Undirected network models

SEM 2 2020

Sacha Epskamp

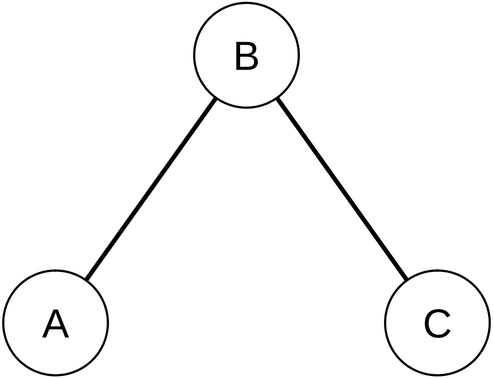
- Directed networks are well usable to display and test a confirmatory theory
  - E.g., Structural equation modeling

- Directed networks are well usable to display and test a confirmatory theory
  - E.g., Structural equation modeling
- But exploratively finding such a model is hard without stringent assumptions
  - Acyclicity (A node does not cause itself directly or indirectly)
  - Many equivalent models, all differently parameterized

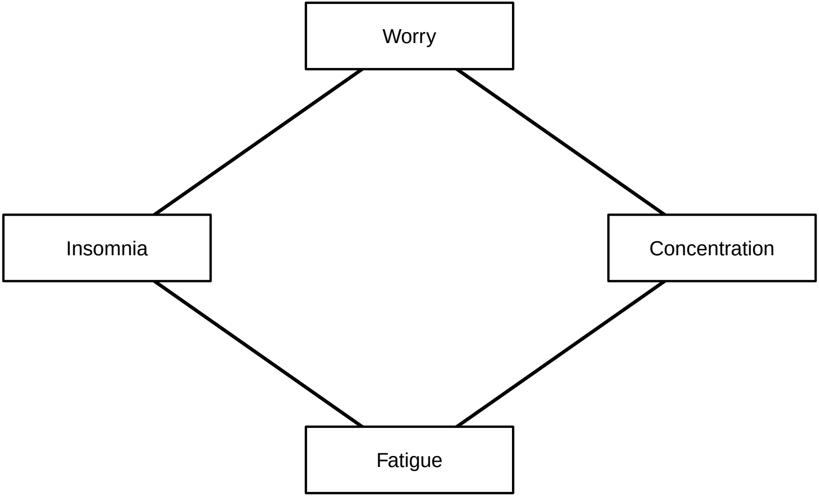
- Directed networks are well usable to display and test a confirmatory theory
  - E.g., Structural equation modeling
- But exploratively finding such a model is hard without stringent assumptions
  - Acyclicity (A node does not cause itself directly or indirectly)
  - Many equivalent models, all differently parameterized
- Exploratory estimation of undirected networks (not DAG skeletons) is an attractive alternative
  - No equivalent models
  - No assumption of acyclicity
  - Straightforward to estimate
  - Well parameterized

- Directed networks are well usable to display and test a confirmatory theory
  - E.g., Structural equation modeling
- But exploratively finding such a model is hard without stringent assumptions
  - Acyclicity (A node does not cause itself directly or indirectly)
  - Many equivalent models, all differently parameterized
- Exploratory estimation of undirected networks (not DAG skeletons) is an attractive alternative
  - No equivalent models
  - No assumption of acyclicity
  - Straightforward to estimate
  - Well parameterized
- When data are multivariate Gaussian, we term the model a Gaussian graphical model and can use partial correlation coefficients to encode the strength of connection in a network

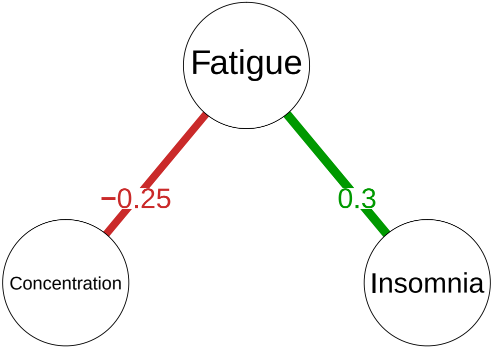
Markov Random Fields



- Two nodes are connected if they are not independent conditional on all other nodes.
- $A \perp\!\!\!\perp C \mid B$



- Worrying and fatigue separate Insomnia and Concentration



- Assuming $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Similar model to SEM:
    - $\boldsymbol{\Sigma} = \boldsymbol{\Delta} (\boldsymbol{I} - \boldsymbol{\Omega})^{-1} \boldsymbol{\Delta}$
- $\boldsymbol{\Omega}$ is a symmetrical matrix that encodes partial correlations, also used as edge-weights

