ARTICLE

Structural Equation Models with Social Networks

Ziqian Xu[a] and Zhiyong Zhang[a]

[a]Department of Psychology, University of Notre Dame

ARTICLE HISTORY
Compiled March 30, 2025

Abstract
Structural Equation Modeling (SEM) is a useful tool to investigate relationships between observed and latent variables in social sciences, yet so far its application is primarily on individual-level data such as psychometric traits and demographics data collected through surveys. However, the rapid development of technologies recently has brought attention to more complex forms of data in social sciences such as network data where relationships between pairs of individuals are of interest. Therefore, it becomes important to form a unified framework that can allow researchers to explore relationships between individual-level data and network data. This paper attempts to address this gap by proposing methods of analyzing network data together with individual-level attributes in the SEM framework. An R package and a corresponding web application are also introduced to simplify the applications.

KEYWORDS
Structural equation modeling; network data; social network analysis

1. Introduction

To understand human thoughts and behaviors, social sciences often rely on surveys designed specifically to collect information from human participants (Groves et al., 2011). Information collected in such fashion tends to suffer from measurement error problems (Biemer, Groves, Lyberg, Mathiowetz, & Sudman, 2013; Bound, Brown, & Mathiowetz, 2001), and often, the construct of interest cannot be adequately captured by a single survey question (Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012). Because of this, it is quite common to see multiple items (i.e., observed variables) being used to describe the same underlying construct (i.e., latent variable), making the Structural Equation Modeling (SEM) approach an ideal framework of analysis (Bollen, 1989; Kline, 2023). Under the SEM approach, models not only consider the measurement relationships between observed and latent variables, but also take into account the structural or regression relationships among the latent variables themselves.

Traditionally, SEM analysis uses survey data which are individual-level responses on a number of target variables (Bollen, 1989; Hair Jr et al., 2021; Kline, 2023). In recent years, social science research has increasingly collected data with more complex structures than the individual-level survey data. For example, text data has become a popular medium in social sciences for identifying major topics in discussions and

CONTACT Ziqian Xu. Email: zxu9@nd.edu

people's sentiments (Baden, Pipal, Schoonvelde, & van der Velden, 2022; Grimmer, Roberts, & Stewart, 2022); biometric data such as facial image and speech pattern can be used to understand traits like emotions and attitudes (Fairhurst, Li, & Da Costa-Abreu, 2017; Kim & Kim, 2022); and network data can provide insights in understanding relationships among entities of interest (Che, Jin, & Zhang, 2020; Clifton & Webster, 2017; Liu, Jin, Zhang, & Yuan, 2021; Sweet, 2016). Within the various types of network data, social network data is particularly relevant to social sciences. For instance, studying student behavior requires understanding the context of their actions because students are not independent entities but are usually connected with one another. This naturally leads to the collection and analysis of classroom friendship network data (Allan, 2021; Van der Horst & Coffé, 2012). The popularity of various social media platforms has further made it easier to obtain data on social relationships that exist in the form of networks (Singh et al., 2022; Yeung, Liccardi, Lu, Seneviratne, & Berners-Lee, 2023).

Due to the growing popularity of network data, it is crucial to develop corresponding methods of analysis, especially when network data are combined with non-network individual-level data. Although existing network models such as the Exponential Random Graph Model and the Latent Space Model both have the capacity to incorporate non-network covariates as part of the network formation process (Hoff, Raftery, & Handcock, 2002; Robins, Pattison, Kalish, & Lusher, 2007), their primary focus remains on modeling the network structure rather than explicitly examining the relationships between networks and non-network variables. Similarly, Liu, Jin, and Zhang (2018) explored one potential option in how non-network data could interact and contribute to network formation, but the focus is still on using non-network data as covariates for network formation instead of examining the relationship between network and non-network data. Thus, it is still unclear how more complex relationships between network and non-network data can be assessed comprehensively. Given the widespread use of SEM in social science research on individual-level survey data, we suggest treating networks as unique variables within the SEM framework. This approach enables the specification and estimation of models incorporating both network and non-network data.

In this paper, we will introduce several approaches to analyzing network data in the SEM framework. We will first discuss the models, and then introduce our implementation of the framework in the R package `networksem` Zhang and Xu (2025). Several examples will be provided to illustrate how the package `networksem` can be used.

2. Approaches for Using Network Data with SEM

2.1. Network Data and SEM Notations

Networks are composed of two primary components: nodes (also known as actors or vertices) and edges (also known as links). In a social network, nodes typically represent individuals, while edges represent social relationships between individuals, such as friendships (Tabassum, Pereira, Fernandes, & Gama, 2018).

To analyze network data, the adjacency matrix format is commonly used. An adjacency matrix, denoted as $\mathbf{M}$ in this paper, has rows that represent the nodes from which edges depart, and columns that represent the nodes to which edges terminate. In this setup, the element m_{ij} signifies the value of the edge connecting node i to node j. If the network is non-directional, the matrix $\mathbf{M}$ is symmetric, meaning that

$m_{ij} = m_{ji}$ for all i and j. Conversely, if the network is directed, the matrix $\mathbf{M}$ is often not symmetric.

SEMs are used to analyze regression relationships between latent variables, which are defined by multiple observed variables. In this paper, we employ SEM notation as outlined in Bentler and Weeks (1980) and in Equation 1 for explaining our methods but other specifications can equally work. In this specification, η denotes all endogenous variables, while ξ represents all exogenous variables. This approach treats latent and observed variables equally, without distinguishing between them. Consequently, β represents all coefficients that relate endogenous variables to other endogenous variables, while γ relates exogenous variables to endogenous variables.

$$\boldsymbol{\eta} = \boldsymbol{\beta}\boldsymbol{\eta} + \boldsymbol{\gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{1}$$

In this study, we will incorporate network data into the SEM framework in a two-step approach by first converting network data into SEM-compatible variables, and then using the transformed data in SEM together with non-network variables. We will focus on two main approaches. The first approach involves extracting information from a network based on each node or participant. This information is then used as variables in a SEM model. In this method, each node in the network serves as the primary unit of analysis. The second approach involves extracting information from a network based on each edge or relationship. In this method, each pair of nodes is used as the primary unit of analysis. In both approaches, network data can be used without assuming an underlying model or with a model specification. This paper will discuss both options.

