Structural Equation Models with Social Network Data

Zhiyong Zhang
Lab for Big Data Methodology
University of Notre Dame

APA Convention
August 8, 2025

Outline

- Structural equation models

Outline

- Structural equation models
- Social network analysis

Outline

- Structural equation models
- Social network analysis
- An example data set

Outline

- Structural equation models
- Social network analysis
- An example data set
- Structural equation modeling with social networks

Outline

- Structural equation models
- Social network analysis
- An example data set
- Structural equation modeling with social networks
- Four examples with code

Outline

- Structural equation models
- Social network analysis
- An example data set
- Structural equation modeling with social networks
- Four examples with code
- Discussion and future directions

Structural Equation Models

- Structural equation models are a collection of models:
  - Regression models
  - Mediation models
  - Factor models
  - MIMIC models

Structural Equation Models

- Structural equation models are a collection of models:
  - Regression models
  - Mediation models
  - Factor models
  - MIMIC models

- It synchronizes different models in the same general framework and allows flexible extension of them.

Structural Equation Models

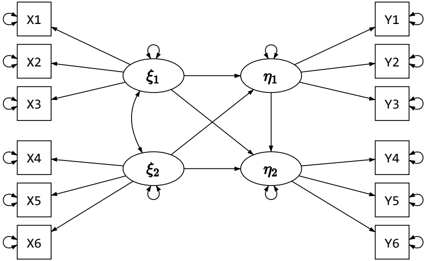
- Structural equation models are a collection of models:
  - Regression models
  - Mediation models
  - Factor models
  - MIMIC models

- It synchronizes different models in the same general framework and allows flexible extension of them.

- It frees researchers from estimating a model to focus on “building a model or theory.”

Path diagram

- A graphical representation of a SEM.
- Squares or rectangles: observed variables, data
- Circles or ovals: latent variables, factors, errors
- One-headed arrows: factor loadings, regression coefficients
- Two-headed arrows: variances, error variances, covariance



Social network analysis

- Social network analysis is a popular interdisciplinary research topic in statistics, sociology, political science, and recently psychology (e.g., Hoff, Raftery, & Handcock, 2002; Saul & Filkov, 2007; Schaefer, Adams, & Haas, 2013; Wasserman & Faust, 1994).

Social network analysis

- Social network analysis is a popular interdisciplinary research topic in statistics, sociology, political science, and recently psychology (e.g., Hoff, Raftery, & Handcock, 2002; Saul & Filkov, 2007; Schaefer, Adams, & Haas, 2013; Wasserman & Faust, 1994).
- Can assess structures or relationships through connections between entities/nodes/subjects in a bounded network.
  - Economics: How are the social, economic, and technological worlds are connected?
  - Politics: How do social networks influence individual's political preference?
  - Epidemiology: Social network analysis was used to analyze the emergence of infectious diseases.
  - Sociology and Psychology: What are the factors that explain the patterns in a social network?
  - Education: Social network analysis is useful in detecting and preventing bullying among students.

Example data

- Data were collected from 180 college students in a 4-year college in Shandong, China.

Example data

- Data were collected from 180 college students in a 4-year college in Shandong, China.
- A sample of 162 participants: 90 female and 72 male students.

Example data

- Data were collected from 180 college students in a 4-year college in Shandong, China.
- A sample of 162 participants: 90 female and 72 male students.
- Basic information
  - Three waves: 2017, 2018, 2019 (1 year after graduation)
  - Average age: 21.64 years (SD=0.86) at wave 1
  - Weight and height
  - Number of WeChat friends
  - Academic performance

Psychological and behavior data

- Big five personality measured by the 20-item Mini-IPIP (Donnellan et al., 2006)
- Depression measured by the Personal Health Questionnaire (7 items, Kroenke et al., 2009)
- Loneliness measured by the UCLA loneliness scale (10 items, Russell et al., 1978)
- Happiness measured by the subjective happiness scale (4 items, Lyubomirsky and Lepper, 1999).

Psychological and behavior data

- Big five personality measured by the 20-item Mini-IPIP (Donnellan et al., 2006)
- Depression measured by the Personal Health Questionnaire (7 items, Kroenke et al., 2009)
- Loneliness measured by the UCLA loneliness scale (10 items, Russell et al., 1978)
- Happiness measured by the subjective happiness scale (4 items, Lyubomirsky and Lepper, 1999).
- Alcohol use
  - Do you drink alcohol?
  - How many times have you drunk alcohol in the past 30 days?
- Smoking
  - Do you smoke?
  - If you smoke, how many cigars on average each day do you smoke in the past 30 days?

Descriptive statistics

Name	Mean	Median	SD	Minimum	Maximum
Gender	Male 74 (45%)			Female 91 (55%)	
Age	21.64	22	0.855	18	24
BMI	21.51	20.31	3.848	15.4	39.52
GPA	3.273	3.285	0.488	1.173	4.22
WeChat friends	165	106	182	23	1000
Extroversion	2.914	3	0.786	1	5
Agreeableness	3.556	3.5	0.613	1.75	5
Conscientiousness	3.532	3.5	0.697	2	5
Neuroticism	2.876	2.75	0.638	1	4.75
Imagination	3.538	3.5	0.687	1.5	5
Depression	0.780	0.714	0.418	0	1.857
Loneliness	1.128	1.1	0.567	0	2.6
Happiness	4.935	4.75	0.868	2.5	7
Smoking	Yes 43 (36%)			No 122 (64%)	
Alcohol use	Yes 68 (41%)			No 97 (59%)	

Friendship network

Each student was given a roster of all the students in the study and was asked to report his/her acquaintanceship with every others. The friendship is measured on a 5-point Likert scale

Friendship network

Each student was given a roster of all the students in the study and was asked to report his/her acquaintanceship with every others. The friendship is measured on a 5-point Likert scale

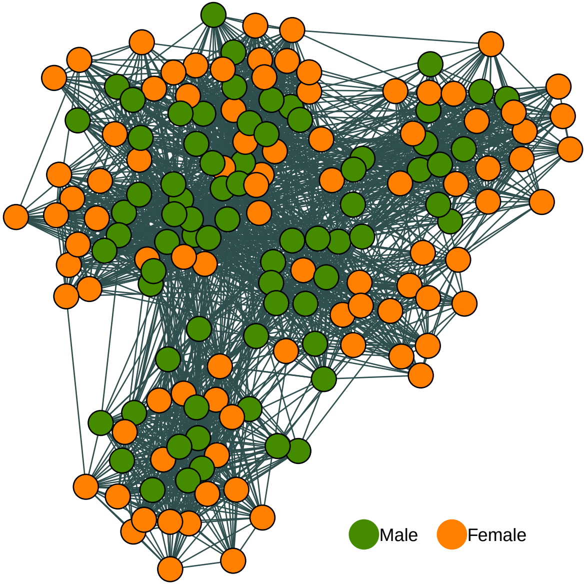
1. I have never heard about the student.
2. I heard about the student but had no personal interaction with her/him.
3. I have met the student a few times but he/she is not a friend of mine.
4. The student is a friend of mine.
5. The student is one of my best friends.

