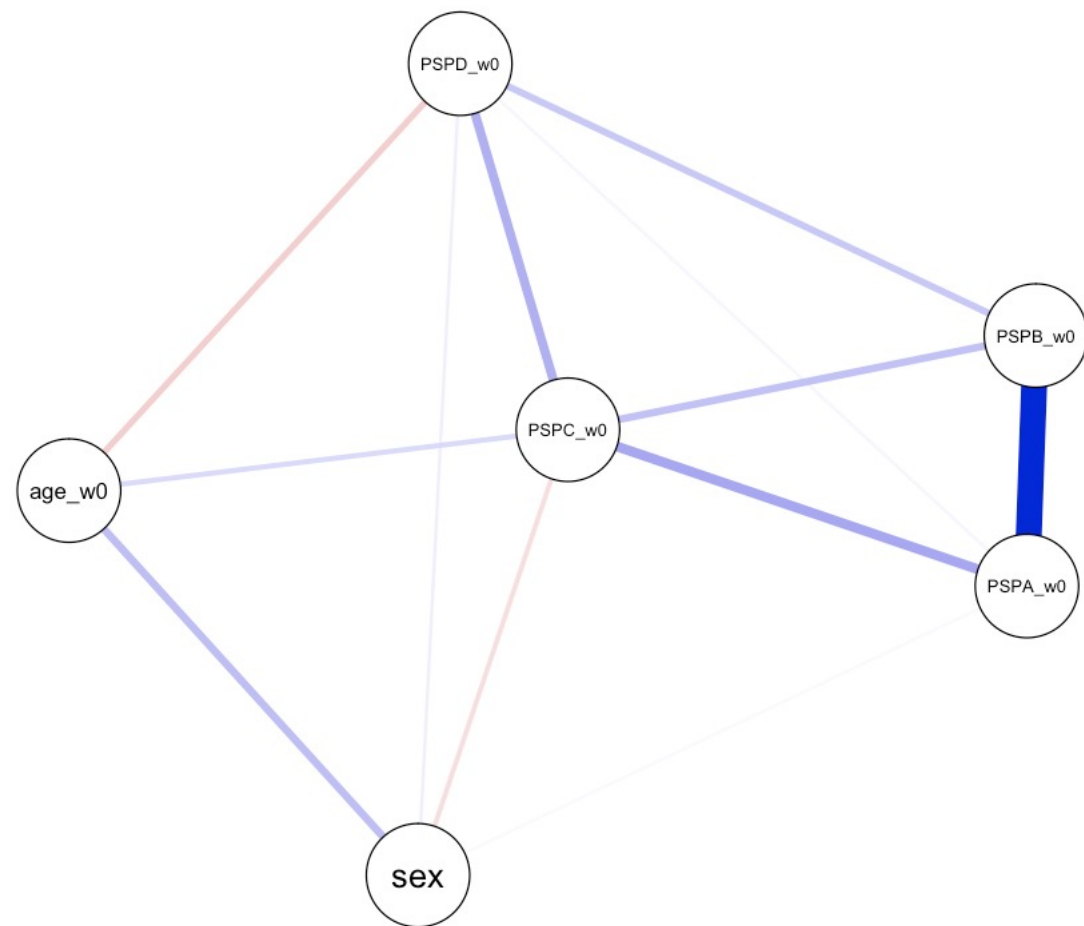