2.2. SEM with Node-Based Network Data

In the first method, each node is the analysis unit in SEM. The non-network component of SEM is the same as in traditional data analysis in Equation 1. However, network data cannot be directly modeled in traditional SEM. Instead, SEM with network data can be summarized in Equation 2. In this equation, $\boldsymbol{\eta}$ is a vector of endogenous variables (both observed and latent), and $\boldsymbol{\xi}$ is a vector of exogenous variables (both observed and latent). In particular, neither $\boldsymbol{\eta}$ nor $\boldsymbol{\xi}$ incorporates network information. The network variables represented by $\mathbf{n}$ typically do not appear as both endogenous and exogenous variables in our model. To avoid confusion, we separated them into two sets–$\mathbf{n}^+$ for endogenous variables and $\mathbf{n}^-$ for exogenous variables. $\boldsymbol{\beta}$ is a matrix of coefficients determining the relationship among endogenous variables and $\boldsymbol{\gamma}$ is a coefficient matrix governing the relationship between endogenous variables and exogenous variables. Because network variables are not measurement items, they will only appear in the structural part of SEM.

$$\begin{pmatrix} \boldsymbol{\eta} \\ \mathbf{n}^+ \end{pmatrix} = \boldsymbol{\beta} \begin{pmatrix} \boldsymbol{\eta} \\ \mathbf{n}^+ \end{pmatrix} + \boldsymbol{\gamma} \begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{n}^- \end{pmatrix} \tag{2}$$

We now go into two distinct approaches to extracting node information from social networks.

2.2.1. Network Statistics Approach

Based on the adjacency matrix $\mathbf{M} = [m_{ij}]$ of a network, many node-based network statistics can be defined (Brinkmeier & Schank, 2005; Wasserman & Faust, 1994). For example, the degree statistic is a centrality measure that simply counts how many other nodes that a specific node connects to in the network: in a friendship network, if a person has a large degree centrality, the person is usually more popular in their social circle and has more friends. Additionally, the betweenness statistic measures the extent to which a node lies on the paths between other node. Nodes with high betweenness influence how the information flows in the network. Both degree and betweenness quantify the importance of a subject in a network. In our model, we use $\mathbf{t}_i(\mathbf{M})$ to represent a vector of network statistics for node i.

When analyzing network statistics within the SEM framework, as illustrated in Equation 2, we have $\mathbf{n} = \mathbf{t}_i(\mathbf{M})$. Since network statistics are node-based, the resulting data will align with the non-network data, which are also based on each individual. Consequently, they can be combined and utilized in SEM. While our current notations are intended for individuals, we will introduce the subscript i to distinguish between the node-based model (with subscript i) and the edge-based model (with subscript ij) from now on.

$$\begin{pmatrix} \boldsymbol{\eta}_i \\ \mathbf{t}_i^+ \end{pmatrix} = \boldsymbol{\beta} \begin{pmatrix} \boldsymbol{\eta}_i \\ \mathbf{t}_i^+ \end{pmatrix} + \boldsymbol{\gamma} \begin{pmatrix} \boldsymbol{\xi}_i \\ \mathbf{t}_i^- \end{pmatrix} \tag{3}$$

2.2.2. Latent Space Model Approach

Another node-based strategy is related to the latent space model (Hoff et al., 2002). In this approach, each node assumes a position in a latent space such as a Euclidean space. The distance of two nodes in the latent space is assumed to be related to how likely they are connected in the network. The idea of latent space modeling is similar to that of factor analysis with a latent factor space and factor scores. Let $\mathbf{z}_i$ be a vector of latent positions of subject i in the latent space. For subjects i and j, the Euclidean distance between them is

$$d_{ij}(\mathbf{z}_i, \mathbf{z}_j) = \sqrt{(\mathbf{z}_i - \mathbf{z}_j)^t(\mathbf{z}_i - \mathbf{z}_j)} = \sqrt{\sum_{d=1}^{D} (z_{i,d} - z_{j,d})^2} \tag{4}$$

where the symbol $(\cdot)^t$ is the transpose of a matrix or vector, D is the dimension of the Euclidean latent space, $\mathbf{z}_i = (z_{i,1}, z_{i,2}, \cdots, z_{i,D})^t$ and $\mathbf{z}_j = (z_{j,1}, z_{j,2}, \cdots, z_{j,D})^t$ are the latent positions of subjects i and j, respectively. With the distance, the latent space model can be written as

$$\begin{cases} m_{ij} & \sim \text{Bernoulli}(p_{ij}) \\ \text{logit}[p(m_{ij})] & = \alpha + \boldsymbol{\beta}'\mathbf{h}_{ij} - \kappa \times d_{ij}(\mathbf{z}_i, \mathbf{z}_j) \end{cases} \tag{5}$$

where α is an intercept, $\mathbf{h}_{ij}$ is a vector of covariates and $\boldsymbol{\beta}$ contains the coefficients of the covariates. Note that the network is assumed to be unweighted here. In this study, following the tradition in network analysis, the coefficient κ for d_{ij} is fixed as 1 because κ can be rescaled together with the distance (Hoff et al., 2002). Therefore, the closer of two subjects are in the latent space, the higher the probability is for them to be connected after controlling the covariates in the model.

In this study, we use a form of the latent space model with only the latent space and an intercept as shown in Equation 6.

$$\begin{cases} E(m_{ij}) & = \mu_{ij} \\ g(\mu_{ij}) & = \alpha - d_{ij}(\mathbf{z}_i, \mathbf{z}_j) \end{cases} \tag{6}$$

In this equation, g is a link function. We assume the connection between two subjects is solely explained by the latent space. The edge value can also be of any exponential family of distributions. Using this model, we can extract information from a network with an idea similar to principal component analysis. In this case, $\mathbf{n_i} = \mathbf{z_i}$ such that the latent positions will be used along with non-network variables in the SEM framework (Equation 3). While the formation in Equation 6 is for undirected networks, directed networks can also be used according to the projection model in Hoff et al. (2002).

$$\begin{pmatrix} \boldsymbol{\eta}_i \\ \mathbf{z}_i^+ \end{pmatrix} = \boldsymbol{\beta} \begin{pmatrix} \boldsymbol{\eta}_i \\ \mathbf{z}_i^+ \end{pmatrix} + \boldsymbol{\gamma} \begin{pmatrix} \boldsymbol{\xi}_i \\ \mathbf{z}_i^- \end{pmatrix} \tag{7}$$