Friendship network

Each student was given a roster of all the students in the study and was asked to report his/her acquaintanceship with every others. The friendship is measured on a 5-point Likert scale

1. I have never heard about the student.
2. I heard about the student but had no personal interaction with her/him.
3. I have met the student a few times but he/she is not a friend of mine.
4. The student is a friend of mine.
5. The student is one of my best friends.

Network plot - friends / best friends



Data structure

id	p1	p2	p3	p4	p5	p6	gender	age	smoke	alcohol	extraversion
p1	NA	0	1	1	0	1	M	20	Y	Y	1.9
p2	0	NA	1	0	1	0	F	21	N	Y	0.8
p3	1	1	NA	0	1	1	F	20	N	N	0.7
p4	1	0	0	NA	0	1	M	19	N	Y	0.5
p5	0	1	1	0	NA	1	M	20	N	Y	2.1
p6	1	0	1	1	1	NA	M	21	Y	N	2.3

Data structure

id	p1	p2	p3	p4	p5	p6	gender	age	smoke	alcohol	extraversion
p1	NA	0	1	1	0	1	M	20	Y	Y	1.9
p2	0	NA	1	0	1	0	F	21	N	Y	0.8
p3	1	1	NA	0	1	1	F	20	N	N	0.7
p4	1	0	0	NA	0	1	M	19	N	Y	0.5
p5	0	1	1	0	NA	1	M	20	N	Y	2.1
p6	1	0	1	1	1	NA	M	21	Y	N	2.3

Network Data

Data structure

id	p1	p2	p3	p4	p5	p6	gender	age	smoke	alcohol	extraversion
p1	NA	0	1	1	0	1	M	20	Y	Y	1.9
p2	0	NA	1	0	1	0	F	21	N	Y	0.8
p3	1	1	NA	0	1	1	F	20	N	N	0.7
p4	1	0	0	NA	0	1	M	19	N	Y	0.5
p5	0	1	1	0	NA	1	M	20	N	Y	2.1
p6	1	0	1	1	1	NA	M	21	Y	N	2.3

Predictors

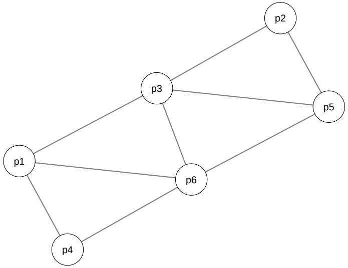
Data structure

id	p1	p2	p3	p4	p5	p6	gender	age	smoke	alcohol	extraversion
p1	NA	0	1	1	0	1	M	20	Y	Y	1.9
p2	0	NA	1	0	1	0	F	21	N	Y	0.8
p3	1	1	NA	0	1	1	F	20	N	N	0.7
p4	1	0	0	NA	0	1	M	19	N	Y	0.5
p5	0	1	1	0	NA	1	M	20	N	Y	2.1
p6	1	0	1	1	1	NA	M	21	Y	N	2.3

Predictors or outcomes

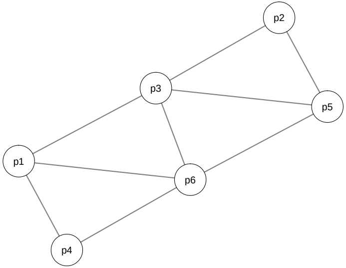
Modeling networks

- Traditional network analysis often focuses on modeling the network itself.



Modeling networks

- Traditional network analysis often focuses on modeling the network itself.



- Exponential random graph model (ERGM; Anderson, Wasserman, & Crouch, 1999; Frank & Strauss, 1986) treats the entire social network as a random variable and explains the probability of networks using their local features such as triangle counts and node degrees.

Latent space models (Hoff, 2002)

$$\begin{cases} y_{ij} & \sim \text{Bernoulli}(p_{ij}) \\ \text{logit}(p_{ij}) & = \alpha - ||\mathbf{z}_i - \mathbf{z}_j|| \end{cases}$$

where

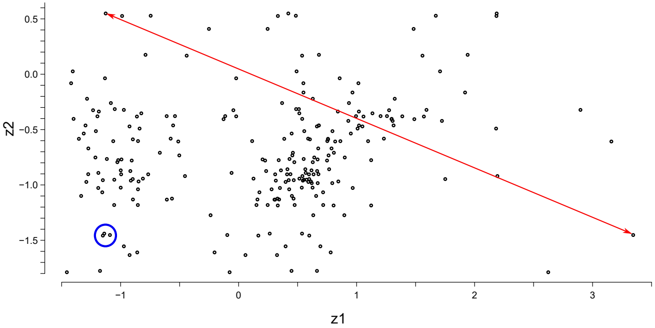
- α is a constant,
- $\mathbf{z}_i$ is the latent position vector of node i, and
- $||\mathbf{z}_i - \mathbf{z}_j||$ is the distance of node i and node j.

Latent space models (Hoff, 2002)

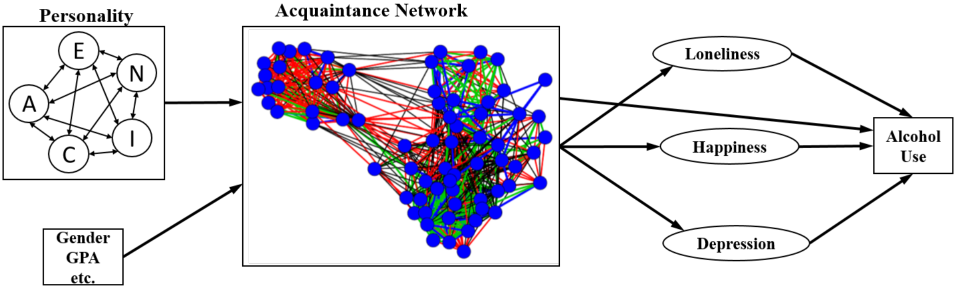
$$\begin{cases} y_{ij} & \sim \text{Bernoulli}(p_{ij}) \\ \text{logit}(p_{ij}) & = \alpha - ||\mathbf{z}_i - \mathbf{z}_j|| \end{cases}$$

where

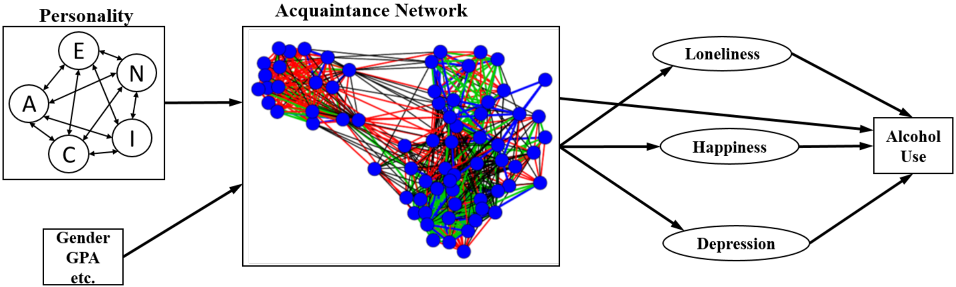
- α is a constant,
- $\mathbf{z}_i$ is the latent position vector of node i, and
- $||\mathbf{z}_i - \mathbf{z}_j||$ is the distance of node i and node j.



