Structure Estimation for Mixed Graphical Models

Jonas Haslbeck Lourens Waldorp

Psychological Methods Group, University of Amsterdam

February 18, 2016

Overview

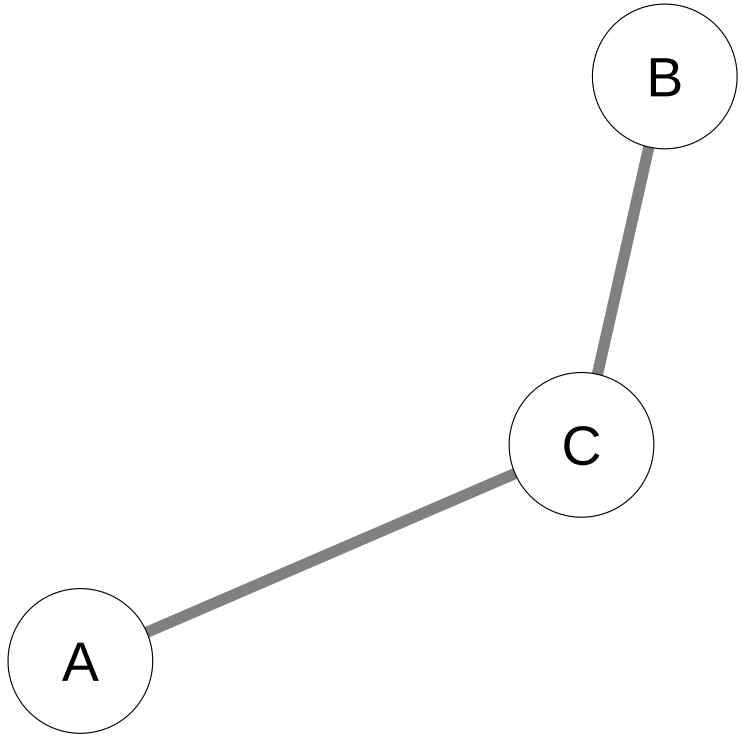
Undirected Graphical Models (or Markov Random Fields)