2.3. SEM with Edge-Based Network Data

Alternatively to using nodes as the basis for analyzing network data with SEM, we can also use edges in network data as the unit of interest. In this case, non-network data must be reformatted for analysis to be based on pairs of individuals. In this scenario, we define $\mathbf{c_{ij}} = f(\mathbf{c_i}, \mathbf{c_j})$, where $\mathbf{c_i}$ and $\mathbf{c_j}$ are non-network, subject-based observed variables. The function f can be freely chosen. For instance, $\mathbf{c_{ij}}$ could be the mean of $\mathbf{c_i}$ and $\mathbf{c_j}$ corresponding to hypotheses using the joint levels of non-network covariates, or it could be the difference, corresponding to hypothesis attempting to understand homophily effects. Subsequently, these pairwise non-network variables can be utilized as either endogenous ($\boldsymbol{\eta_{ij}}$) or exogenous ($\boldsymbol{\xi_{ij}}$) variables. The subscripts of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are modified to accommodate the transition from node-based to edge-based analysis.

$$\begin{pmatrix} \boldsymbol{\eta}_{ij} \\ \mathbf{n}_{ij}^+ \end{pmatrix} = \boldsymbol{\beta} \begin{pmatrix} \boldsymbol{\eta}_{ij} \\ \mathbf{n}_{ij}^+ \end{pmatrix} + \boldsymbol{\gamma} \begin{pmatrix} \boldsymbol{\xi}_{ij} \\ \mathbf{n}_{ij}^- \end{pmatrix} \tag{8}$$

2.3.1. Network Statistics Approach

In the edge-based method, similar to in the node-based method, network statistics that are obtained free from assuming underlying social network models can be used in SEM. However, these statistics are now based on each pair of subjects, rather than each subject individually. We'll refer to these statistics as $\mathbf{n} = \mathbf{t_{ij}}(\mathbf{M})$. For instance, the shortest path between each pair of nodes can be used as an edge-based network statistic. A more intuitive way to model networks in SEM using edges as units is to use the edge value itself, so $\mathbf{n_{ij}} = \mathbf{t_{ij}}(\mathbf{M}) = \mathbf{M_{ij}}$. With this approach, the network data can be combined with the reformatted covariates data $\boldsymbol{\xi_{ij}}$ and $\boldsymbol{\eta_{ij}}$ in SEM.

$$\begin{pmatrix} \boldsymbol{\eta}_{ij} \\ \mathbf{t}^{+}_{ij} \end{pmatrix} = \boldsymbol{\beta} \begin{pmatrix} \boldsymbol{\eta}_{ij} \\ \mathbf{t}^{+}_{ij} \end{pmatrix} + \boldsymbol{\gamma} \begin{pmatrix} \boldsymbol{\xi}_{ij} \\ \mathbf{t}^{-}_{ij} \end{pmatrix} \tag{9}$$

2.3.2. Latent Space Model Approach

The latent space modeling approach can again be used when using a pair of subjects as the unit of analysis. In this case, the latent distance between two subjects $\mathbf{d_{ij}}(\mathbf{z_i}, \mathbf{z_j})$ can be used in SEM instead of the latent positions $\mathbf{z_i}$ and $\mathbf{z_j}$, so $\mathbf{n_{ij}} = \mathbf{d_{ij}}$.

$$\begin{pmatrix} \boldsymbol{\eta}_{ij} \\ \mathbf{d}^{+}_{ij} \end{pmatrix} = \boldsymbol{\beta} \begin{pmatrix} \boldsymbol{\eta}_{ij} \\ \mathbf{d}^{+}_{ij} \end{pmatrix} + \boldsymbol{\gamma} \begin{pmatrix} \boldsymbol{\xi}_{ij} \\ \mathbf{d}^{-}_{ij} \end{pmatrix} \tag{10}$$

3. Model Estimation and Software Implementation

We propose a two-stage method to estimate SEM with network data. In the first stage, methods pertaining to network analysis are used to extract network information, either through the network statistics or the latent space model. In the second stage, the extracted information is then inputted to the SEM models as observed variables to estimate the SEM-related model parameters. Through the two-stage method, many existing methods for network and SEM analysis can be applied directly.

For example, in the first stage, network statistics $\mathbf{t}$ can be obtained from the existing R package `sna` (Butts, 2020) for network analysis. To use the latent space modeling approach, the latent positions can be estimated using the R package `latentnet` (Krivitsky & Handcock, 2008). In this stage, one can determine the dimension of the latent space by fitting models with different numbers of dimensions. In the second stage, SEM analysis with the obtained network statistics can be conducted using the R package `lavaan` (Rosseel, 2012). The main advantages of the two-stage method include ease of use, flexibility, and the ease of adopting new developments in both SEM and network analysis techniques.

To ease the use of SEM with network data, we have developed an R package `networksem` (Zhang & Xu, 2025). It can be installed from GitHub or CRAN.

3.1. The R Package `networksem`

The package `networksem` provides four separate functions corresponding to the four approaches discussed in the previous section.

- `sem.net`: Fit SEM with both network and non-network data by incorporating node-level network statistics as variables.
- `sem.net.edge`: Fit SEM with both network and non-network data by transforming non-network data into paired values corresponding to network edge values.
- `sem.net.lsm`: Fit SEM with both network and non-network data by incorporating network latent positions as variables.
- `sem.net.edge.lsm`: Fit SEM with both network and non-network data by transforming non-network data into paired values corresponding to network latent distance pairs.

There are some common arguments shared by these four functions, which are listed below.

- `model`: A model specified using the `lavaan` model syntax.
- `data`: A list containing both the network data and the non-network data.
- `netstats.rescale`: A logical value (TRUE or FALSE) indicating whether to rescale network statistics to have mean 0 and sd 1.
- `data.rescale`: A logical value (TRUE or FALSE) indicating whether to rescale all data to have mean 0 and sd 1.

For the function `sem.net`, the arguments pertaining network statistics including `netstats` can be used to specify network statistics. The `degree`, `betweenness`, `closeness`, `evcent`, `infocent`, and `stresscent` options for network statistics are adapted from the R package `sna` (Butts, 2020) and the options `ivi`, `hubeness.score`, `spreading.score`, and `clusterRank` are adapted from the R package `influential` (Salavaty, Ramialison, & Currie, 2020). Internally, when specifying the corresponding network statistics terms in the `networksem` package, the functions calls the `sna` or `influential` package and retrieve corresponding values from there.