A general SEM framework with networks

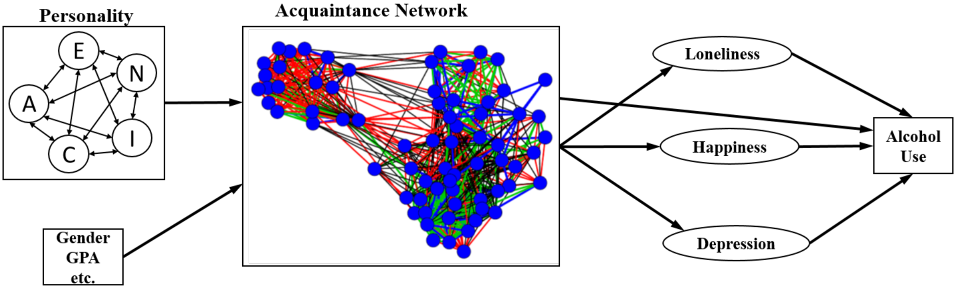


A general SEM framework with networks



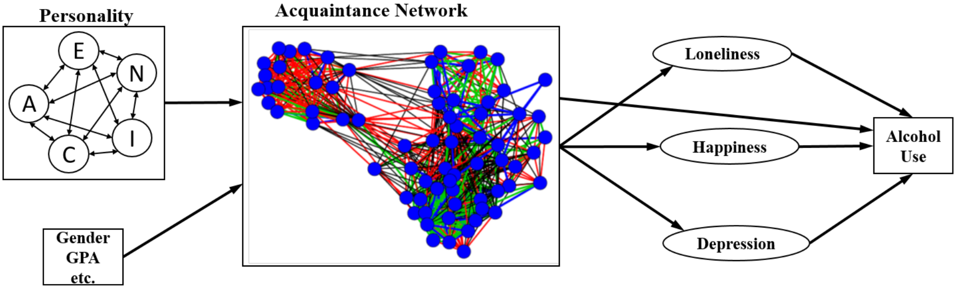
- Network as outcomes: Friendship network can be predicted by gender, age, personality and other variables.

A general SEM framework with networks



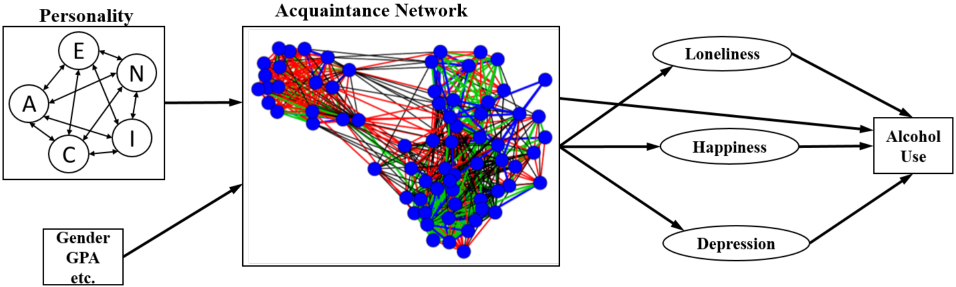
- Network as outcomes: Friendship network can be predicted by gender, age, personality and other variables.
- Network as predictors: Friendship network can predict depression, loneliness, alcohol use, and other outcomes.

A general SEM framework with networks



- Network as outcomes: Friendship network can be predicted by gender, age, personality and other variables.
- Network as predictors: Friendship network can predict depression, loneliness, alcohol use, and other outcomes.
- Network as mediators: Friendship network can server as a mediator or intermediate variable between two variables.

A general SEM framework with networks



- Network as outcomes: Friendship network can be predicted by gender, age, personality and other variables.
- Network as predictors: Friendship network can predict depression, loneliness, alcohol use, and other outcomes.
- Network as mediators: Friendship network can server as a mediator or intermediate variable between two variables.
- Easy to extend: Longitudinal and dynamic models.

Model estimation

- A two-stage method is used to estimate the model.
- The key is to match the dimensions of the network data and the non-network data.
- Both node-based method and edge-based method can be used.

Model estimation

- A two-stage method is used to estimate the model.
- The key is to match the dimensions of the network data and the non-network data.
- Both node-based method and edge-based method can be used.
- We have developed both an R package networksem and an online app to facilitate the model estimation.

Node-based method

id	p1	p2	p3	p4	p5	p6
p1	NA	0	1	1	0	1
p2	0	NA	1	0	1	0
p3	1	1	NA	0	1	1
p4	1	0	0	NA	0	1
p5	0	1	1	0	NA	1
p6	1	0	1	1	1	NA

degree	closeness	gender	age	smoke	alcohol	extraversion
2	0.143	M	20	Y	Y	1.9
2	0.111	F	21	N	Y	0.8
4	0.167	F	20	N	N	0.7
2	0.111	M	19	N	Y	0.5
3	0.143	M	20	N	Y	2.1
4	0.167	M	21	Y	N	2.3

Edge-based method

id	p1	p2	p3	p4	p5	p6	gender	age	smoke	alcohol	extraversion
p1	NA	0	1	1	0	1	M	20	Y	Y	1.9
p2	0	NA	1	0	1	0	F	21	N	Y	0.8
p3	1	1	NA	0	1	1	F	20	N	N	0.7
p4	1	0	0	NA	0	1	M	19	N	Y	0.5
p5	0	1	1	0	NA	1	M	20	N	Y	2.1
p6	1	0	1	1	1	NA	M	21	Y	N	2.3

↓

Node1	Node 2	Edge	age.diff	ext.avg
p1	p2	0	1	1.35
p1	p3	1	0	1.3
p1	p4	1	-1	1.2
p1	p5	0	0	2
p1	p6	1	1	2.1
p2	p3	1	1	1.5
p2	p4	0	2	1.3
…	…	…	…	…
p5	p6	1	-1	2.2

Example: Node-based method through network statistics

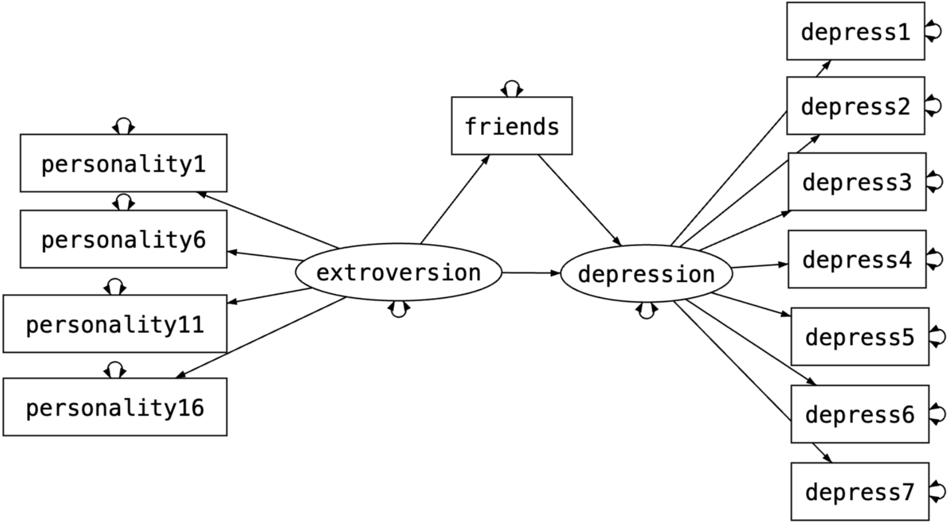
- Degree: number of connections to other nodes, a measure of popularity
- Closeness: how close a node is to all other nodes, socially central or well-integrated
- Betweenness: how often a node lies on the shortest paths between other nodes, social bridge, influence between groups