$$X_A \perp\!\!\!\perp X_B | X_C$$

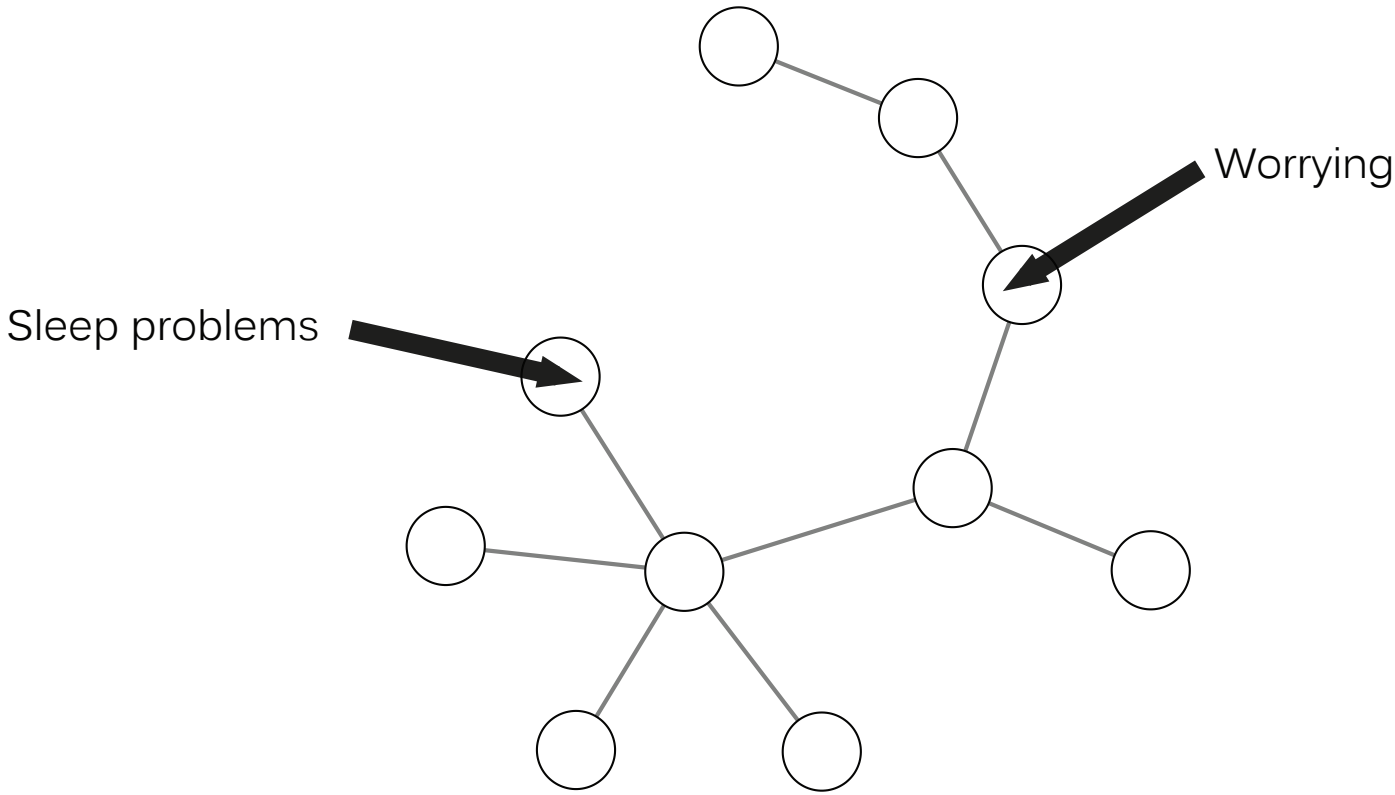
$$X_A \not\perp\!\!\!\perp X_C | X_B$$

$$X_C \not\perp\!\!\!\perp X_B | X_A$$

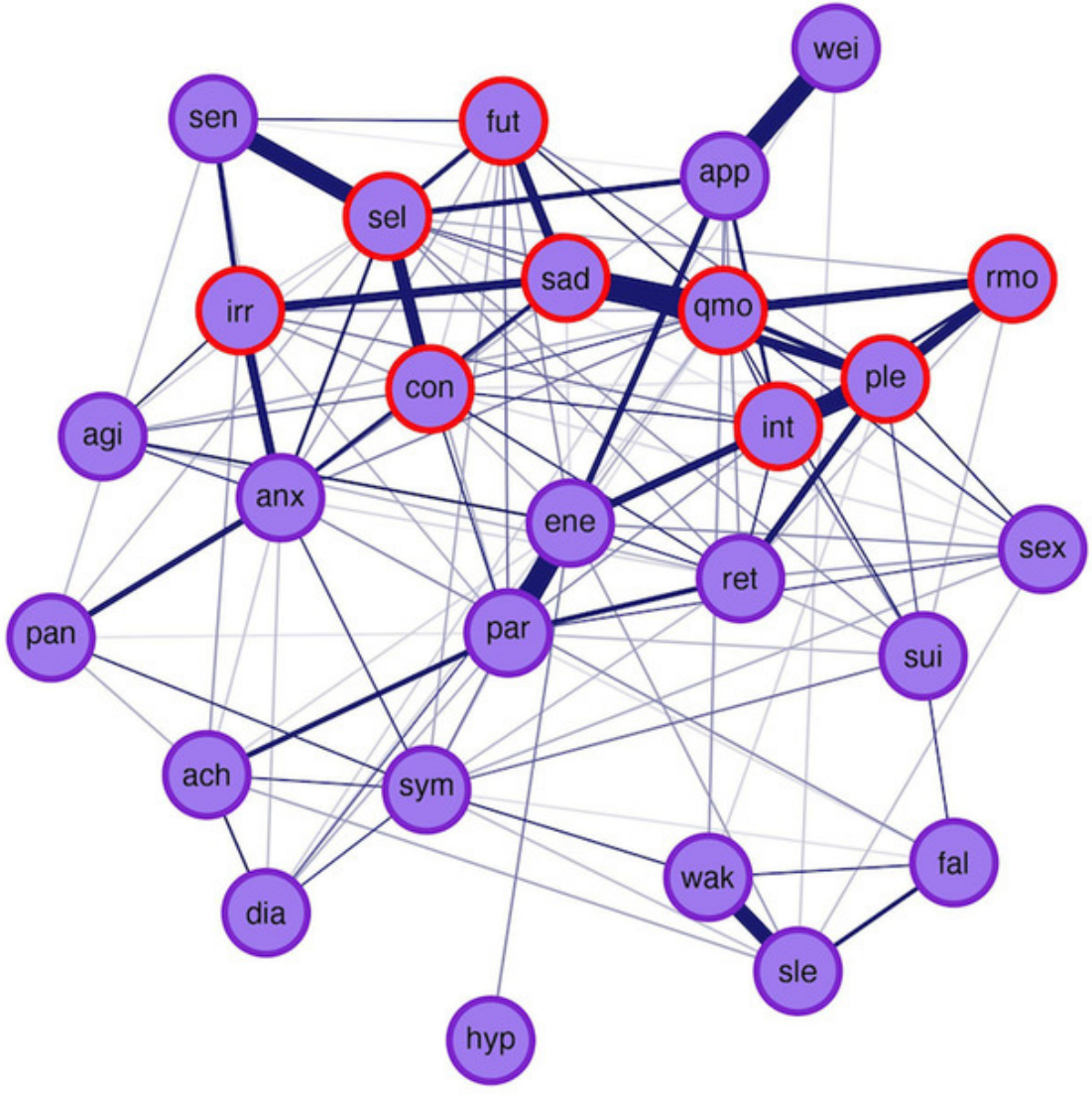
$\Longleftrightarrow$



Why Graphical Models?

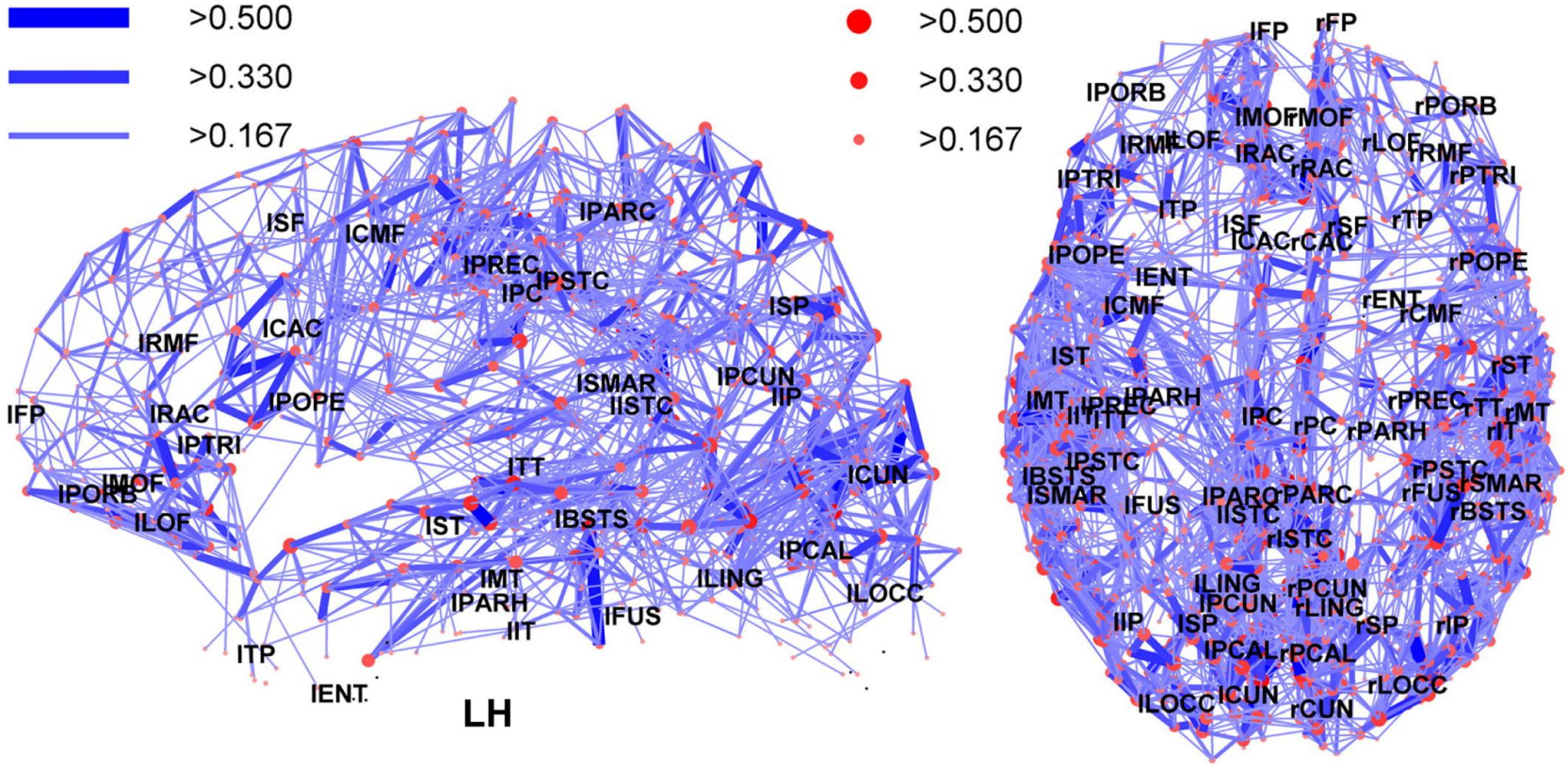


Why Graphical Models?