For the function `sem.net.edge` and `sem.net.edge.lsm`, an argument `type` specifying whether to use average or difference across non-network variables with network edge values in SEM can be used. For `sem.net.lsm` and `sem.net.edge.lsm`, an argument called `latent.dim` can be used to specify how many latent dimensions should be used in the latent space model.

In the `networksem` package, we provide a `summary()` function on objects returned by the previously mentioned main functions to view the results directly. The function will return SEM results and LSM results in proper situations. Additionally, the function `path.networksem()` can take the `networkse` object as well as the intended predictor, mediator, and outcome as argument to output a calculated indirect effect from the model.

3.2. The BigSEM Web Application

An online wrapper for the `networksem` package is also developed and can be assessed via the website `https://bigsem.psychstat.org/`. The web app has a graphical interface that allows users to draw path diagrams as models, and the models can be fitted based on the diagrams. Specific instructions on how to use the interface can be found on the website and will not be discussed in the paper.

4. Examples and Applications

4.1. Node-Based Network Statistics Approach

4.1.1. Friendship Network Data

In this example, a friendship network collected as part of a larger data collection project will be used. The dataset contains 165 participants including 45% males and 55% females with an average age of 21.64 years (standard deviation = 0.85). Besides a friendship network, responses from 4 items measuring extroversion from the mini-IPIP scale (Donnellan, Oswald, Baird, & Lucas, 2006) and 7 items measuring depression from the Patient Health Questionnaire (Kroenke, Spitzer, & Williams, 2001) are included. The dataset is part of the `networksem` package.

The dataset is already cleaned to match format of `networksem`. After loading, the list `friend_data` contains two sub-lists: `network` includes the friendship network of interest in the adjacency matrix format, and `nonnetwork` contains all individual-level non-network data in a variable by individual dataframe format. In this example, the network and non-network data are already in the formats desired so they do not need to be transformed. However, if network data are in alternative formats such as an edge list and non-network data are in other formats, then they need to be reformatted first.

```
data('friend_data')
head(friend_data$non_network)
View(friend_data$network)
```

4.1.2. Model Fitting

In this example, we apply the node-based approach with network statistics to the friendship network data from Section 4.1.1. The **measurement component** of this model consists of 4 observed variables `personality1`, `personality6`, `personality11` and `personality16` informing the latent personality trait of `Extroversion`. Additionally, the latent variable of `Depress` denoting depression level is formed by 7 observed items inquiring respondents about their depression status. A social network mediation model is considered in the **structural component** of the model where extroversion acts as the predictor, the friendship network acts as the mediator, and depression acts as the outcome. Based on this setup, the friendship network can be viewed as a mediator linking personalities' effect to happiness. Model specification corresponding to the above description can be put in `lavaan` syntax as a string variable shown below.

```
model <-'
  Extroversion =~ personality1 + personality6
                + personality11 + personality16
  Depress =~ depress1 + depress2 + depress3 + depress4 + depress5 + depress6 +
      depress7
  friends ~ Extroversion
  Depress ~ friends + Extroversion
'
```

The function `sem.net` in the `networksem` package can be used to fit the current model as shown by the code below. Data in the format specified in Section 4.1.1 is used. The desired network statistics to be used can also be specified as in the argument `netstats=c("degree")`. For the degree centrality measure, we can further designate a mode for it using the command `netstats.options=list("degree"=list("cmode"="freeman"))`. The argument `netstats.rescale = T` is used so that the network statistics will be scaled to have means of 0 and standard deviations of 1. Results are saved in the `res` variable here.

```
res <- sem.net(model=model, data=friend_data,
               netstats=c("degree"),
               netstats.rescale = T,
               netstats.options=list("degree"=list("cmode"="freeman")))
```

A summary function `summary(res)` can be used to inspect results from the `sem.net` output. Results are in `lavaan` format and users can refer to Rosseel (2014). Model fit can be assessed from the output as shown below. The chi-square test shows a non-significant p-value, suggesting the model does not deviate significantly from the data ($\chi^2 = 64.549, df = 52, p = 0.114$). Alternatively, fit indices can be used, suggesting acceptable fit ($CFI = 0.955, TLI = 0.943, RMSEA = 0.038, SRMR = 0.062$).

```
Model Test User Model:

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

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)              -2263.328
  Loglikelihood unrestricted model (H1)      -2231.053

  Akaike (AIC)                                4578.655
  Bayesian (BIC)                              4659.410
  Sample-size adjusted Bayesian (SABIC)       4577.093

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:

  SRMR                                           0.062
```

Parameter estimates can also be obtained from the results as shown below. The factor loadings for depressions and extroversion are all significant, with depression items in general have higher loadings than extroversion items. For the mediation model, only the path predicting friendship network degree statistics with extroversion is significant ($a = 0.403, p = 0.001$). The effect of extroversion predicting depression and the effect of the friendship network degree predicting depression are both not significant in the results.

```
Parameter Estimates:

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

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  Depress =~
    depress7          1.000
    depress6          1.175    0.269    4.362    0.000
    depress5          1.275    0.277    4.605    0.000
    depress4          1.309    0.270    4.856    0.000
    depress3          1.385    0.309    4.474    0.000
    depress2          1.054    0.235    4.490    0.000
    depress1          0.709    0.190    3.732    0.000
  Extroversion =~
    personality16     1.000
```

```
    personality11      0.763     0.176     4.341     0.000
    personality6      -0.648     0.156    -4.141     0.000
    personality1      -0.549     0.142    -3.863     0.000

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  Depress ~
    Extroversion      0.006    0.046    0.131    0.896
  friends.degree ~
    Extroversion      0.403    0.125    3.214    0.001
  Depress ~
    friends.degree    0.008    0.031    0.263    0.793
```

4.2. Node-Based LSM Approach

4.2.1. UK Faculty Network Data

The UK faculty network dataset includes personal friendship data of 81 faculty members and can be found publicly as part of the `igraphdata` package (Csardi & Csardi, 2015; Nepusz, Petróczi, Négyessy, & Bazsó, 2008). It contains unweighted edges of 1 indicating friendship and 0 indicating no friendship. It also contains one nodal covariate of which school the faculty belongs to. The data is converted to formats required by `networksem` with the code below. Because `Group` is a categorical variable with three categories, we create two dummy variables `Group2` and `Group3` for it, with two missing values treated as in the reference category which has most participants.

```
library(igraphdata)
data(UKfaculty)
nonnet <- as.data.frame(igraph::get.vertex.attribute(UKfaculty))
nonnet$Group2 <- ifelse(nonnet$Group == 2, 1, 0)
nonnet$Group3 <- ifelse(nonnet$Group == 3, 1, 0)
net <- as.matrix(as_adjacency_matrix(UKfaculty))
uknet = list(network = list(net = net), nonnetwork = nonnet)
```