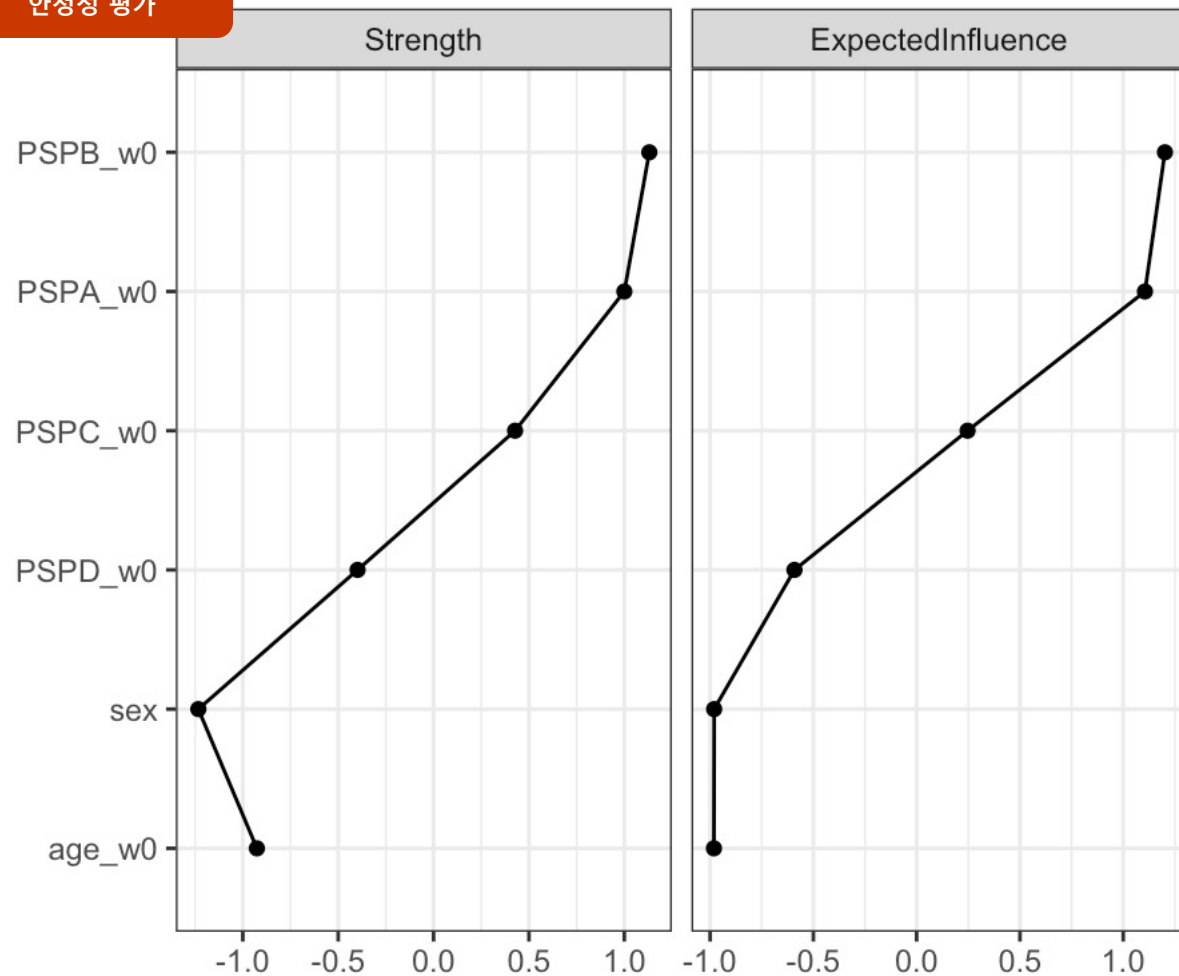