id	p1	p2	p3	p4	p5	p6
p1	NA	0	1	1	0	1
p2	0	NA	1	0	1	0
p3	1	1	NA	0	1	1
p4	1	0	0	NA	0	1
p5	0	1	1	0	NA	1
p6	1	0	1	1	1	NA

degree	closeness	gender	age	smoke	alcohol	extraversion
2	0.143	M	20	Y	Y	1.9
2	0.111	F	21	N	Y	0.8
4	0.167	F	20	N	N	0.7
2	0.111	M	19	N	Y	0.5
3	0.143	M	20	N	Y	2.1
4	0.167	M	21	Y	N	2.3

Example: path diagram

- Use the degree from the friendship network as a mediator.



Data

- Mini-IPIP (International Personality Item Pool; Donnellan et al., 2006)
- Extroversion factor:
  - 1: Don't talk a lot.
  - 6: Keep in the background.
  - 11: Am the life of the party.
  - 16: Talk to a lot of different people at parties.
- Depression: Personal Health Questionnaire (PHQ; Kroenke et al., 2009). Only 7 items (removed item 6 & 9).
- Friendship: self-reported, either friend or not.

Data organization

- To use our R package networksem for analysis, data need to be organized as a list.

```
> str(friend_data)
List of 2
 $ network    :List of 1
  ..$ friends: num [1:165, 1:165] 0 1 1 1 1 1 1 1 1 1 ...
 $ nonnetwork:'data.frame':     165 obs. of  11 variables:
  ..$ personality1 : int [1:165] 3 3 4 3 1 3 2 3 4 3 ...
  ..$ personality6 : int [1:165] 3 1 4 3 1 4 3 5 4 5 ...
  ..$ personality11: int [1:165] 3 4 4 4 2 1 5 3 2 1 ...
  ..$ personality16: int [1:165] 3 3 3 4 4 3 3 1 2 1 ...
  ..$ depress1     : int [1:165] 1 0 1 1 1 1 0 2 2 1 ...
  ..$ depress2     : int [1:165] 0 0 1 1 2 2 0 0 0 1 ...
  ..$ depress3     : int [1:165] 0 1 0 1 1 1 0 0 1 0 ...
  ..$ depress4     : int [1:165] 0 0 1 1 1 2 0 0 1 1 ...
  ..$ depress5     : int [1:165] 0 0 1 1 2 1 0 0 1 0 ...
  ..$ depress6     : int [1:165] 0 0 0 0 3 1 0 0 0 1 ...
  ..$ depress7     : int [1:165] 0 0 0 0 0 2 0 3 1 1
```

Model specification

- The model specification used by the R package lavaan can be utilized here.
- In the model, a network variable can be directly used. The name should match the one used in the network list of the data.

```
ex1.model <- '
 extroversion =~ personality1 + personality6
                + personality11 +personality16
 depression =~ depress1 + depress2 + depress3
               + depress4 + depress5 + depress6 + depress7
 friends ~ extroversion
 depression ~ friends + extroversion
 '
```

Model estimation

- To estimate the model, the function sem.net from the networksem package can be used.

```
fit <- sem.net(model = ex1.model, data = friend_data, std.
   lv = T, netstats = 'degree', netstats.rescale = TRUE)
```

- sem.net conducts the node-based analysis.
- The required inputs include the “model” and the “data”.
- Different network statistics can be used, here, “degree”. Multiple network statistics can be used at the same time.
- std.lv specifies whether to standardize the latent variables.

Model results I

```
> summary(fit)
The SEM output:
lavaan 0.6-19 ended normally after 43 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        26

  Number of observations                           165

Model Test User Model:

  Test statistic                                64.549
  Degrees of freedom                                52
  P-value (Chi-square)                           0.114
```

Model results II

```
Model Test Baseline Model:

  Test statistic                               343.181
  Degrees of freedom                                66
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.955
  Tucker-Lewis Index (TLI)                       0.943

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -2758.904
```

Model results III

```
  Loglikelihood unrestricted model (H1)      -2726.629

  Akaike (AIC)                                5569.807
  Bayesian (BIC)                              5650.562
  Sample-size adjusted Bayesian (SABIC)       5568.246

Root Mean Square Error of Approximation:

  RMSEA                                          0.038
  90 Percent confidence interval - lower         0.000
  90 Percent confidence interval - upper         0.066
  P-value H_0: RMSEA <= 0.050                    0.730
  P-value H_0: RMSEA >= 0.080                    0.004

Standardized Root Mean Square Residual:
```

Model results IV

```
  SRMR                                           0.062

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  depression =~
    depress7          0.319    0.059    5.362    0.000
    depress6          0.375    0.057    6.562    0.000
    depress5          0.407    0.054    7.509    0.000
```

Model results V

```
    depress4            0.418    0.047    8.911    0.000
    depress3            0.442    0.063    6.970    0.000
    depress2            0.336    0.048    7.031    0.000
    depress1            0.226    0.047    4.849    0.000
  extroversion =~
    personality16       0.835    0.117    7.165    0.000
    personality11       0.637    0.112    5.698    0.000
    personality6       -0.541    0.104   -5.183    0.000
    personality1       -0.458    0.099   -4.629    0.000

Regressions:
                     Estimate  Std.Err  z-value  P(>|z|)
  depression ~
    extroversion        0.016    0.119    0.131    0.896
  friends.degree ~
```

Model results VI

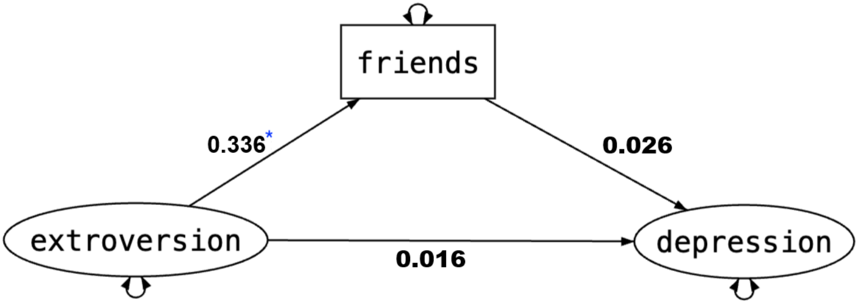
```
    extroversion        0.336    0.093    3.598    0.000
  depression ~
    friends.degree      0.026    0.098    0.263    0.793

Variances:
                     Estimate  Std.Err  z-value  P(>|z|)
   .depress7            0.399    0.048    8.390    0.000
   .depress6            0.343    0.043    7.961    0.000
   .depress5            0.286    0.038    7.488    0.000
   .depress4            0.182    0.028    6.438    0.000
   .depress3            0.409    0.053    7.775    0.000
   .depress2            0.232    0.030    7.745    0.000
   .depress1            0.252    0.029    8.531    0.000
   .personality16       0.741    0.162    4.586    0.000
   .personality11       1.032    0.145    7.111    0.000
```

Model results VII

```
.personality6      0.968    0.127    7.606    0.000
.personality1      0.921    0.115    7.990    0.000
.friends.degree    0.881    0.104    8.482    0.000
.depression        1.000
 extroversion      1.000
```

Conclusions

- The model fits the data well ($\chi^2 = 64.549, df = 52, p = 0.114,$ $CFI = 0.955, TLI = 0.943, RMSEA = 0.038, SRMR = 0.062$).