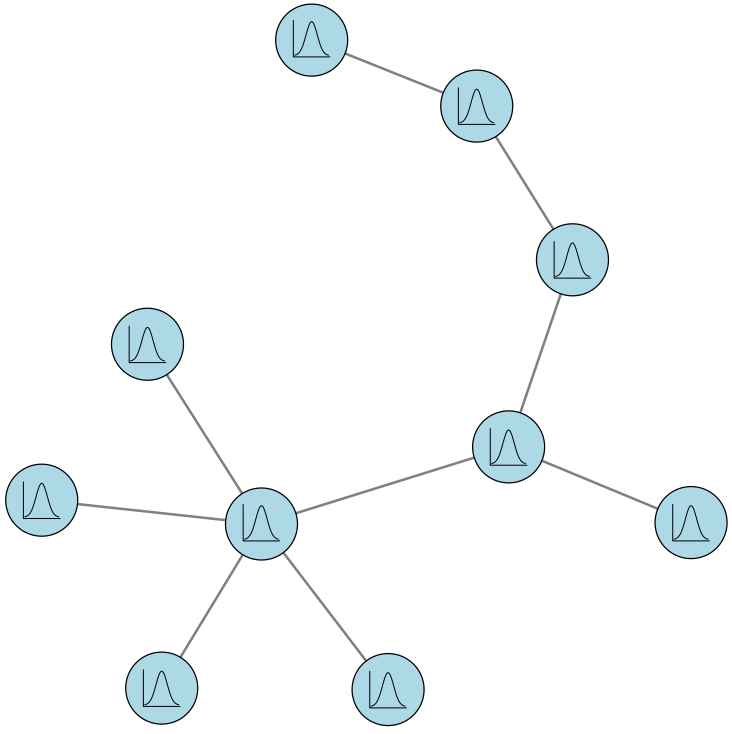


ach: aches and pains
agi: psychomotor agitation
anx: feeling anxious
app: change of appetite
con: concentration problems
dia: diarrhea/constipation
ene: energy level
fal: falling asleep
fut: view of myself
hyp: hypersomnia
int: general interest
irr: feeling irritable
pan: panic/phobic symptoms
par: leaden paralysis
ple: capacity for pleasure (not sex)
qmo: quality of mood
ret: psychomotor retardation
rmo: respons of mood
sad: feeling sad
sel: view of oneself
sen: interpersonal sensitivity
sex: interest in sex
sle: sleep during the night
sui: suicidal thoughts
sym: other bodily symptoms
wak: waking up too early
wei: change of weight

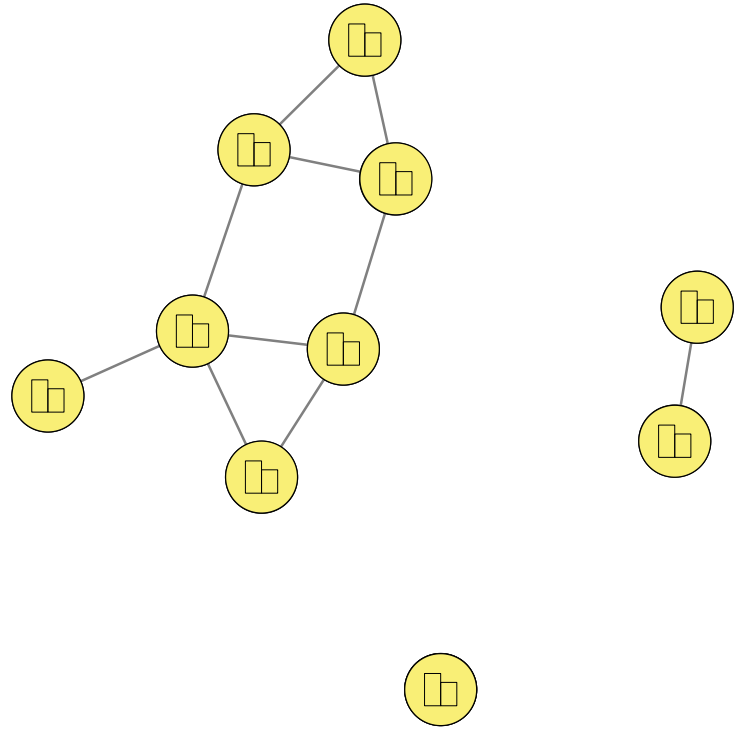
Why Graphical Models?



Existing structure estimation methods

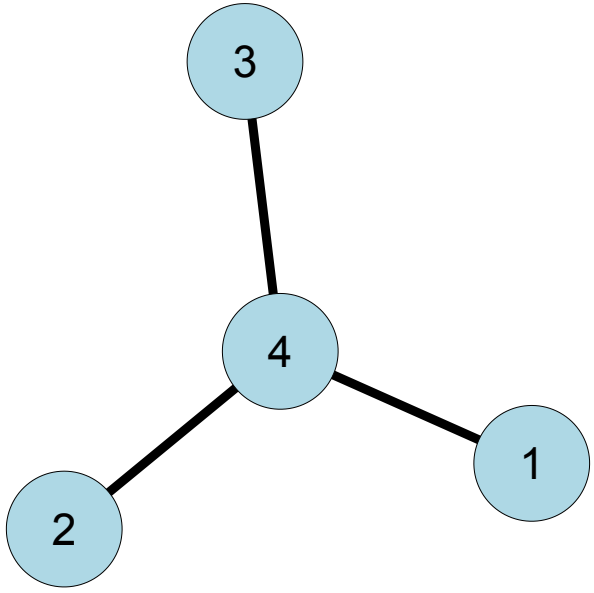


Gaussian



Binary (Ising model)