<https://velog.io/@mymelody/심리적-네트워크-및-정확성-평가-논문-공부>

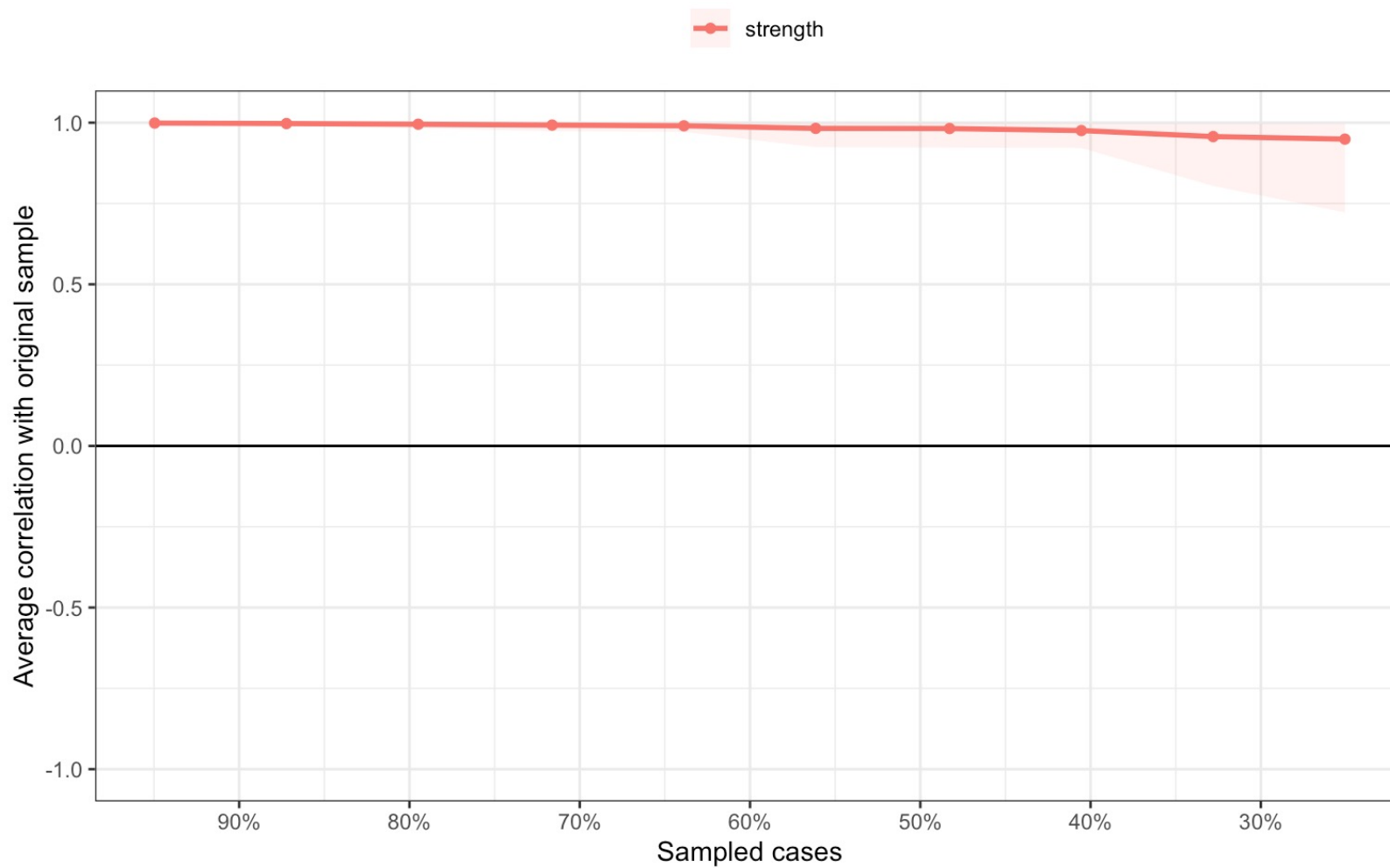