- Extrovert personality is associated with popularity (degree statistic).
- Neither extroversion nor popularity is related to depression.

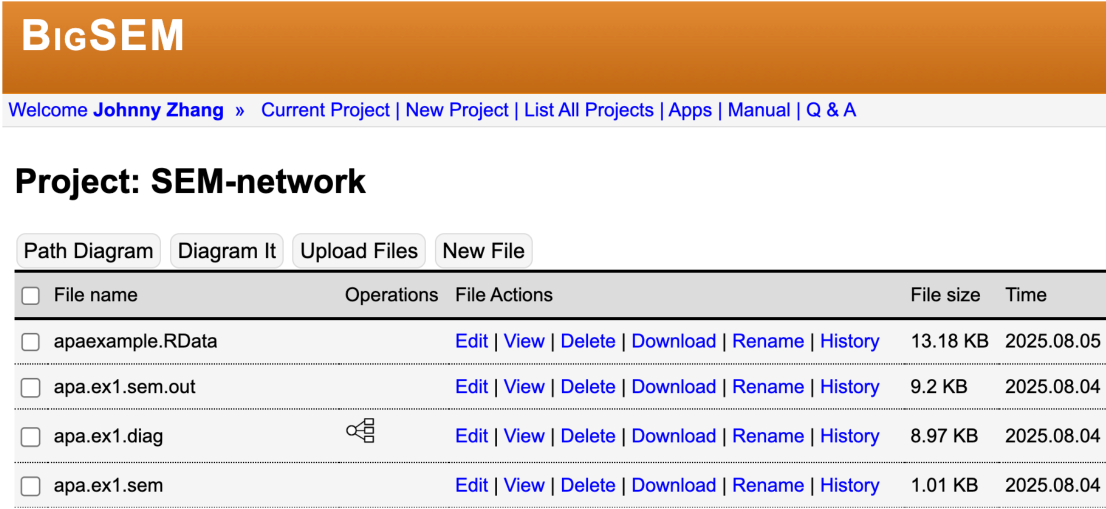
Mediation/indirect effect calculation and testing

```
> path.networksem(fit, 'extroversion', 'friends.degree',
    'depression')

predictor   "extroversion"
mediator    "friends.degree"
outcome     "depression"
apath       "0.3362325"
bpath       "0.02566646"
indirect    "0.008629899"
indirect_se "0.03150774"
indirect_z  "0.2738978"
```

Use of online app

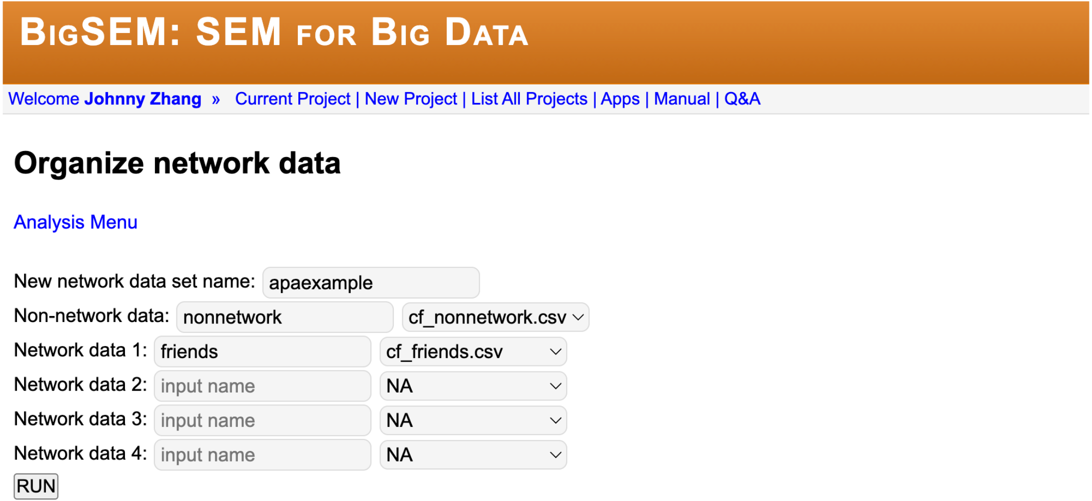
- The same analysis can be conducted using an online app we developed.
- `https://bigsem.psychstat.org/app/`
- It allows the analysis through drawing a path diagram.
  - Organize data
  - Draw a path diagram
  - Conduct the analysis

Interface after login

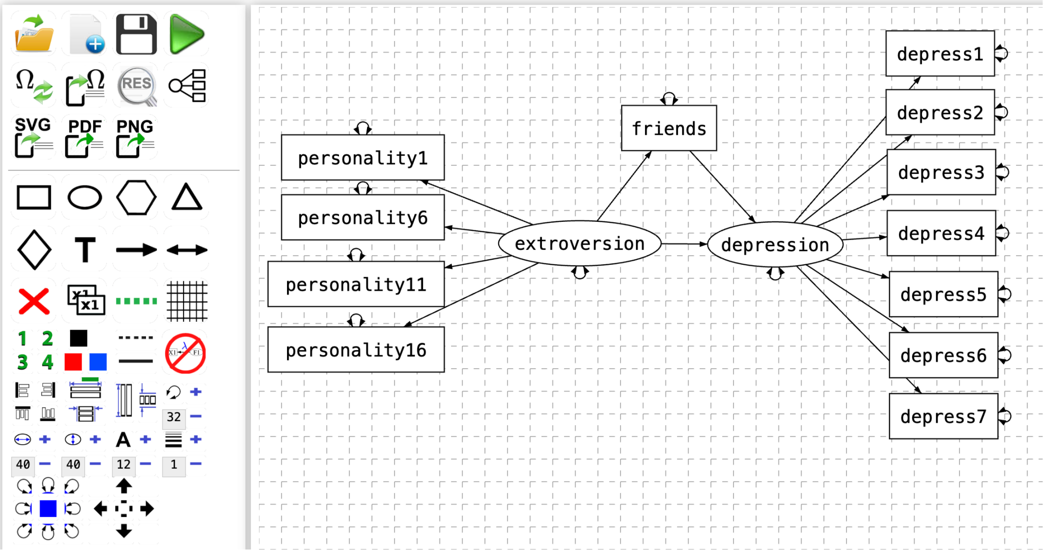


Organize data

- Both non-network and network data can be uploaded as separated .csv files.
- They can then be combined to a list and saved for use with our online tool.