Structure estimation in the Gaussian case

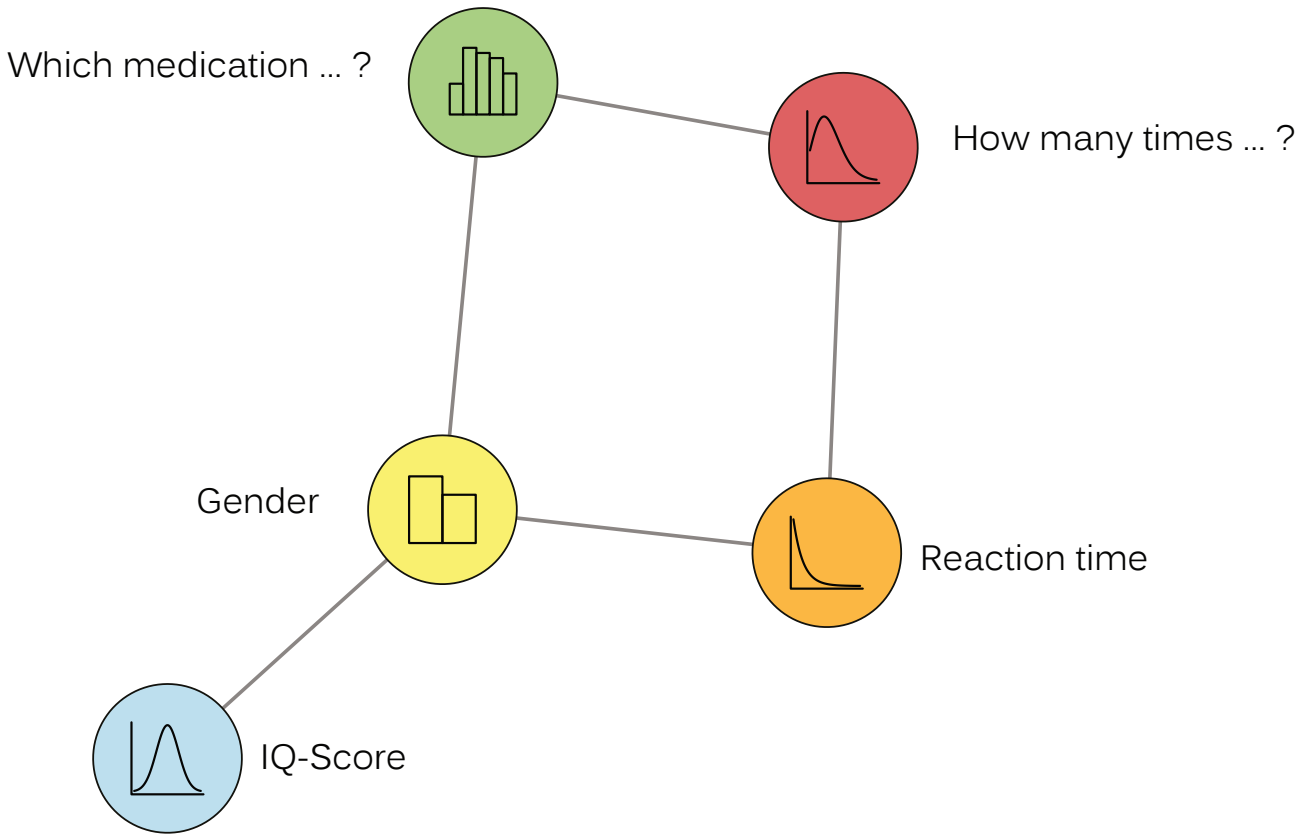


$$\Longleftrightarrow \quad \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{pmatrix} 3.45 & 0 & 0 & 3.18 \\ 0 & 2.14 & 0 & 0.82 \\ 0 & 0 & 3.21 & 1.05 \\ 3.18 & 0.82 & 1.05 & 8.77 \end{pmatrix} \end{array}$$

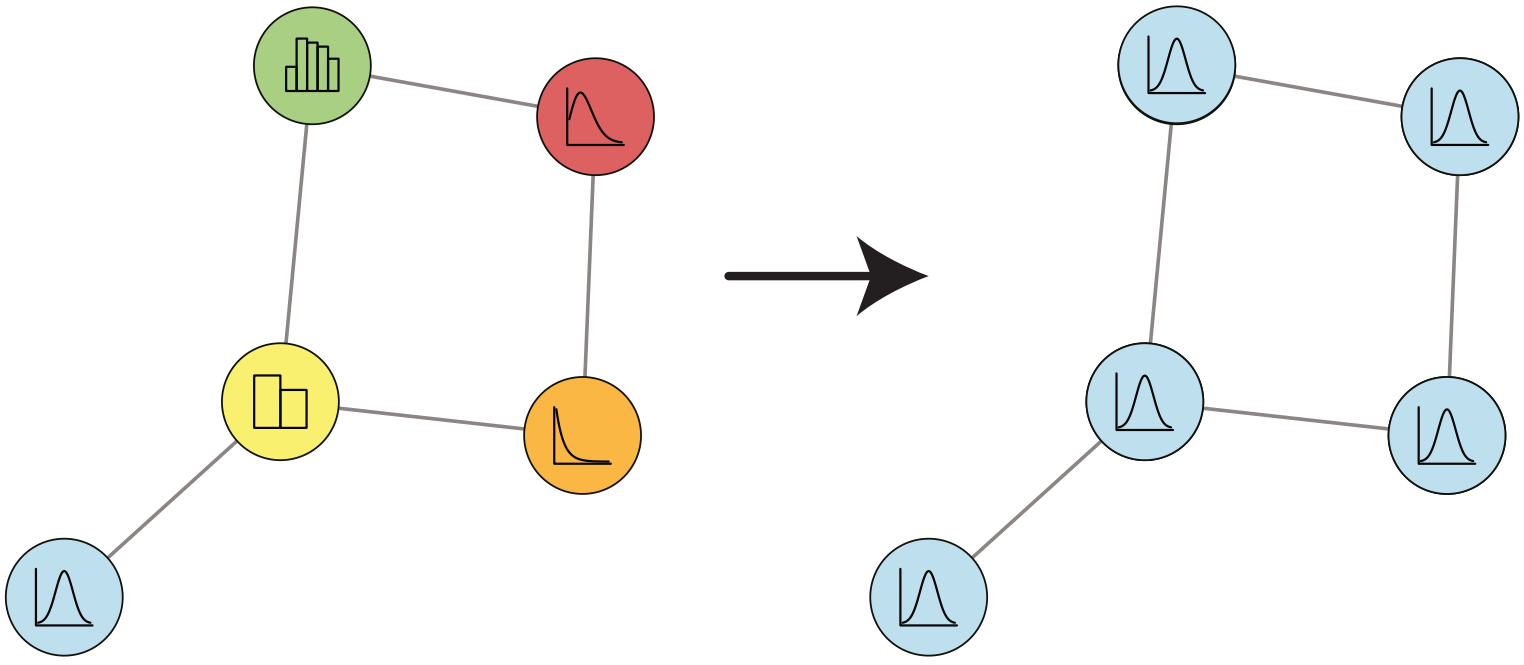
Estimation:

- glasso (Friedman et al., 2008)
- Nodewise methods (Meinshausen & Bühlmann, 2006)

Mixed Graphical Models



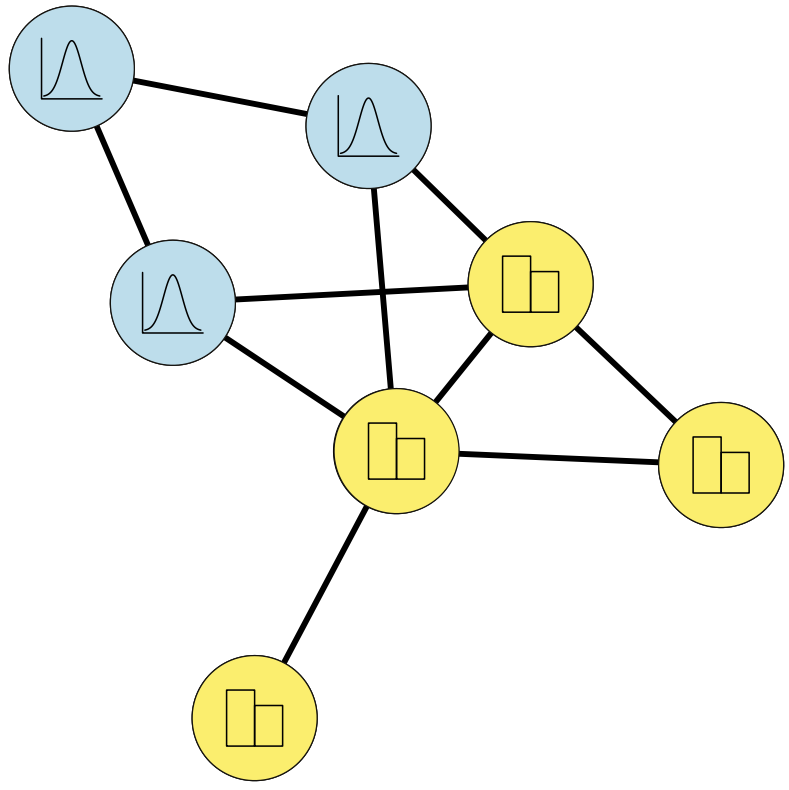
Gaussianizing variables



Two approaches:

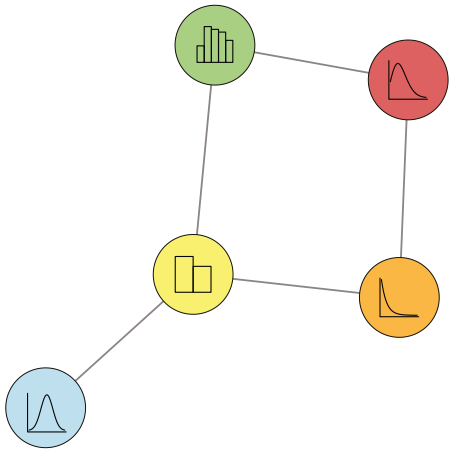
- Copula-based (Dobra and Lenosti, 2001; Liu et al. 2012)
- Non-paranormal (Liu et al., 2009; Lafftery et al. 2012)

Conditional Gaussian



Multivariate Gaussian conditioned on $2^{|\textit{Binary nodes}|}$ configurations.