PSP 척도



Edge Weight  
신뢰구간 추정

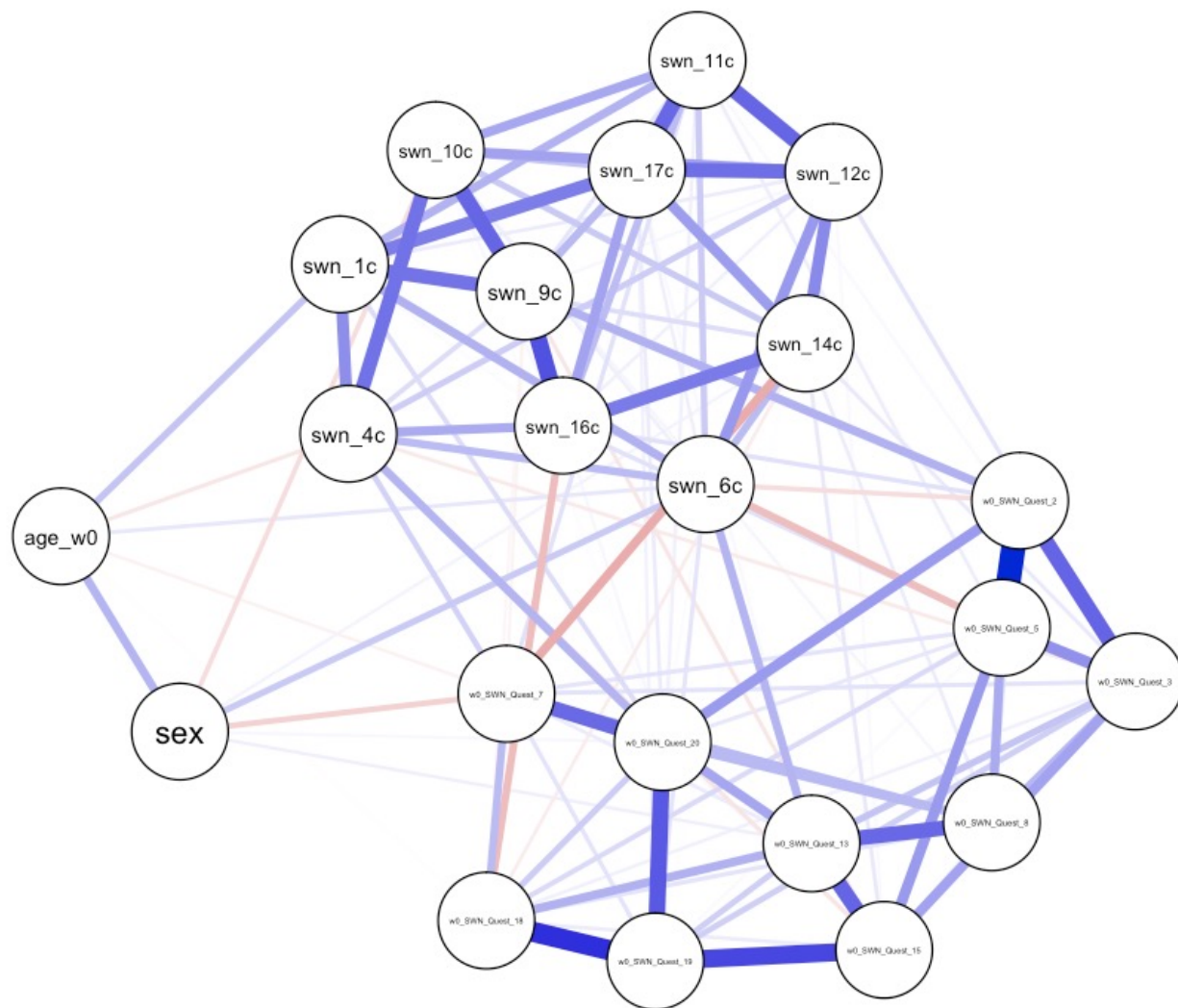