The Gaussian graphical model uses partial correlation coefficients (correlation between two variables after conditioning on all other variables) as edge weights. Suppose $\boldsymbol{y}_1$ is a set containing variables i and j and $\boldsymbol{y}_2$ contains all other variables. We can arbitrarily reorder the varibles such that i and j are the first two:

$$\boldsymbol{y}^\top = \begin{bmatrix} \boldsymbol{y}_1^\top & \boldsymbol{y}_2^\top \end{bmatrix} = \begin{bmatrix} y_i & y_j & \boldsymbol{y}_2^\top \end{bmatrix}$$

The Gaussian graphical model uses partial correlation coefficients (correlation between two variables after conditioning on all other variables) as edge weights. Suppose $\boldsymbol{y}_1$ is a set containing variables i and j and $\boldsymbol{y}_2$ contains all other variables. We can arbitrarily reorder the varibles such that i and j are the first two:

$$\boldsymbol{y}^\top = \begin{bmatrix} \boldsymbol{y}_1^\top & \boldsymbol{y}_2^\top \end{bmatrix} = \begin{bmatrix} y_i & y_j & \boldsymbol{y}_2^\top \end{bmatrix}$$

For notational clarity, let:

$$\begin{aligned} \mathrm{Var}\left(\boldsymbol{y}_1\right) &= \boldsymbol{\Sigma}_1 \\ \mathrm{Var}\left(\boldsymbol{y}_2\right) &= \boldsymbol{\Sigma}_2 \\ \mathrm{Cov}\left(\boldsymbol{y}_1, \boldsymbol{y}_2\right) &= \boldsymbol{\Sigma}_{12} \\ \mathrm{Cov}\left(\boldsymbol{y}_2, \boldsymbol{y}_1\right) &= \boldsymbol{\Sigma}_{21} \end{aligned}$$

The Gaussian graphical model uses partial correlation coefficients (correlation between two variables after conditioning on all other variables) as edge weights. Suppose $\boldsymbol{y}_1$ is a set containing variables i and j and $\boldsymbol{y}_2$ contains all other variables. We can arbitrarily reorder the varibles such that i and j are the first two:

$$\boldsymbol{y}^\top = \begin{bmatrix} \boldsymbol{y}_1^\top & \boldsymbol{y}_2^\top \end{bmatrix} = \begin{bmatrix} y_i & y_j & \boldsymbol{y}_2^\top \end{bmatrix}$$

For notational clarity, let:

$$\begin{aligned} \mathrm{Var}\left(\boldsymbol{y}_1\right) &= \boldsymbol{\Sigma}_1 \\ \mathrm{Var}\left(\boldsymbol{y}_2\right) &= \boldsymbol{\Sigma}_2 \\ \mathrm{Cov}\left(\boldsymbol{y}_1, \boldsymbol{y}_2\right) &= \boldsymbol{\Sigma}_{12} \\ \mathrm{Cov}\left(\boldsymbol{y}_2, \boldsymbol{y}_1\right) &= \boldsymbol{\Sigma}_{21} \end{aligned}$$

This makes $\mathrm{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma}$ a block matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix}$$

From

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix}$$

and the Schur complement, it follows that the conditional covariance of block $\boldsymbol{y}_1$ given $\boldsymbol{y}_2$ takes the form:

$$\mathrm{Var}\left(\boldsymbol{y}_1 \mid \boldsymbol{y}_2\right) = \boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_{21} = \begin{bmatrix} \mathrm{Var}(y_i \mid \boldsymbol{y}_2) & \mathrm{Cov}(y_i, y_j \mid \boldsymbol{y}_2) \\ \mathrm{Cov}(y_j, y_i \mid \boldsymbol{y}_2) & \mathrm{Var}(y_j \mid \boldsymbol{y}_2) \end{bmatrix}$$

From

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix}$$

and the Schur complement, it follows that the conditional covariance of block $\boldsymbol{y}_1$ given $\boldsymbol{y}_2$ takes the form:

$$\mathrm{Var}\left(\boldsymbol{y}_1 \mid \boldsymbol{y}_2\right) = \boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_{21} = \begin{bmatrix} \mathrm{Var}(y_i \mid \boldsymbol{y}_2) & \mathrm{Cov}(y_i, y_j \mid \boldsymbol{y}_2) \\ \mathrm{Cov}(y_j, y_i \mid \boldsymbol{y}_2) & \mathrm{Var}(y_j \mid \boldsymbol{y}_2) \end{bmatrix}$$

The partial correlation can then be obtained as usual:

$$\mathrm{Cor}\left(y_i, y_j \mid \boldsymbol{y}_2\right) = \frac{\mathrm{Cov}(y_i, y_j \mid \boldsymbol{y}_2)}{\sqrt{\mathrm{Var}(y_i \mid \boldsymbol{y}_2)\mathrm{Var}(y_j \mid \boldsymbol{y}_2)}}$$

Let:

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{K} = \begin{bmatrix} \boldsymbol{K}_1 & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_2 \end{bmatrix}$$

Let:

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{K} = \begin{bmatrix} \boldsymbol{K}_1 & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_2 \end{bmatrix}$$

Any block matrix can be inverted as follows:

$$\left(\begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix}\right)^{-1} = \begin{bmatrix} \boldsymbol{K}_1 & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_2 \end{bmatrix} = \begin{bmatrix} \left(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_{21}\right)^{-1} & \text{something} \\ \text{something} & \text{something} \end{bmatrix}$$

https://en.wikipedia.org/wiki/Invertible_matrix#Blockwise_inversion

Let:

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{K} = \begin{bmatrix} \boldsymbol{K}_1 & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_2 \end{bmatrix}$$

Any block matrix can be inverted as follows:

$$\left(\begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix} \right)^{-1} = \begin{bmatrix} \boldsymbol{K}_1 & \boldsymbol{K}_{12} \\ \boldsymbol{K}_{21} & \boldsymbol{K}_2 \end{bmatrix} = \begin{bmatrix} \left(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\Sigma}_{21} \right)^{-1} & \text{something} \\ \text{something} & \text{something} \end{bmatrix}$$

https://en.wikipedia.org/wiki/Invertible_matrix#Blockwise_inversion

Thus it follows that:

$$\boldsymbol{K}_1 = (\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\Sigma}_{21})^{-1} = \boldsymbol{\Sigma}_{1|2}^{-1}$$

Which must be true for all possible orderings of $\boldsymbol{y}$; a block of the inverse of $\boldsymbol{\Sigma}$ equaties the inverse of a conditional variance–covariance matrix $\boldsymbol{\Sigma}_{1|2}$!

A 2×2 matrix, such as $\boldsymbol{\Sigma}_{1|2}$ can readily be inverted:

$$\boldsymbol{K}_1 = \boldsymbol{\Sigma}_{1|2}^{-1} = \frac{1}{\mathrm{Det}(\boldsymbol{\Sigma}_{1|2})} \begin{bmatrix} \mathrm{Var}(y_j \mid \boldsymbol{y}_2) & -\mathrm{Cov}(y_j, y_i \mid \boldsymbol{y}_2) \\ -\mathrm{Cov}(y_i, y_j \mid \boldsymbol{y}_2) & \mathrm{Var}(y_i \mid \boldsymbol{y}_2) \end{bmatrix}$$

A 2×2 matrix, such as $\boldsymbol{\Sigma}_{1|2}$ can readily be inverted:

$$\boldsymbol{K}_1 = \boldsymbol{\Sigma}_{1|2}^{-1} = \frac{1}{\mathrm{Det}(\boldsymbol{\Sigma}_{1|2})} \begin{bmatrix} \mathrm{Var}(y_j \mid \boldsymbol{y}_2) & -\mathrm{Cov}(y_j, y_i \mid \boldsymbol{y}_2) \\ -\mathrm{Cov}(y_i, y_j \mid \boldsymbol{y}_2) & \mathrm{Var}(y_i \mid \boldsymbol{y}_2) \end{bmatrix}$$

This means that element κ_{ij} from $\boldsymbol{K}$ is proportional to the partial correlation:

$$\kappa_{ij} \propto -\mathrm{Cov}\left(y_i, y_j \mid \boldsymbol{y}^{(-(i,j))}\right) \propto -\mathrm{Cor}\left(y_i, y_j \mid \boldsymbol{y}^{(-(i,j))}\right)$$

A 2×2 matrix, such as $\boldsymbol{\Sigma}_{1|2}$ can readily be inverted:

$$\boldsymbol{K}_1 = \boldsymbol{\Sigma}_{1|2}^{-1} = \frac{1}{\mathrm{Det}(\boldsymbol{\Sigma}_{1|2})} \begin{bmatrix} \mathrm{Var}(y_j \mid \boldsymbol{y}_2) & -\mathrm{Cov}(y_j, y_i \mid \boldsymbol{y}_2) \\ -\mathrm{Cov}(y_i, y_j \mid \boldsymbol{y}_2) & \mathrm{Var}(y_i \mid \boldsymbol{y}_2) \end{bmatrix}$$

This means that element κ_{ij} from $\boldsymbol{K}$ is proportional to the partial correlation:

$$\kappa_{ij} \propto -\mathrm{Cov}\left(y_i, y_j \mid \boldsymbol{y}^{(-(i,j))}\right) \propto -\mathrm{Cor}\left(y_i, y_j \mid \boldsymbol{y}^{(-(i,j))}\right)$$