How to estimate a Mixed Graphical Model?



Probability distribution over mixed data

+

$$\Gamma = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{pmatrix} 4.04 & 1.05 & 0 & 3.18 & 0 \\ 1.05 & 1.44 & 2.75 & 0.82 & 0.79 \\ 0 & 2.75 & 5.52 & 1.05 & 4.22 \\ 3.18 & 0.82 & 1.05 & 4.50 & 3.88 \\ 0 & 0.79 & 4.22 & 3.88 & 3.18 \end{pmatrix} \end{array}$$

Feasible estimation procedure

Mixed exponential Markov random fields

Conditional univariate members of the exponential family

$$P(X_s|X_{\setminus s}) = \exp\left\{E_s(X_{\setminus s})\phi_s(X_s) + C_s(X_s) - \Phi(X_{\setminus s})\right\},$$

factorize to a global multivariate distribution which factors according the graph defined by the node-neighborhoods if and only if $E_s(X_{\setminus s})$ has the form:

$$\theta_s + \sum_{t \in N(s)} \theta_{st}\phi_t(X_t) + \ldots + \sum_{t_2,\ldots,t_k \in N(s)} \theta_{t_2,\ldots,t_k} \prod_{j=2}^{k} \phi_{t_j}(X_{t_j}),$$

(Yang et al., 2014)

Mixed exponential Markov random fields

Conditional univariate members of the exponential family

$$P(X_s|X_{\setminus s}) = \exp\left\{E_s(X_{\setminus s})\phi_s(X_s) + C_s(X_s) - \Phi(X_{\setminus s})\right\},$$

factorize to a global multivariate distribution which factors according the graph defined by the node-neighborhoods if and only if $E_s(X_{\setminus s})$ has the form:

$$\theta_s + \sum_{t \in N(s)} \theta_{st}\phi_t(X_t) + \ldots + \sum_{t_2,\ldots,t_k \in N(s)} \theta_{t_2,\ldots,t_k} \prod_{j=2}^{k} \phi_{t_j}(X_{t_j}),$$

(Yang et al., 2014)

Mixed exponential Markov random fields

Conditional univariate members of the exponential family

$$P(X_s|X_{\setminus s}) = \exp\left\{E_s(X_{\setminus s})\phi_s(X_s) + C_s(X_s) - \Phi(X_{\setminus s})\right\},$$

factorize to a global multivariate distribution which factors according the graph defined by the node-neighborhoods if and only if $E_s(X_{\setminus s})$ has the form:

$$\theta_s + \sum_{t \in N(s)} \theta_{st}\phi_t(X_t) + \ldots + \sum_{t_2,\ldots,t_k \in N(s)} \theta_{t_2,\ldots,t_k} \prod_{j=2}^{k} \phi_{t_j}(X_{t_j}),$$

(Yang et al., 2014)

Mixed exponential Markov random fields

Conditional univariate members of the exponential family

$$P(X_s|X_{\setminus s}) = \exp\left\{E_s(X_{\setminus s})\phi_s(X_s) + C_s(X_s) - \Phi(X_{\setminus s})\right\},$$

factorize to a global multivariate distribution which factors according the graph defined by the node-neighborhoods if and only if $E_s(X_{\setminus s})$ has the form:

$$\theta_s + \sum_{t \in N(s)} \theta_{st}\phi_t(X_t) + \ldots + \sum_{t_2,\ldots,t_k \in N(s)} \theta_{t_2,\ldots,t_k} \prod_{j=2}^{k} \phi_{t_j}(X_{t_j}),$$

(Yang et al., 2014)

Mixed exponential Markov random fields

Conditional univariate members of the exponential family

$$P(X_s|X_{\setminus s}) = \exp\left\{E_s(X_{\setminus s})\phi_s(X_s) + C_s(X_s) - \Phi(X_{\setminus s})\right\},$$

factorize to a global multivariate distribution which factors according the graph defined by the node-neighborhoods if and only if $E_s(X_{\setminus s})$ has the form:

$$\theta_s + \sum_{t \in N(s)} \theta_{st}\phi_t(X_t) + \ldots + \sum_{t_2,\ldots,t_k \in N(s)} \theta_{t_2,\ldots,t_k} \prod_{j=2}^{k} \phi_{t_j}(X_{t_j}),$$

(Yang et al., 2014)

Mixed exponential Markov random fields

Conditional univariate members of the exponential family

$$P(X_s|X_{\setminus s}) = \exp\left\{E_s(X_{\setminus s})\phi_s(X_s) + C_s(X_s) - \Phi(X_{\setminus s})\right\},$$

factorize to a global multivariate distribution which factors according the graph defined by the node-neighborhoods if and only if $E_s(X_{\setminus s})$ has the form:

$$\theta_s + \sum_{t \in N(s)} \theta_{st}\phi_t(X_t) + \ldots + \sum_{t_2,\ldots,t_k \in N(s)} \theta_{t_2,\ldots,t_k} \prod_{j=2}^{k} \phi_{t_j}(X_{t_j}),$$

(Yang et al., 2014)

Mixed exponential Markov random fields

Conditional univariate members of the exponential family

$$P(X_s|X_{\setminus s}) = \exp\left\{E_s(X_{\setminus s})\phi_s(X_s) + C_s(X_s) - \Phi(X_{\setminus s})\right\},$$

factorize to a global multivariate distribution which factors according the graph defined by the node-neighborhoods if and only if $E_s(X_{\setminus s})$ has the form:

$$\theta_s + \sum_{t \in N(s)} \theta_{st}\phi_t(X_t) + \ldots + \sum_{t_2,\ldots,t_k \in N(s)} \theta_{t_2,\ldots,t_k} \prod_{j=2}^{k} \phi_{t_j}(X_{t_j}),$$

(Yang et al., 2014)

Mixed exponential Markov random fields

Conditional univariate members of the exponential family

$$P(X_s|X_{\setminus s}) = \exp\left\{E_s(X_{\setminus s})\phi_s(X_s) + C_s(X_s) - \Phi(X_{\setminus s})\right\},$$

factorize to a global multivariate distribution which factors according the graph defined by the node-neighborhoods if and only if $E_s(X_{\setminus s})$ has the form:

$$\theta_s + \sum_{t \in N(s)} \theta_{st}\phi_t(X_t) + \ldots + \sum_{t_2,\ldots,t_k \in N(s)} \theta_{t_2,\ldots,t_k} \prod_{j=2}^{k} \phi_{t_j}(X_{t_j}),$$

(Yang et al., 2014)