중심성 지수  
안정성 평가

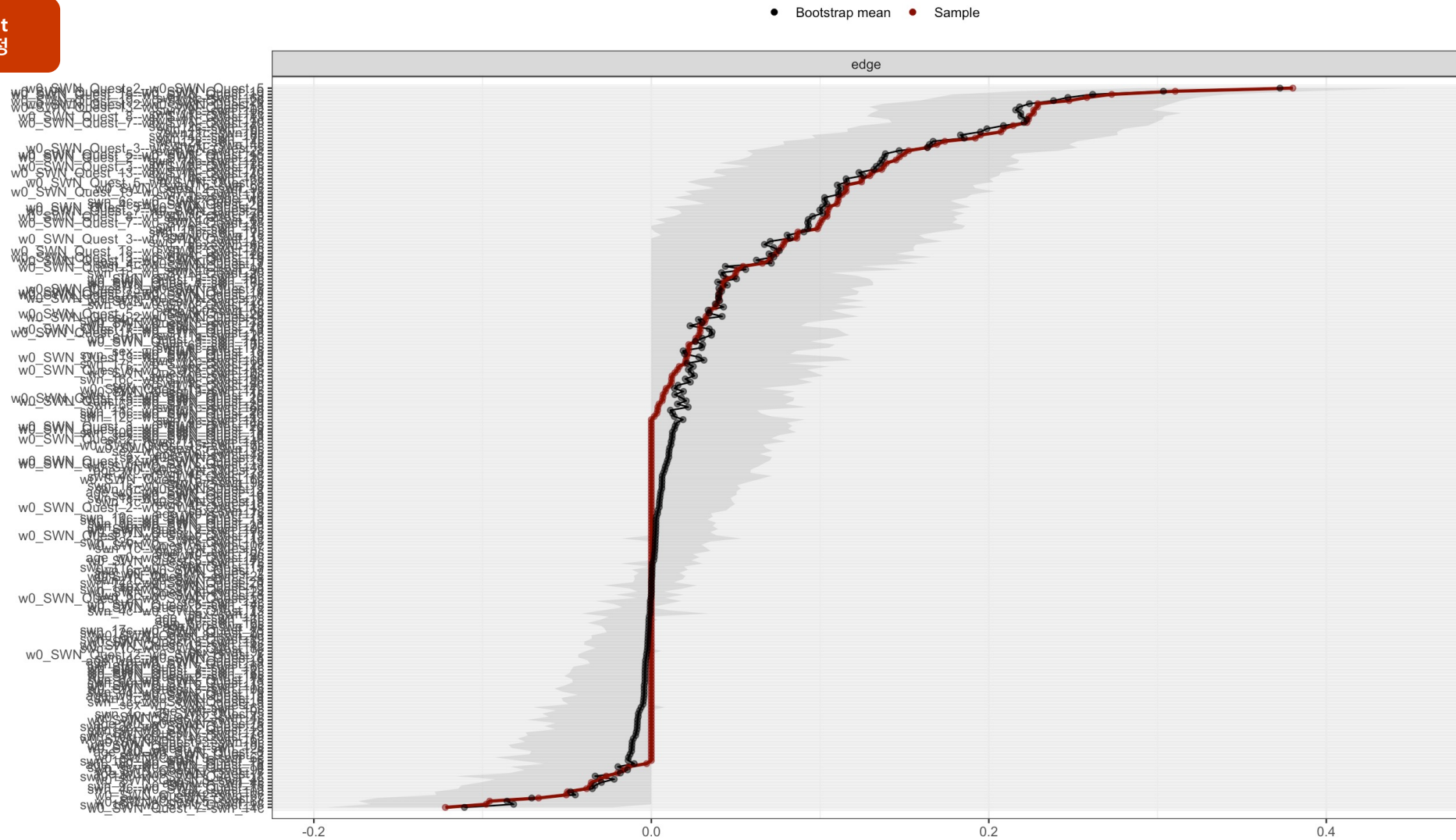
중심성 지수  
안정성 평가