4.2.2. Model Fitting

A regression model using school affiliation to predict faculty network is used here. Instead of network statistics, for the function `sem.net.lsm`, the argument `latent.dim` can be used to specify the number of latent dimensions. The LSM has a random component in estimation, so a random seed is used here for replication purposes.

```
model <-'
  net ~ Group2 + Group3
'
set.seed(100)
res <- sem.net.lsm(model = model, data = uknet, latent.dim = 2)
```

Again, the function `summary(res)` can be used to obtain estimates. Because this model assumes no latent variables, SEM model fit from the output becomes irrelevant, and only the regression coefficient estimates and the LSM fits are shown below. In this case, school affiliation's effect on the latent position are significant for both latent dimensions. This means that faculty members of schools 2 and 3 both have different latent positions compared to members of school 1. For the LSM components, the intercept estimate of 1.906 translates to $g(\mu_{ij} = 1.906 - |z_i, z_j|^2$, with a BIC of 3606.878.

```
The SEM output:
Regressions:
```

```
                 Estimate  Std.Err  z-value  P(>|z|)
  net.Z1 ~
    Group2           4.518    0.523    8.645    0.000
  net.Z2 ~
    Group2           3.889    0.469    8.293    0.000
  net.Z1 ~
    Group3          -2.112    0.581   -3.633    0.000
  net.Z2 ~
    Group3           7.911    0.522   15.163    0.000

The LSM output:

==========================
Summary of model fit
==========================

Formula:   network::network(data$network[[latent.network[i]]]) ~ euclidean(d =
    latent.dim)
<environment: 0x7fe22a49af68>
Attribute: edges
Model:     Bernoulli
MCMC sample of size 4000, draws are 10 iterations apart, after burnin of 10000
    iterations.
Covariate coefficients posterior means:
            Estimate   2.5%  97.5% 2*min(Pr(>0),Pr(<0))
(Intercept)   1.9055 1.6736 2.1386            < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Overall BIC:        3606.878
Likelihood BIC:     2783.433
Latent space/clustering BIC:     823.4445

Covariate coefficients MKL:
            Estimate
(Intercept) 1.421039
```

4.3. Edge Values Approach

4.3.1. Attorney Network Data

This dataset includes the cowork and advice networks from 71 attorneys in the law firm called SG&R, collected in the year 1988, and is available at the SIENA website `https://www.stats.ox.ac.uk/~snijders/siena/siena_datasets.htm`. The first wave of network data will be used in the analysis in the current study. The cowork network is collected by asking the company employees to select people who have worked on the same case with them. Information on an advice network is collected via asking respondents who they seek advice from at work. In addition to the networks, several non-network attributes such as gender and the number of years spent with the firm are collected.

In preparing this dataset, we still use the `non_network` variable to store non-network data and the `network` variable to store the two networks, `advice` for the advice network and `cowork` for the cowork network.

```
non_network <- read.table("data/attorney/ELattr.dat")[,c(3,5)]
colnames(non_network) <- c('gender', 'years')
non_network$gender <- non_network$gender - 1 # change gender to 0 and 1.
network <- list()
network$advice <- read.table("data/attorney/ELadv.dat")
```

```
network$cowork <- read.table("data/attorney/ELwork.dat")
attorneynet <- list(network = network, nonnetwork = non_network)
```

4.3.2. Model Fitting

In this example, the advice network is predicted by gender and years in practice, whereas the cowork network is predicted by the advice network, gender, and years in practice all together. In this case, the advice network acts as a mediator, while gender and years in practice exert indirect effect onto the cowork network through the advice network in addition to having direct effects. The model string is shown below.

```
model <-"
  advice ~ gender + years
  cowork ~ advice + gender + years
"
```

The corresponding function to run an edge-based model is `sem.net.edge()` with the `type` argument needed to specify whether the covariate values to be run with the social network edge values in SEM should be calculated as the "difference" between two individuals or the "average" across the two individuals. The "difference" option is used here because we hypothesize homophily and heterophily effects here for the cowork network and the advice network: people with same gender are more likely to cowork and advise each other, whereas people with closer years in practice may be more likely to cowork but less likely to advise each other. In other cases when the joint level of attributes are hypothesized to influence network formation or vice versa, the "average" option can be used. The "difference" option here is similar to the "absdiff" term in Exponential Random Graph Models, and the option "average" is similar to the "nodecov" term (Morris, Handcock, & Hunter, 2008). Finally, the argument `ordered = c("cowork", "advice")` is used to tell `lavaan` that the outcome variables cowork and advice are binary, considering edge values are used in this approach.

```
res <- sem.net.edge(model = model, data = attorneynet,
                    data.rescale = F,
                    netstats.rescale = T,
                    network = network, type = "difference",
                    ordered = c("cowork", "advice"))
```

The SEM output can be interpreted similarly as in the previous examples using the `summary(res)` function, except that now the unit of analysis is each pair of individuals. Because no latent variables are used, we will only look at the regression output from the results as shown below. Years in practice's influence on the advice network is significant and negative ($a = -0.014, p = 0.000$), meaning that people with closer years in practice tend to advise each other more; and years in practice also has a positive effect on cowork ($c = 0.020, p = 0.000$), suggesting that people with larger difference in years of practice tend to cowork more. This is opposite to what we hypothesized but not impossible given that lawyers may find it easier to ask for advice from someone with similar experience level, whereas more experienced lawyers may bring less experienced lawyers on their case for learning purposes. Gender's effect on the advice network is significant ($a = -0.248, p = 0.000$), suggesting that people tend to seek advice from the same gender, which is what we expected. Advice network, being the mediator, also has a significant influence on the cowork network ($b = 0.687, p = 0.000$).

```
Regressions:
                  Estimate  Std.Err  z-value  P(>|z|)
  advice ~
    gender          -0.248    0.044   -5.642    0.000
```

```
    years             -0.014    0.003   -5.452    0.000
  cowork ~
    advice             0.687    0.019   36.675    0.000
    gender             0.023    0.043    0.533    0.594
    years              0.020    0.003    7.422    0.000

Thresholds:
                   Estimate  Std.Err  z-value  P(>|z|)
    advice|t1          0.698    0.036   19.516    0.000
    cowork|t1          1.090    0.040   27.072    0.000
```