Mixed MRF: Joint Distribution

The corresponding joint distribution has the form

$$P(X;\theta) = \exp\{\sum_{s\in V}\theta_s\phi_s(X_s) + \sum_{s\in V}\sum_{t\in N(s)}\theta_{st}\phi_s(X_s)\phi_t(X_t) +$$

$$\cdots + \sum_{t_1,\ldots,t_k\in\mathcal{C}}\theta_{t_1,\ldots,t_k}\prod_{j=1}^{k}\phi_{t_j}(X_{t_j}) + \sum_{s\in V}C_s(X_s) - \Phi(\theta)\}.$$

Mixed MRF: Joint Distribution

The corresponding joint distribution has the form

$$P(X;\theta) = \exp\{\sum_{s\in V} \theta_s \phi_s(X_s) + \sum_{s\in V} \sum_{t\in N(s)} \theta_{st}\phi_s(X_s)\phi_t(X_t) +$$

$$\cdots + \sum_{t_1,\ldots,t_k\in\mathcal{C}} \theta_{t_1,\ldots,t_k} \prod_{j=1}^{k} \phi_{t_j}(X_{t_j}) + \sum_{s\in V} C_s(X_s) - \Phi(\theta)\}.$$

Mixed MRF: Joint Distribution

The corresponding joint distribution has the form

$$P(X;\theta) = \exp\{\sum_{s\in V} \theta_s \phi_s(X_s) + \sum_{s\in V} \sum_{t\in N(s)} \theta_{st}\phi_s(X_s)\phi_t(X_t) +$$

$$\cdots + \sum_{t_1,\ldots,t_k\in\mathcal{C}} \theta_{t_1,\ldots,t_k} \prod_{j=1}^{k} \phi_{t_j}(X_{t_j}) + \sum_{s\in V} C_s(X_s) - \Phi(\theta)\}.$$

Mixed MRF: Joint Distribution

The corresponding joint distribution has the form

$$P(X;\theta) = \exp\{\sum_{s\in V}\theta_s\phi_s(X_s) + \sum_{s\in V}\sum_{t\in N(s)}\theta_{st}\phi_s(X_s)\phi_t(X_t) +$$

$$\cdots + \sum_{t_1,\ldots,t_k\in\mathcal{C}}\theta_{t_1,\ldots,t_k}\prod_{j=1}^{k}\phi_{t_j}(X_{t_j}) + \sum_{s\in V}C_s(X_s) - \Phi(\theta)\}.$$

Mixed MRF: Ising-Gaussian

$$P(Y,Z) \propto \exp\Big\{ \sum_{s\in V_Y} \frac{\theta_s^y}{\sigma_s} Y_s + \sum_{r\in V_Z} \theta_r^z Z_r + \sum_{(s,t)\in E_Y} \frac{\theta_{st}^{yy}}{\sigma_s\sigma_t} Y_s Y_t + \sum_{(r,q)\in E_Z} \theta_{rq}^{zz} Z_r Z_q + \sum_{(s,r)\in E_{YZ}} \frac{\theta_{sr}^{yz}}{\sigma_s} Y_s Z_r - \sum_{s\in V_Y} \frac{Y_s^2}{2\sigma_s^2} \Big\}$$

If X_s Bernoulli, the node-conditional has the form:

$$P(X_s|X_{\setminus s}) \propto \exp\Big\{\theta_r^z Z_r + \sum_{q\in N(r)_Z} \theta_{rq}^{zz} Z_r Z_q + \sum_{t\in N(r)_Y} \frac{\theta_{rt}^{yz}}{\sigma_t} Z_r Y_t \Big\}$$

If X_s Gaussian, the node-conditional has the form:

$$P(X_s|X_{\setminus s}) \propto \exp\Big\{\frac{\theta_s^y}{\sigma_s} Y_s + \sum_{t\in N(s)_Y} \frac{\theta_{st}^{yy}}{\sigma_s\sigma_t} Y_s Y_t + \sum_{r\in N(s)_Z} \frac{\theta_{sr}^{yz}}{\sigma_s} Y_s Z_r - \frac{Y_s^2}{2\sigma_s^2} \Big\}$$

Mixed MRF: Ising-Gaussian

$$P(Y,Z) \propto \exp\Big\{ \sum_{s\in V_Y} \frac{\theta_s^y}{\sigma_s} Y_s + \sum_{r\in V_Z} \theta_r^z Z_r + \sum_{(s,t)\in E_Y} \frac{\theta_{st}^{yy}}{\sigma_s\sigma_t} Y_s Y_t + \sum_{(r,q)\in E_Z} \theta_{rq}^{zz} Z_r Z_q + \sum_{(s,r)\in E_{YZ}} \frac{\theta_{sr}^{yz}}{\sigma_s} Y_s Z_r - \sum_{s\in V_Y} \frac{Y_s^2}{2\sigma_s^2} \Big\}$$

If X_s Bernoulli, the node-conditional has the form:

$$P(X_s|X_{\setminus s}) \propto \exp\Big\{\theta_r^z Z_r + \sum_{q\in N(r)_Z} \theta_{rq}^{zz} Z_r Z_q + \sum_{t\in N(r)_Y} \frac{\theta_{rt}^{yz}}{\sigma_t} Z_r Y_t \Big\}$$

If X_s Gaussian, the node-conditional has the form:

$$P(X_s|X_{\setminus s}) \propto \exp\Big\{\frac{\theta_s^y}{\sigma_s} Y_s + \sum_{t\in N(s)_Y} \frac{\theta_{st}^{yy}}{\sigma_s\sigma_t} Y_s Y_t + \sum_{r\in N(s)_Z} \frac{\theta_{sr}^{yz}}{\sigma_s} Y_s Z_r - \frac{Y_s^2}{2\sigma_s^2} \Big\}$$

Mixed MRF: Ising-Gaussian

$$P(Y,Z) \propto \exp\Big\{ \sum_{s\in V_Y} \frac{\theta_s^y}{\sigma_s} Y_s + \sum_{r\in V_Z} \theta_r^z Z_r + \sum_{(s,t)\in E_Y} \frac{\theta_{st}^{yy}}{\sigma_s\sigma_t} Y_s Y_t + \sum_{(r,q)\in E_Z} \theta_{rq}^{zz} Z_r Z_q + \sum_{(s,r)\in E_{YZ}} \frac{\theta_{sr}^{yz}}{\sigma_s} Y_s Z_r - \sum_{s\in V_Y} \frac{Y_s^2}{2\sigma_s^2} \Big\}$$

If X_s Bernoulli, the node-conditional has the form:

$$P(X_s|X_{\setminus s}) \propto \exp\Big\{\theta_r^z Z_r + \sum_{q\in N(r)_Z} \theta_{rq}^{zz} Z_r Z_q + \sum_{t\in N(r)_Y} \frac{\theta_{rt}^{yz}}{\sigma_t} Z_r Y_t \Big\}$$

If X_s Gaussian, the node-conditional has the form:

$$P(X_s|X_{\setminus s}) \propto \exp\Big\{\frac{\theta_s^y}{\sigma_s} Y_s + \sum_{t\in N(s)_Y} \frac{\theta_{st}^{yy}}{\sigma_s\sigma_t} Y_s Y_t + \sum_{r\in N(s)_Z} \frac{\theta_{sr}^{yz}}{\sigma_s} Y_s Z_r - \frac{Y_s^2}{2\sigma_s^2} \Big\}$$