Draw a path diagram



Control the analysis

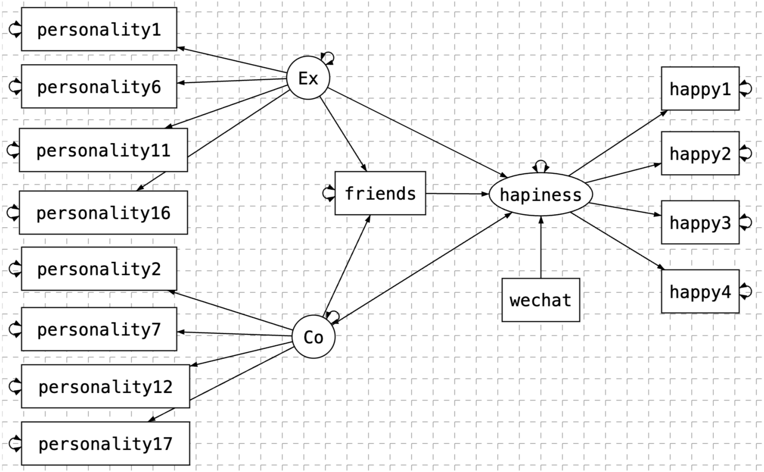
Output

- Slight difference

```
  extroversion =~
    personality16      0.694    0.097    7.165    0.000
    personality11      0.530    0.093    5.698    0.000
    personality6      -0.480    0.093   -5.183    0.000
    personality1      -0.430    0.093   -4.629    0.000
    friends.degree     0.336    0.093    3.598    0.000

Regressions:
                    Estimate  Std.Err  z-value  P(>|z|)
  depression ~
    extroversion       0.016    0.119    0.131    0.896
    friends.degree     0.026    0.098    0.263    0.793
```

Multiple networks (example 2)

- Self-reported friendship and social media
- Degree, closeness, and betweenness



R code I

```
load("network.RData")

## specify the model
ex2.model <-'
  extroversion =~ personality1 + personality6
            + personality11 +personality16
  conscientiousness =~ personality2 + personality7
                  + personality12 +personality17
  happiness =~ happy1 + happy2 + happy3 + happy4
  friends ~ extroversion + conscientiousness
  happiness ~ friends + wechat + extroversion +
      conscientiousness
'
```

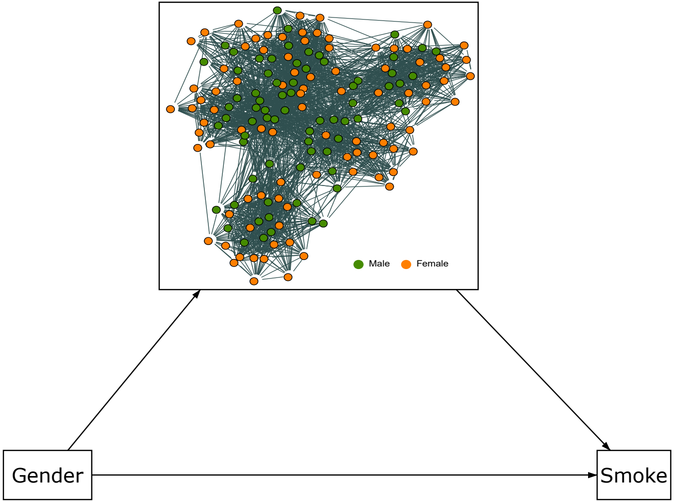
R code II

```
## fit the model
fit2 <- sem.net(ex2.model, data = network, std.lv=T,
  netstats = c('degree', 'closeness', 'betweenness'),
  netstats.rescale = TRUE)
summary(fit2)

Regressions:
                        Estimate  Std.Err  z-value  P(>|z|)
  happiness ~
    conscientisnss         2.132   12.241    0.174    0.862
    extroversion          -2.368   12.622   -0.188    0.851
  friends.degree ~
    extroversion           1.756    0.484    3.627    0.000
  friends.closeness ~
    extroversion           1.630    0.462    3.528    0.000
```

R code III

```
friends.betweenness ~
  extroversion             1.557    0.446    3.495    0.000
friends.degree ~
  conscientisnss          -1.682    0.433   -3.887    0.000
friends.closeness ~
  conscientisnss          -1.604    0.413   -3.882    0.000
friends.betweenness ~
  conscientisnss          -1.541    0.399   -3.864    0.000
happiness ~
  friends.degree           0.505    4.220    0.120    0.905
  friends.clsnss           0.695    1.778    0.391    0.696
  frinds.btwnnss           0.234    1.161    0.201    0.840
  wechat.degree            0.098    0.240    0.411    0.681
  wechat.closnss          -0.433    0.240   -1.806    0.071
  wechat.btwnnss           0.350    0.210    1.669    0.095
```

Example 3: mediation between gender and smoking behavior



Node-based analysis through latent space models

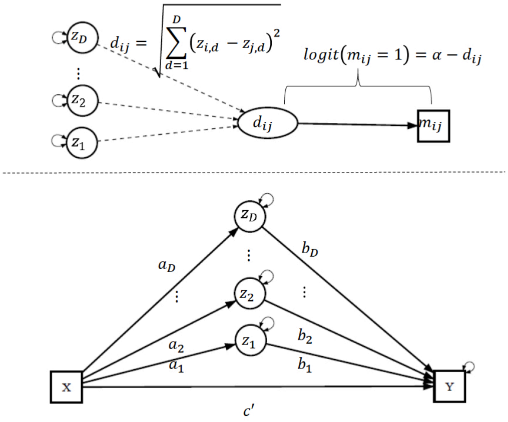
$$\begin{cases} y_{ij} & \sim \text{Bernoulli}(p_{ij}) \\ \text{logit}(p_{ij}) & = \alpha - ||\mathbf{z}_i - \mathbf{z}_j|| \end{cases}$$

id	p1	p2	p3	p4	p5	p6
p1	NA	0	1	1	0	1
p2	0	NA	1	0	1	0
p3	1	1	NA	0	1	1
p4	1	0	0	NA	0	1
p5	0	1	1	0	NA	1
p6	1	0	1	1	1	NA

id	z1	z2	gender	age	smoke	alcohol	extraversion
p1	-5.65	-1.15	M	20	Y	Y	1.9
p2	-6.49	-2.27	F	21	N	Y	0.8
p3	1.18	2.66	F	20	N	N	0.7
p4	-4.37	0.05	M	19	N	Y	0.5
p5	1.33	3.52	M	20	N	Y	2.1
p6	-10.45	-1.71	M	21	Y	N	2.3

Mediation effect

- Dimension reduction method
- Overall mediation effect ($\sum a_d b_d$), individual dimension unexplainable