4.4. Edge-Based LSM Approach

4.4.1. Florentine Marriage Data

This dataset of marital relationships from 16 Florentine families is from Breiger and Pattison (1986) and can be found in the R package `ergm` (Hunter, Handcock, Butts, Goodreau, & Morris, 2008). The non-network variables include wealth and priorates (number of seats on the civic council). The data can be prepared using the code below.

```
library(ergm)
data("florentine")
network <- list()
network$flo <- as.matrix.network.adjacency(flomarriage)
nonnetwork <- data.frame(
    name = get.vertex.attribute(flomarriage, "vertex.names"),
    wealth = get.vertex.attribute(flomarriage, "wealth"),
    priorates = get.vertex.attribute(flomarriage, "priorates")
)
```

4.4.2. Model Fitting

A regression model can be used such that wealth predicts the marriage network and the marriage network in turn predicts priorates. The model can be written in the form below.

```
https://www.overleaf.com/project/660420aae1e6bc80a746836d#
model <- '
  flo ~  wealth
  priorates ~ flo + wealth
'
```

When fitting the model , the function to be used is `sem.net.edge.lsm` where the argument `type` and `latent.dim` are needed. We again use `type = "difference"` to capture homophily effects. Here, although the marriage network contains binary edges, the `ordered` argument is not needed since only the continuous latent distances will be used in the SEM. A random seed is needed for using LSM.

```
set.seed(100)
res <- sem.net.edge.lsm(model=model, type = "difference", data=data, latent.dim = 2)
```

To interpret output from the model, only wealth's effect on priorates is significant ($a = 0.387, p = 0.000$). Latent distance related effects are not significant.

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  priorates ~
    wealth             0.387    0.058    6.738    0.000
  flo.dists ~
```

```
    wealth            0.059    0.062    0.942    0.346
  priorates ~
    flo.dists         0.060    0.058    1.046    0.296

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .priorates         0.840    0.074   11.314    0.000
   .flo.dists         0.993    0.088   11.314    0.000

The LSM output:

==========================
Summary of model fit
==========================

Formula:   network::network(data$network[[latent.network[i]]]) ~ euclidean(d =
    latent.dim)
<environment: 0x7fbd917d05c0>
Attribute: edges
Model:     Bernoulli
MCMC sample of size 4000, draws are 10 iterations apart, after burnin of 10000
    iterations.
Covariate coefficients posterior means:
            Estimate   2.5%  97.5% 2*min(Pr(>0),Pr(<0))
(Intercept)   5.0133 2.5627 7.9665            < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Overall BIC:        259.7975
Likelihood BIC:     85.53086
Latent space/clustering BIC:     174.2666

Covariate coefficients MKL:
            Estimate
(Intercept) 2.861026
```

4.5. Additional Analysis

The path diagram with estimated parameters can be visualized in R using the following code with the help of the `RAMpath` (Zhang, Hamagami, Grimm, & McArdle, 2015) and `DiagrammeR` packages (Iannone & Roy, 2015). Here, using the first example of applying the node-based approach using Network statistics in SEM with the friendship dataset, the plotting process can be done as shown below.

```
plot.res <- lavaan2ram(res$estimates, ram.out = F)
plot.res.path <- ramPathBridge(plot.res, F, F)
plot(plot.res.path, "ex1", output.type="dot")
grViz("ex1.dot")
```

Furthermore, our package `networksem` also provides the functionality to calculate indirect effects in SEM with network data. This is done through the `path.networksem()` function with an example shown below.