Standardizing $\boldsymbol{K}$ as we would standardize $\boldsymbol{\Sigma}$ cancels this constant out. Multiplying with -1 then gives the partial correlation:

$$\mathrm{Cor}\left(y_i, y_j \mid \boldsymbol{y}^{(-(i,j))}\right) = \omega_{ij} = -\frac{\kappa_{ij}}{\sqrt{\kappa_{ii}\kappa_{jj}}} \quad \text{if } i \neq j$$

A 2×2 matrix, such as $\boldsymbol{\Sigma}_{1|2}$ can readily be inverted:

$$\boldsymbol{K}_1 = \boldsymbol{\Sigma}_{1|2}^{-1} = \frac{1}{\mathrm{Det}(\boldsymbol{\Sigma}_{1|2})} \begin{bmatrix} \mathrm{Var}(y_j \mid \boldsymbol{y}_2) & -\mathrm{Cov}(y_j, y_i \mid \boldsymbol{y}_2) \\ -\mathrm{Cov}(y_i, y_j \mid \boldsymbol{y}_2) & \mathrm{Var}(y_i \mid \boldsymbol{y}_2) \end{bmatrix}$$

This means that element κ_{ij} from $\boldsymbol{K}$ is proportional to the partial correlation:

$$\kappa_{ij} \propto -\mathrm{Cov}\left(y_i, y_j \mid \boldsymbol{y}^{(-(i,j))}\right) \propto -\mathrm{Cor}\left(y_i, y_j \mid \boldsymbol{y}^{(-(i,j))}\right)$$

Standardizing $\boldsymbol{K}$ as we would standardize $\boldsymbol{\Sigma}$ cancels this constant out. Multiplying with -1 then gives the partial correlation:

$$\mathrm{Cor}\left(y_i, y_j \mid \boldsymbol{y}^{(-(i,j))}\right) = \omega_{ij} = -\frac{\kappa_{ij}}{\sqrt{\kappa_{ii}\kappa_{jj}}} \quad \text{if } i \neq j$$

Which we may capture in a matrix $\boldsymbol{\Omega}$ with zeroes on the diagonal. Using a diagonal scaling matrix $\boldsymbol{\Delta}$ with $1/\sqrt{\kappa_{kk}}$ on its kth diagonal element, we can write this in a matrix expression:

$$\boldsymbol{\Omega} = \boldsymbol{I} - \boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}$$

We can now rearange this expression to form a psychometric model for $\boldsymbol{\Sigma} = \boldsymbol{K}^{-1}$ as a function of the partial correlations in $\boldsymbol{\Omega}$:

$$
\begin{aligned}
\boldsymbol{I} - \boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta} &= \Omega \\
\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta} &= \boldsymbol{I} - \Omega \\
\boldsymbol{K} &= \boldsymbol{\Delta}^{-1}\left(\boldsymbol{I} - \Omega\right)\boldsymbol{\Delta}^{-1}
\end{aligned}
$$

We can now rearange this expression to form a psychometric model for $\boldsymbol{\Sigma} = \boldsymbol{K}^{-1}$ as a function of the partial correlations in $\boldsymbol{\Omega}$:

$$\begin{aligned}\boldsymbol{I} - \boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta} &= \boldsymbol{\Omega} \\ \boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta} &= \boldsymbol{I} - \boldsymbol{\Omega} \\ \boldsymbol{K} &= \boldsymbol{\Delta}^{-1} \left(\boldsymbol{I} - \boldsymbol{\Omega}\right) \boldsymbol{\Delta}^{-1}\end{aligned}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Delta} \left(\boldsymbol{I} - \boldsymbol{\Omega}\right)^{-1} \boldsymbol{\Delta}$$

This expression can be used to fit a GGM in exactly the same way you would fit a SEM! It only requires a different matrix expression to be used.

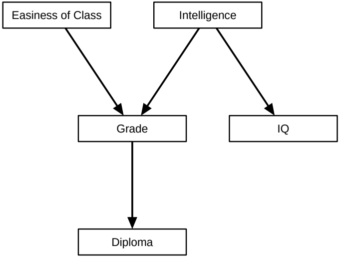
A Gaussian Graphical Model, like a SEM model, implies a variance–covariance structure:

$$\boldsymbol{\Sigma} = \boldsymbol{\Delta} \left(\boldsymbol{I} - \boldsymbol{\Omega} \right)^{-1} \boldsymbol{\Delta}$$

And can be fitted like a SEM. In this case, $DF = 1$ and the model matrices are:

$$\boldsymbol{\Omega} = \begin{bmatrix} 0 & & \\ \omega_{21} & 0 & \\ 0 & \omega_{32} & 0 \end{bmatrix}, \boldsymbol{\Delta} = \begin{bmatrix} \delta_{11} & & \\ 0 & \delta_{22} & \\ 0 & 0 & \delta_{33} \end{bmatrix}$$

Causal model



Markov Random Field