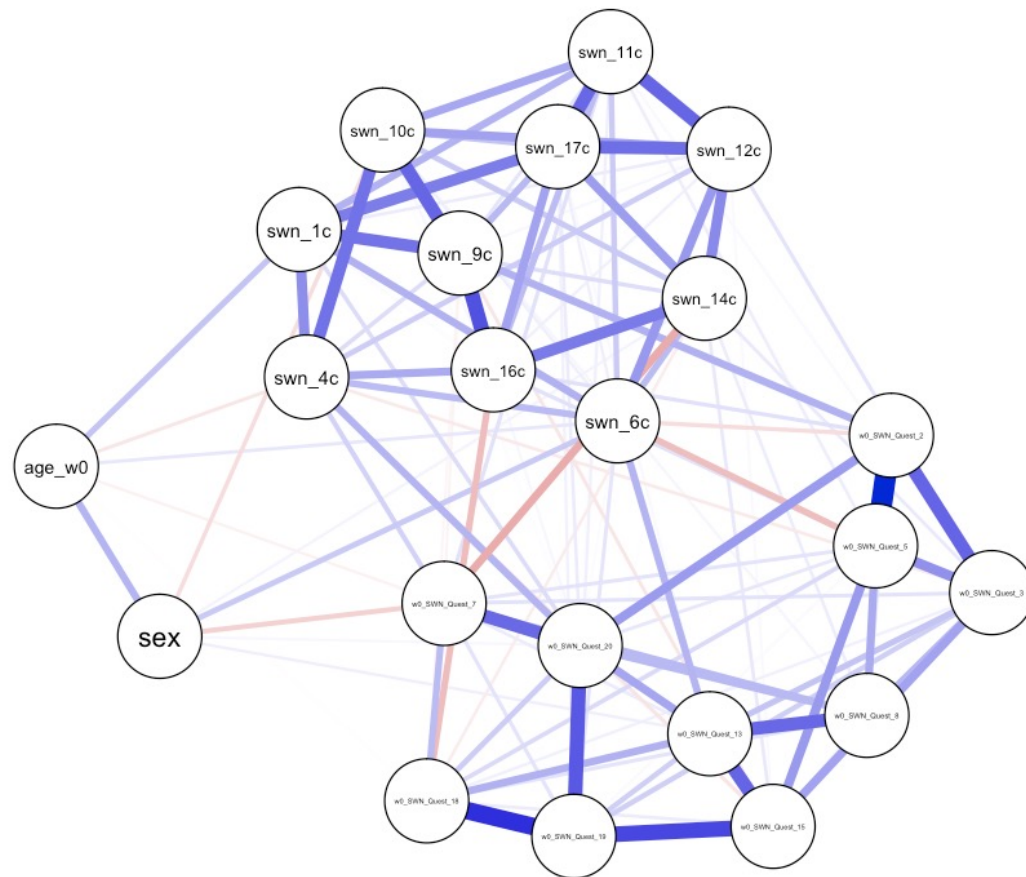
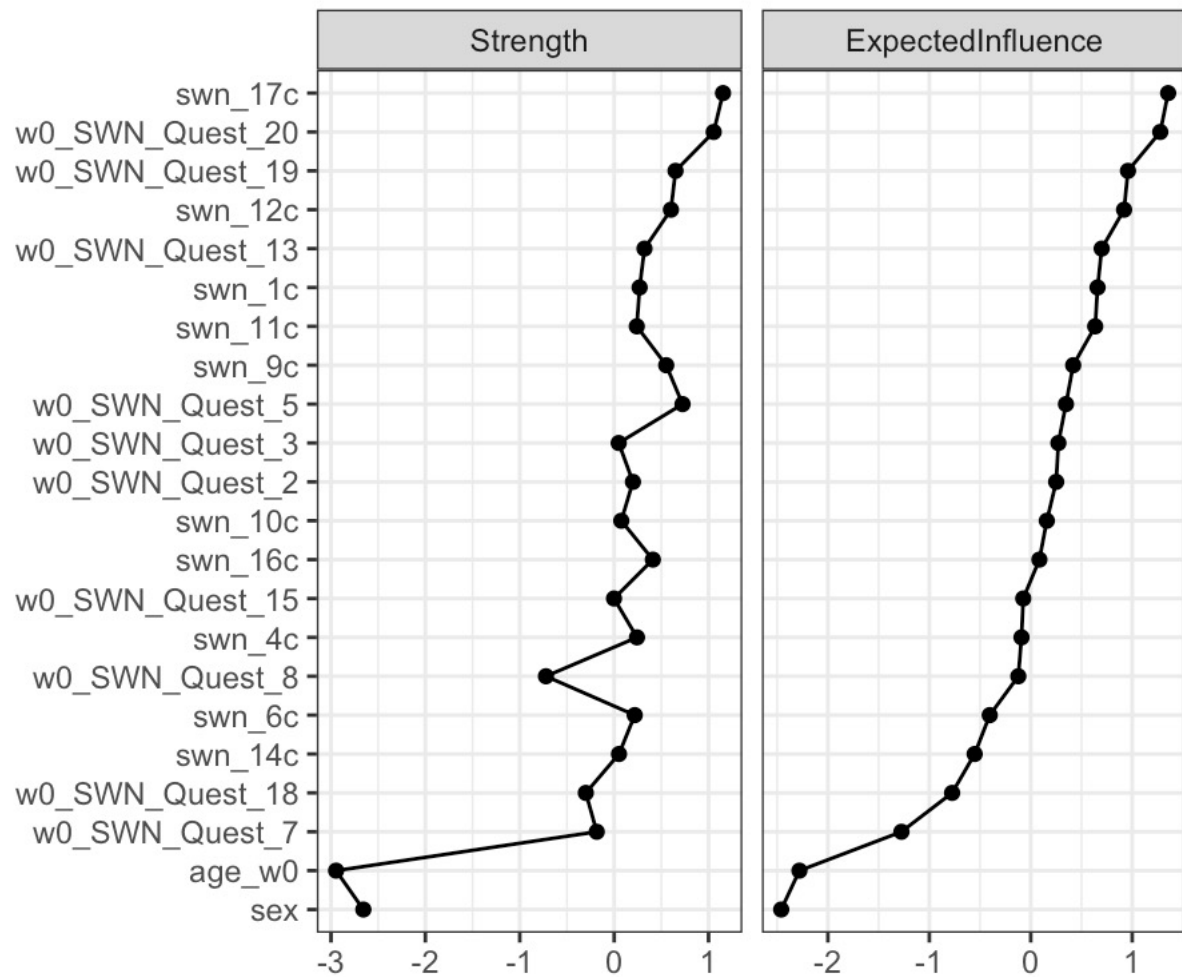