Model estimation I

- Five latent dimensions seemed to work best.
- The function sem.net.lsm can be used.
- The outcome "smoke" is a binary variable.

```
ex3.model <- '
  friends ~ gender
  smoke ~ friends + gender
'

fit3 <- sem.net.lsm(ex3.model, data = network, latent.dim=5,
     ordered='smoke')
summary(fit3)
```

Model estimation II

- The total mediation effect can be calculated.

```
## Get the parameter information
parest <- fit3$estimates$sem.es@ParTable
## indirect effect
indirect_effects <- sum(parest$est[2:6]*parest$est[7:11])
## se of the indirect effect
indirect_se <- sqrt(sum(parest$se[2:6]^2 * parest$est[7:11]^2
      + parest$se[7:11]^2 * parest$est[2:6]^2))
## Z-score
z_score <- indirect_effects / indirect_se
## p-values
p_value <- 2 * (1 - pnorm(abs(z_score)))
```

Model estimation III

```
> indirect_effects
[1] -0.4185963
> z_score
[1] -2.285045
> p_value
[1] 0.02231016
```

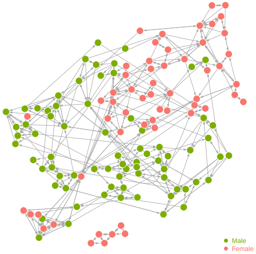
Results and conclusions

- Gender –> Friendship network –> smoking
- The estimated mediation effect was -0.42, which was significant based on the Sobel test.
- The estimated direct effect was –1.21, also significant.
- Therefore, the friendship network partially mediates the relationship between gender and smoking.

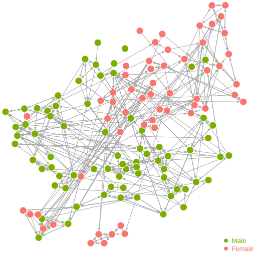
Example 4: Longitudinal network analysis

- Network data can be collected across time.
- Teenage Friends and Lifestyle Study (Michell and West, 1996; Pearson and Michell, 2000)
  - A total of 129 pupils with 73 boys and 56 girls.
  - Networks
    - Friendship network formed by asking each student to name up to six friends.
    - Only 13 out of the 129 students named the maximum number of six friends, with the median number of named friends to be 3 and the average number of friends to be 3.5.
    - 2 standing for "best friend", 1 for "just a friend", and code 0 for "no friend". The average number of “best friend” was 0.67 and the average number of “just a friend” was 2.81 for the first wave of data.
    - A binary network was created with 1 being best friend or just a friend.

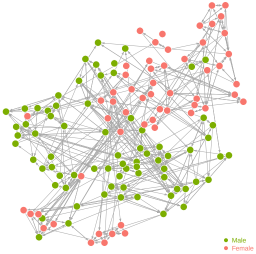
Network plot



(a) Friendship at Time 1.

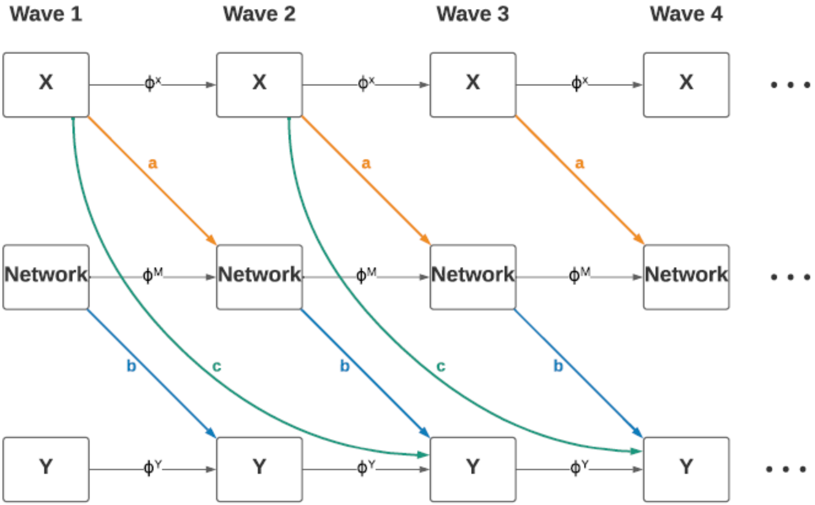


(b) Friendship at Time 2.



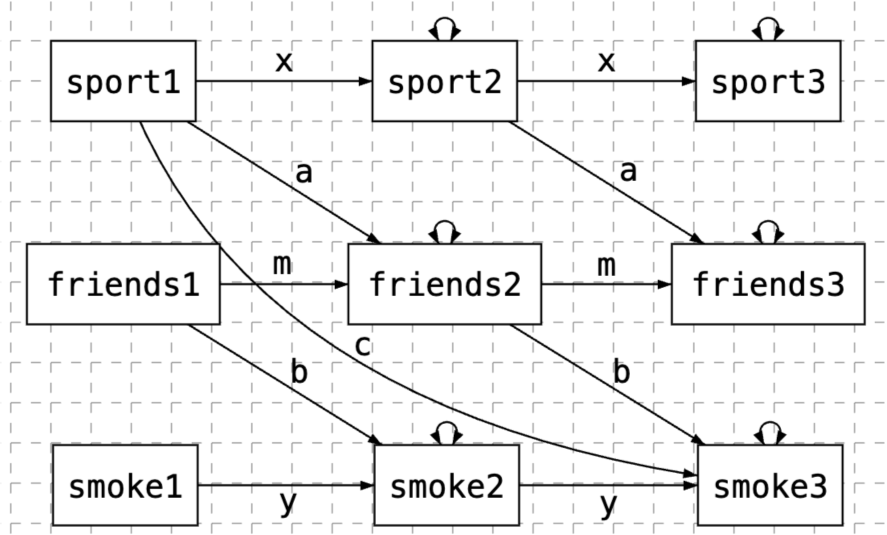
(c) Friendship at Time 3.

A longitudinal mediation model



Teenage Friends and Lifestyle Study

- Sport activity –> friendship –> smoking



Edge-based method

id	p1	p2	p3	p4	p5	p6	gender	age	smoke	alcohol	extraversion
p1	NA	0	1	1	0	1	M	20	Y	Y	1.9
p2	0	NA	1	0	1	0	F	21	N	Y	0.8
p3	1	1	NA	0	1	1	F	20	N	N	0.7
p4	1	0	0	NA	0	1	M	19	N	Y	0.5
p5	0	1	1	0	NA	1	M	20	N	Y	2.1
p6	1	0	1	1	1	NA	M	21	Y	N	2.3

$\downarrow$

Node1	Node 2	Edge	age.diff	ext.avg
p1	p2	0	1	1.35
p1	p3	1	0	1.3
p1	p4	1	-1	1.2
p1	p5	0	0	2
p1	p6	1	1	2.1
p2	p3	1	1	1.5
p2	p4	0	2	1.3
...	...	...	...	...
p5	p6	1	-1	2.2