Mixed MRF: Ising-Gaussian

$$P(Y,Z) \propto \exp\Big\{ \sum_{s \in V_Y} \frac{\theta_s^y}{\sigma_s} Y_s + \sum_{r \in V_Z} \theta_r^z Z_r + \sum_{(s,t) \in E_Y} \frac{\theta_{st}^{yy}}{\sigma_s \sigma_t} Y_s Y_t + \sum_{(r,q) \in E_Z} \theta_{rq}^{zz} Z_r Z_q + \sum_{(s,r) \in E_{YZ}} \frac{\theta_{sr}^{yz}}{\sigma_s} Y_s Z_r - \sum_{s \in V_Y} \frac{Y_s^2}{2\sigma_s^2} \Big\}$$

If X_s Bernoulli, the node-conditional has the form:

$$P(X_s | X_{\setminus s}) \propto \exp\Big\{ \theta_r^z Z_r + \sum_{q \in N(r)_Z} \theta_{rq}^{zz} Z_r Z_q + \sum_{t \in N(r)_Y} \frac{\theta_{rt}^{yz}}{\sigma_t} Z_r Y_t \Big\}$$

If X_s Gaussian, the node-conditional has the form:

$$P(X_s | X_{\setminus s}) \propto \exp\Big\{ \frac{\theta_s^y}{\sigma_s} Y_s + \sum_{t \in N(s)_Y} \frac{\theta_{st}^{yy}}{\sigma_s \sigma_t} Y_s Y_t + \sum_{r \in N(s)_Z} \frac{\theta_{sr}^{yz}}{\sigma_s} Y_s Z_r - \frac{Y_s^2}{2\sigma_s^2} \Big\}$$

Mixed MRF: Ising-Gaussian

$$P(Y,Z) \propto \exp\Big\{ \sum_{s \in V_Y} \frac{\theta_s^y}{\sigma_s} Y_s + \sum_{r \in V_Z} \theta_r^z Z_r + \sum_{(s,t) \in E_Y} \frac{\theta_{st}^{yy}}{\sigma_s \sigma_t} Y_s Y_t + \sum_{(r,q) \in E_Z} \theta_{rq}^{zz} Z_r Z_q + \sum_{(s,r) \in E_{YZ}} \frac{\theta_{sr}^{yz}}{\sigma_s} Y_s Z_r - \sum_{s \in V_Y} \frac{Y_s^2}{2\sigma_s^2} \Big\}$$

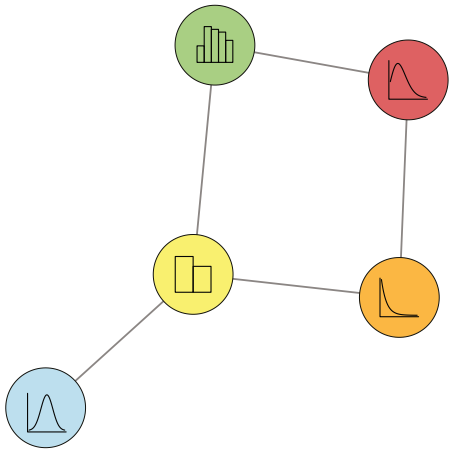
If X_s Bernoulli, the node-conditional has the form:

$$P(X_s | X_{\setminus s}) \propto \exp\Big\{ \theta_r^z Z_r + \sum_{q \in N(r)_Z} \theta_{rq}^{zz} Z_r Z_q + \sum_{t \in N(r)_Y} \frac{\theta_{rt}^{yz}}{\sigma_t} Z_r Y_t \Big\}$$

If X_s Gaussian, the node-conditional has the form:

$$P(X_s | X_{\setminus s}) \propto \exp\Big\{ \frac{\theta_s^y}{\sigma_s} Y_s + \sum_{t \in N(s)_Y} \frac{\theta_{st}^{yy}}{\sigma_s \sigma_t} Y_s Y_t + \sum_{r \in N(s)_Z} \frac{\theta_{sr}^{yz}}{\sigma_s} Y_s Z_r - \frac{Y_s^2}{2\sigma_s^2} \Big\}$$

How to estimate a Mixed Graphical Model?



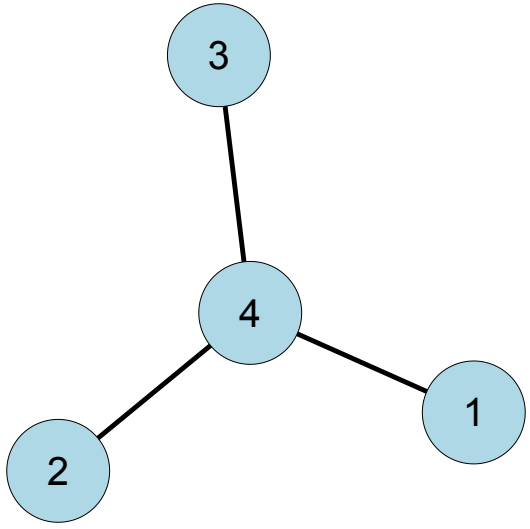
+

$$\Gamma = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{pmatrix} 4.04 & 1.05 & 0 & 3.18 & 0 \\ 1.05 & 1.44 & 2.75 & 0.82 & 0.79 \\ 0 & 2.75 & 5.52 & 1.05 & 4.22 \\ 3.18 & 0.82 & 1.05 & 4.50 & 3.88 \\ 0 & 0.79 & 4.22 & 3.88 & 3.18 \end{pmatrix} \end{array}$$

Probability distribution over mixed data

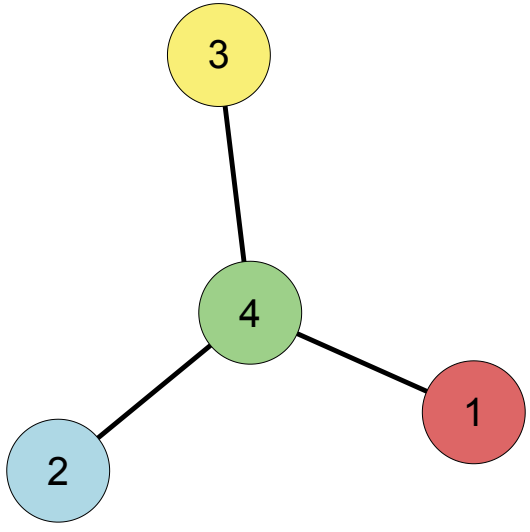
Feasible estimation procedure

Inverse covariance matrices and graph structure



$\Longleftrightarrow$

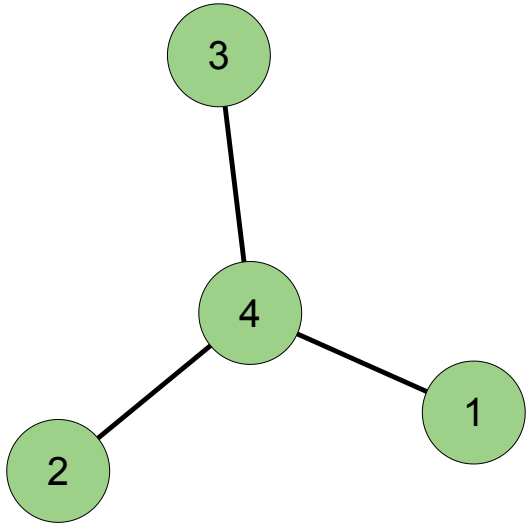
$$\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{pmatrix} 3.45 & 0 & 0 & 3.18 \\ 0 & 2.14 & 0 & 0.82 \\ 0 & 0 & 3.21 & 1.05 \\ 3.18 & 0.82 & 1.05 & 8.77 \end{pmatrix} \end{array}$$