SWN 척도

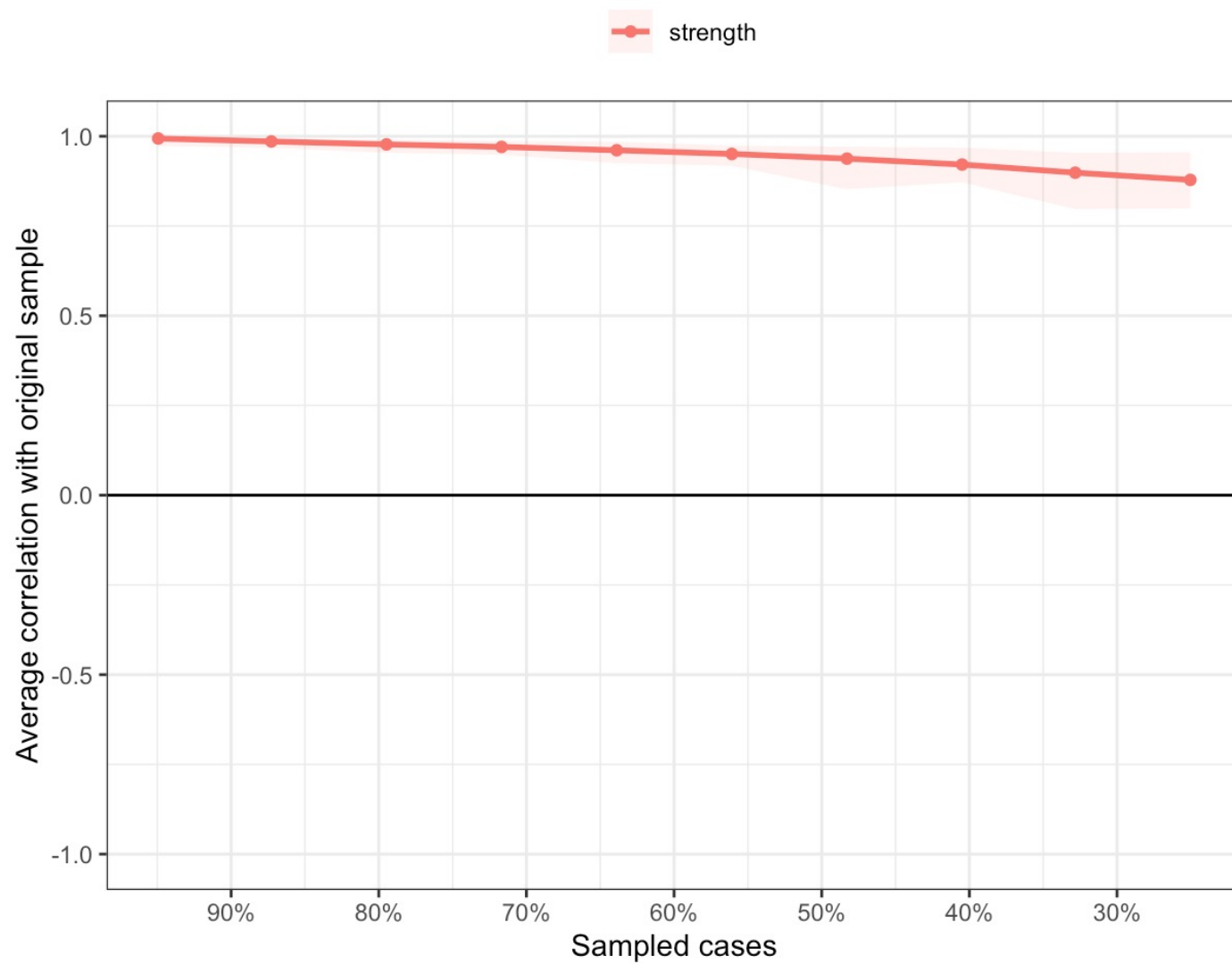


Edge Weight  
신뢰구간 추정

중심성 지수  
안정성 평가



중심성 지수  
안정성 평가



## 질문

1. Study Column?
2. 결측값 처리 방법?
3. 더욱 집중해서 공부해야 될 부분들?  
(수학적 개념, 방법론...)

## todo

1. Node 24개
2. 결측값 제거
3. 네트워크 만드는 과정, 개념에 대해 공부!  
가우시안 - Lasso (집중)  
Networkx 그래프이론, 통계적 접근  
시각화 방법 수정

**THANK YOU!**

**감사합니다!**