```
> path.networksem(res, 'Extroversion', 'friends.degree', 'Depress')
     predictor       mediator outcome     apath       bpath    indirect indirect_se
1 Extroversion friends.degree Depress 0.4026832 0.008188015 0.003297176  0.05045369
  indirect_z
1 0.06535054
```

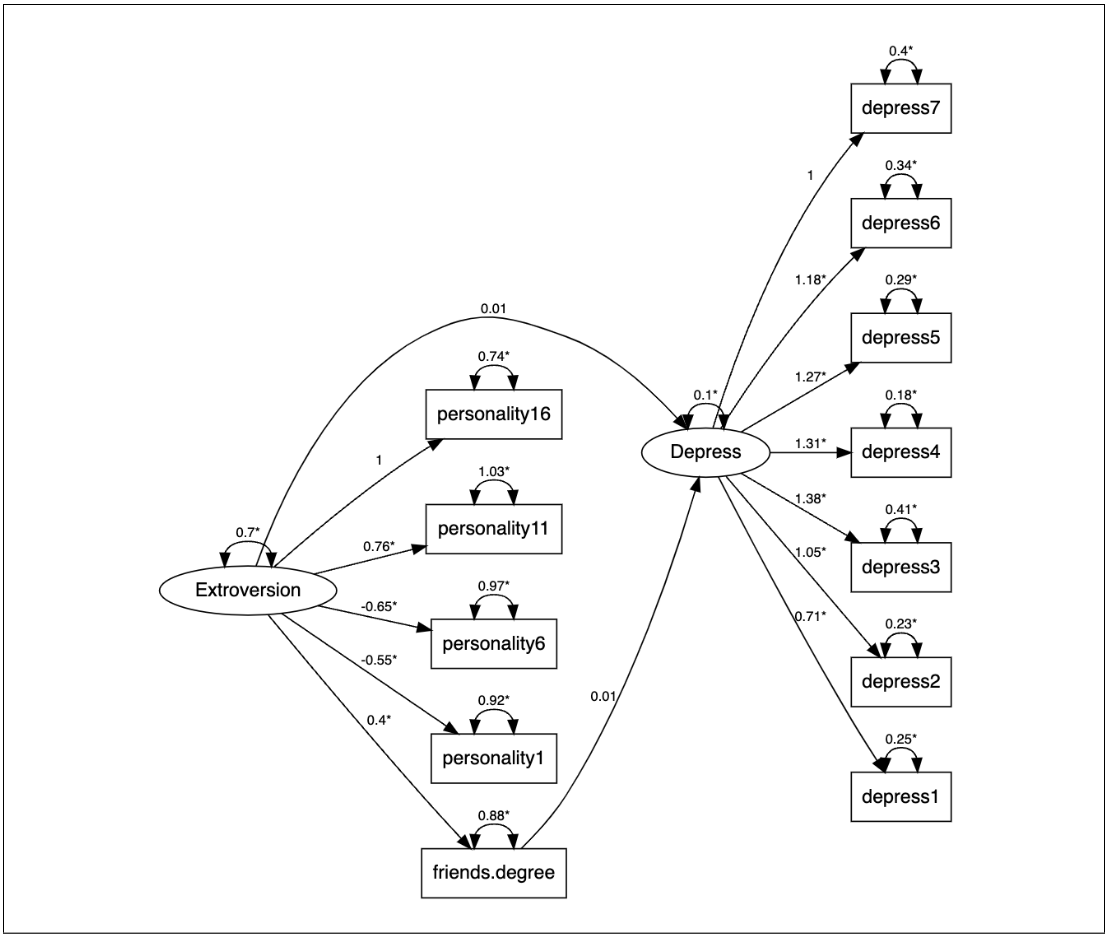

Figure 1. Path diagram of SEM with node-based network statistics approach for the friendship network.

5. Discussion

To address the growing popularity of complex data forms in social sciences and the lack of systematic approaches for integrating such data with more traditional individual-based survey data, the current paper introduces a two-stage method to incorporate network data into SEM analysis. Network data can be transformed into vector forms compatible with non-network data either based on the nodes, or based on the edges between pairs of nodes. We propose four methods to transform network data in this paper. The node-based network statistics approach calculates network statistics such as degree and betweenness that are for each node, and then use them in the SEM analysis with non-network data. The node-based LSM approach finds latent positions of the nodes in a network, and then use the position coordinates as variables in SEM. The edge value approach keeps network edges, but transforms non-network variables either into pairwise differences or pairwise averages to be used with an edge as the unit. The edge-based LSM approach finds the pairwise latent distance between nodes, and use that along with the pairwise transformed non-network data, in the SEM analysis. Implementations of the four different methods can be found in our R package `networksem` and in our web app BigSEM.

The two-stage methods introduced in the current paper are flexible and can be easily extended. For example, instead of using the LSM to represent the network data, relevant parameters in an Exponential Random Graph Model or a Stochastic Block Model can be used. However, the flexibility comes at some cost such that if a one-stage integrated estimation method is used, the resulting estimates may be more accurate. In addition, for the edge-based approaches, the number of observations will increase after data transformation that takes the data from being based on each individual to being based on each pair of individuals. The resulting transformed data does not have independence among observations, and the increased number of observations can also lead to longer computation time for the models. Further, it is difficult to distinguish which model underlies the networks observed and thus which of the four methods to use may be largely dependent on theory.

Future research could expand methods in the current paper by substituting alternative network models into the framework. It would also be of interest to develop one-stage estimation methods such as Bayesian methods for the joint estimation of network data and non-network data in SEM. Additionally, model comparison metrics can be explored and evaluated in the context of network data in SEM analysis. We also recognize that network data are not the only form of complex data being adopted into social science research, and a generalizable framework to integrate different data forms could be highly useful.

References

Allan, G. A. (2021). *A sociology of friendship and kinship.* Routledge.

Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures*, *16*(1), 1–18.

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, *45*(3), 289–308.

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (2013). *Measurement errors in surveys.* John Wiley & Sons.

Bollen, K. A. (1989). *Structural equations with latent variables.* John Wiley & Sons.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In *Handbook of econometrics* (Vol. 5, pp. 3705–3843). Elsevier.

Breiger, R. L., & Pattison, P. E. (1986). Cumulated social roles: The duality of persons and their algebras. *Social networks*, *8*(3), 215–256.

Brinkmeier, M., & Schank, T. (2005). Network statistics. In *Network analysis: methodological foundations* (pp. 293–317). Springer.

Butts, C. T. (2020). sna: Tools for social network analysis [Computer software manual]. (R package version 2.6)

Che, C., Jin, I. H., & Zhang, Z. (2020, February). Network mediation analysis using model-based eigenvalue decomposition. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(1), 148–161.

Clifton, A., & Webster, G. D. (2017, May). An introduction to social network analysis for personality and social psychologists. *Social Psychological and Personality Science*, *8*(4), 442–453.

Csardi, G., & Csardi, M. G. (2015). Package 'igraphdata'. *yeast*, *13*, 1.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, *40*, 434–449.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality. *Psychological assessment*, *18*(2), 192.

Fairhurst, M., Li, C., & Da Costa-Abreu, M. (2017). Predictive biometrics: a review and analysis of predicting personal characteristics from biometric data. *IET Biometrics*, *6*(6), 369–378.

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences.* Princeton University Press.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology.* John Wiley & Sons.

Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., Ray, S., ... others (2021). An introduction to structural equation modeling. *Partial least squares structural equation modeling (PLS-SEM) using R: a workbook*, 1–29.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association*, *97*(460), 1090–1098.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of statistical software*, *24*, 1–29.

Iannone, R., & Roy, O. (2015, January). *Diagrammer: Graph/network visualization.* The R Foundation.

Kim, J., & Kim, N. (2022). Quantifying emotions in architectural environments using biometrics. *Applied Sciences*, *12*(19), 9998.

Kline, R. B. (2023). *Principles and practice of structural equation modeling.* Guilford publications.

Krivitsky, P. N., & Handcock, M. S. (2008). Fitting position latent cluster models for social networks with latentnet. *Journal of statistical software*, *24*(5), 1–23.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The phq-9: validity of a brief depression severity measure. *Journal of general internal medicine*, *16*(9), 606–613.

Liu, H., Jin, I. H., & Zhang, Z. (2018). Structural equation modeling of social networks: Specification, estimation, and application. *Multivariate behavioral research*, *53*(5), 714–730.

Liu, H., Jin, I. H., Zhang, Z., & Yuan, Y. (2021, March). Social network mediation analysis: A latent space approach. *Psychometrika*, *86*(1), 272–298.

Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: terms and computational aspects. *Journal of statistical software*, *24*, 1–24.

Nepusz, T., Petróczi, A., Négyessy, L., & Bazsó, F. (2008). Fuzzy communities and the concept of bridgeness in complex networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, *77*(1), 016107.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social networks*, *29*(2), 173–191.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Rosseel, Y. (2014). The lavaan tutorial. *Department of Data Analysis: Ghent University*.