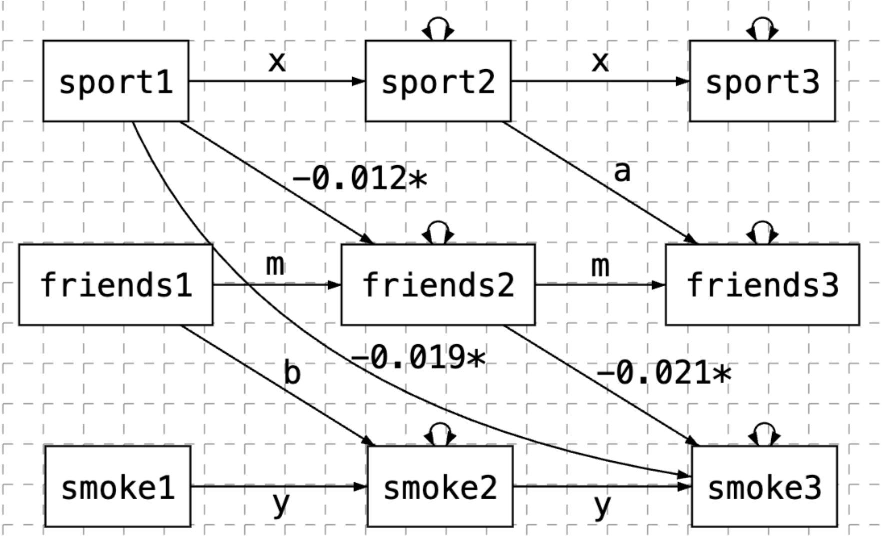
Model estimation

- The function sem.net.edge is used.

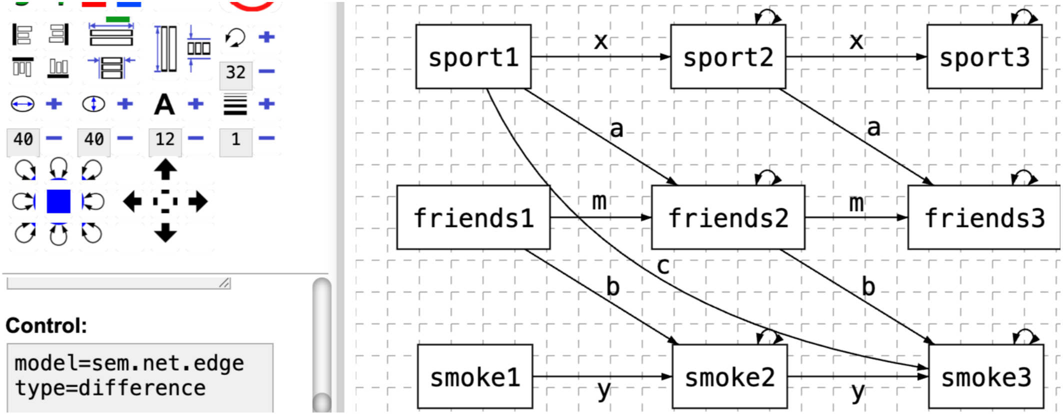
```
load("ex4.data.RData") ## Glasgow data
ex4.model <- '
  friends2 ~ a*sport1 + m*friends1
  friends3 ~ a*sport2 + m*friends2
  sport2 ~ x*sport1
  sport3 ~ x*sport2
  smoke2 ~ y*smoke1 + b*friends1
  smoke3 ~ y*smoke2 + b*friends2 + c*sport1
  ab := a*b
'

fit4 <- sem.net.edge(ex4.model, data = ex4.data,
                     type='difference',
                     netstats.rescale = T)
summary(fit4)
```

Results

- A significant yet small effect.

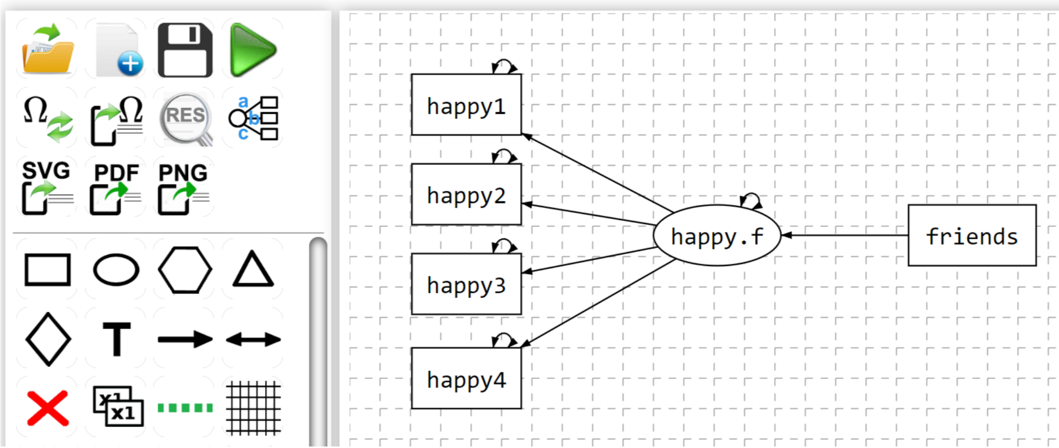


Use of the online app



Software

- General purpose software program for SEM with networks
  - R package neworksem (available on CRAN with the latest development on Github)
  - Online app: `https://bigsem.psychstat.org/app`
  - Manual: `https://bigsem.org` (Xu & Zhang, 2025; SEM)



Discussion

- Simultaneously analyzing network and non-network data can
  - Inform the formation of networks.
  - Understand the effects of networks on behaviors
- We utilized a two-stage method for each model estimation and interpretation.
- Statistical properties still need further investigation, particularly for the edge based method.
- Future directions
  - Brain network
  - Psychometric network
  - Improve software

Other methods

- Traditional network analysis often focuses on modeling the network itself.
  - Zhang et al. (2018): longitudinal clustering
- We have developed methods and models to study the association between networks and non-network variables.
  - Liu, Jin & Zhang (2018): joint modeling network latent space and factor space
  - Che, Jin & Zhang (2021), Liu, Jin & Zhang (2021): Model network as a mediator
  - Xu & Zhang (2025): SEM with networks
  - Xu & Zhang (under review): Dynamic network with covariates
- Other considerations
  - Qu, Liu & Zhang (2020): Permutation test
  - Xu, Hai, Yang & Zhang (2023): Missing data

Acknowledgment

- Current and former students
  - Chang Che, Meta
  - Haiyan Liu, UC Merced
  - Wen Qu, Fudan University
  - Ziqian Xu, University of Notre Dame
- Colleagues
  - Ick-Hoon Jin, Yonsei University
  - Lijuan Wang, University of Notre Dame
  - Ke-Hai Yuan, University of Notre Dame
- Funding
  - Institute of Education Sciences (R305D210023)
  - Notre Dame Global
  - Franco Family Institute for Liberal Arts and the Public Good

We value your feedback

- We need your feedback to improve our software programs.
- If you can fill out our survey here: https://forms.gle/ecExNjimzPonQedE7, you can get a $25 Amazon gift card.

Q & A

- For more information
  - Zhiyong Zhang (zzhang4@nd.edu)
  - Website: `https://bigdatalab.nd.edu`

Lab For BIG DATA Methodology