Salavaty, A., Ramialison, M., & Currie, P. D. (2020). Integrated value of influence: An integrative method for the identification of the most influential nodes within networks. *Patterns*.

Singh, D. K. S., Nithya, N., Rahunathan, L., Sanghavi, P., Vaghela, R. S., Manoharan, P., ... Tunze, G. B. (2022). Social network analysis for precise friend suggestion for twitter by associating multiple networks using ml. *International Journal of Information Technology and Web Engineering (IJITWE)*, *17*(1), 1–11.

Sweet, T. M. (2016, September). Social network methods for the educational and psychological sciences. *Educational Psychologist*, *51*(3–4), 381–394.

Tabassum, S., Pereira, F. S., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(5), e1256.

Van der Horst, M., & Coffé, H. (2012). How friendship network characteristics influence subjective well-being. *Social Indicators Research*, *107*, 509–529.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge university press.

Yeung, C.-m. A., Liccardi, I., Lu, K., Seneviratne, O., & Berners-Lee, T. (2023). Decentralization: The future of online social networking. In *Linking the world's information: Essays on tim berners-lee's invention of the world wide web* (pp. 187–199).

Zhang, Z., Hamagami, F., Grimm, K. J., & McArdle, J. J. (2015). Using r package rampath for tracing sem path diagrams and conducting complex longitudinal data analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 132–147.

Zhang, Z., & Xu, Z. (2025, January). networksem: Network structural equation modeling [Computer software manual]. The R Foundation.

Acknowledgments

This work was supported by a grant from the Department of Education (R305D210023). However, the content of this document does not necessarily represent the policy of the Department of Education. The study was also supported by the Notre Dame Global.

References

Allan, G. A. (2021). *A sociology of friendship and kinship.* Routledge.

Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures*, *16*(1), 1–18.

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, *45*(3), 289–308.

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (2013). *Measurement errors in surveys.* John Wiley & Sons.

Bollen, K. A. (1989). *Structural equations with latent variables.* John Wiley & Sons.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In *Handbook of econometrics* (Vol. 5, pp. 3705–3843). Elsevier.

Breiger, R. L., & Pattison, P. E. (1986). Cumulated social roles: The duality of persons and their algebras. *Social networks*, *8*(3), 215–256.

Brinkmeier, M., & Schank, T. (2005). Network statistics. In *Network analysis: methodological foundations* (pp. 293–317). Springer.

Butts, C. T. (2020). sna: Tools for social network analysis [Computer software manual]. (R package version 2.6)

Che, C., Jin, I. H., & Zhang, Z. (2020, February). Network mediation analysis using model-based eigenvalue decomposition. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(1), 148–161.

Clifton, A., & Webster, G. D. (2017, May). An introduction to social network analysis for personality and social psychologists. *Social Psychological and Personality Science*, *8*(4), 442–453.

Csardi, G., & Csardi, M. G. (2015). Package 'igraphdata'. *yeast*, *13*, 1.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, *40*, 434–449.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality. *Psychological assessment*, *18*(2), 192.

Fairhurst, M., Li, C., & Da Costa-Abreu, M. (2017). Predictive biometrics: a review and analysis of predicting personal characteristics from biometric data. *IET Biometrics*, *6*(6), 369–378.

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences.* Princeton University Press.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology.* John Wiley & Sons.

Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., Ray, S., ... others (2021). An introduction to structural equation modeling. *Partial least squares structural equation modeling (PLS-SEM) using R: a workbook*, 1–29.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association*, *97*(460), 1090–1098.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of statistical software*, *24*, 1–29.

Iannone, R., & Roy, O. (2015, January). *Diagrammer: Graph/network visualization.*

The R Foundation.
Kim, J., & Kim, N. (2022). Quantifying emotions in architectural environments using biometrics. *Applied Sciences*, *12*(19), 9998.
Kline, R. B. (2023). *Principles and practice of structural equation modeling.* Guilford publications.
Krivitsky, P. N., & Handcock, M. S. (2008). Fitting position latent cluster models for social networks with latentnet. *Journal of statistical software*, *24*(5), 1–23.
Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The phq-9: validity of a brief depression severity measure. *Journal of general internal medicine*, *16*(9), 606–613.
Liu, H., Jin, I. H., & Zhang, Z. (2018). Structural equation modeling of social networks: Specification, estimation, and application. *Multivariate behavioral research*, *53*(5), 714–730.
Liu, H., Jin, I. H., Zhang, Z., & Yuan, Y. (2021, March). Social network mediation analysis: A latent space approach. *Psychometrika*, *86*(1), 272–298.
Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: terms and computational aspects. *Journal of statistical software*, *24*, 1–24.
Nepusz, T., Petróczi, A., Négyessy, L., & Bazsó, F. (2008). Fuzzy communities and the concept of bridgeness in complex networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, *77*(1), 016107.
Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social networks*, *29*(2), 173–191.
Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.
Rosseel, Y. (2014). The lavaan tutorial. *Department of Data Analysis: Ghent University*.
Salavaty, A., Ramialison, M., & Currie, P. D. (2020). Integrated value of influence: An integrative method for the identification of the most influential nodes within networks. *Patterns*.
Singh, D. K. S., Nithya, N., Rahunathan, L., Sanghavi, P., Vaghela, R. S., Manoharan, P., ... Tunze, G. B. (2022). Social network analysis for precise friend suggestion for twitter by associating multiple networks using ml. *International Journal of Information Technology and Web Engineering (IJITWE)*, *17*(1), 1–11.
Sweet, T. M. (2016, September). Social network methods for the educational and psychological sciences. *Educational Psychologist*, *51*(3–4), 381–394.
Tabassum, S., Pereira, F. S., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(5), e1256.
Van der Horst, M., & Coffé, H. (2012). How friendship network characteristics influence subjective well-being. *Social Indicators Research*, *107*, 509–529.
Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge university press.
Yeung, C.-m. A., Liccardi, I., Lu, K., Seneviratne, O., & Berners-Lee, T. (2023). Decentralization: The future of online social networking. In *Linking the world's information: Essays on tim berners-lee's invention of the world wide web* (pp. 187–199).
Zhang, Z., Hamagami, F., Grimm, K. J., & McArdle, J. J. (2015). Using r package rampath for tracing sem path diagrams and conducting complex longitudinal

data analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 132–147.

Zhang, Z., & Xu, Z. (2025, January). networksem: Network structural equation modeling [Computer software manual]. The R Foundation.