$\Longleftrightarrow\!\!\!\!\!\!/\;\;$

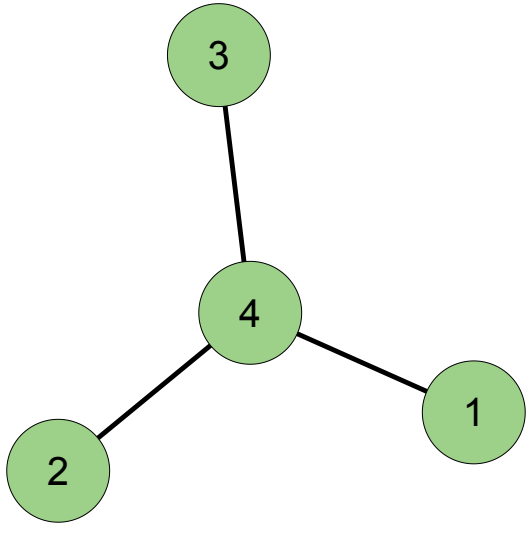
$$\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{pmatrix} 3.45 & 5.12 & 0 & 3.18 \\ 5.12 & 2.14 & 0.51 & 0.82 \\ 0 & 0.51 & 3.21 & 1.05 \\ 3.18 & 0.82 & 1.05 & 8.77 \end{pmatrix} \end{array}$$

Generalized covariance matrices



$\not\Longleftrightarrow$

$$\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{pmatrix} 3.45 & 5.12 & 0 & 3.18 \\ 5.12 & 2.14 & 0.51 & 0.82 \\ 0 & 0.51 & 3.21 & 1.05 \\ 3.18 & 0.82 & 1.05 & 8.77 \end{pmatrix} \end{array}$$

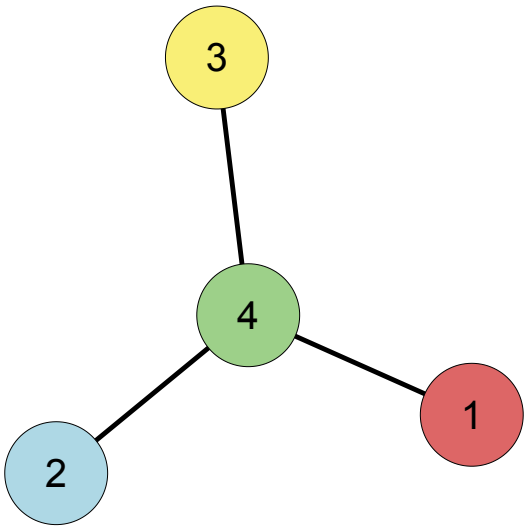


$\Longleftrightarrow$

$$\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_1X_2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_1X_2 & \dots \end{array} \\ \begin{pmatrix} 3.45 & 0 & 0 & 3.18 & 4.98 & \dots \\ 0 & 2.14 & 0 & 0.82 & 1.15 & \dots \\ 0 & 0 & 3.21 & 1.05 & 4.48 & \dots \\ 3.18 & 0.82 & 1.05 & 8.77 & 4.37 & \dots \\ 4.98 & 1.15 & 4.48 & 4.37 & 8.45 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

(Loh and Wainwright, 2013)

Generalized covariance matrices for mixed MRFs

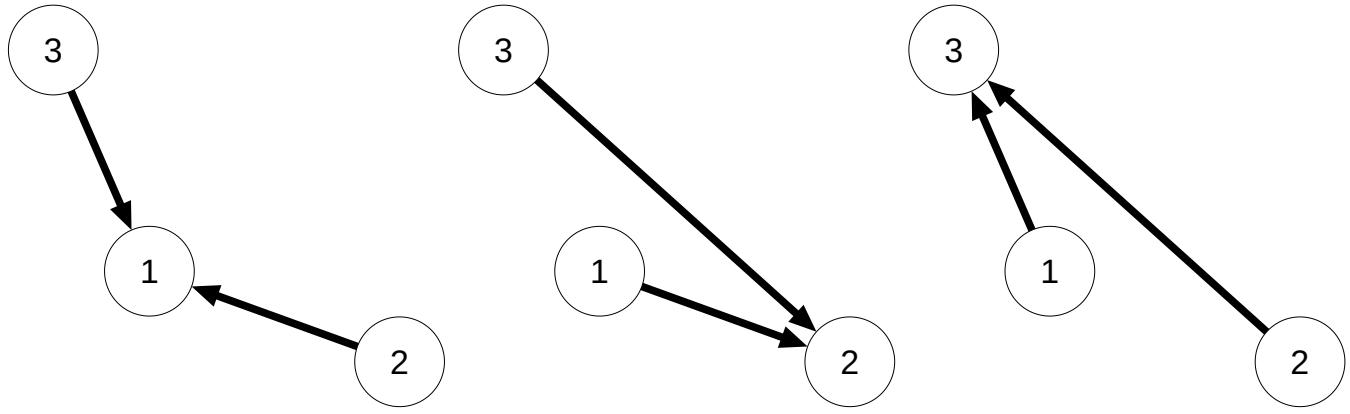


$\Longleftrightarrow$

$$
\begin{array}{c}
 \\
X_1 \\
X_2 \\
X_3 \\
X_4 \\
X_1X_2 \\
\vdots
\end{array}
\begin{array}{c}
\begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_1X_2 & \dots \end{array} \\
\begin{pmatrix}
3.45 & 0 & 0 & 3.18 & 4.98 & \dots \\
0 & 2.14 & 0 & 0.82 & 1.15 & \dots \\
0 & 0 & 3.21 & 1.05 & 4.48 & \dots \\
3.18 & 0.82 & 1.05 & 8.77 & 4.37 & \dots \\
4.98 & 1.15 & 4.48 & 4.37 & 8.45 & \dots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\end{array}
$$

Nodewise estimation algorithm

1. Regress all nodes $V_{\setminus s}$ on node V_s with a ℓ_1-penalty

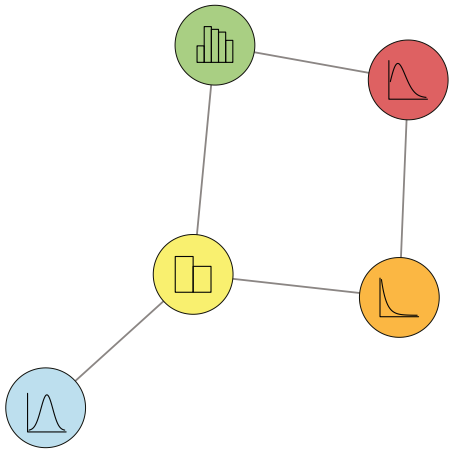


2. Threshold parameters at $\tau_n = \sqrt{d}||\widehat{\beta}||_2 \sqrt{\frac{\log p}{n}}$

3. Combine parameter estimates

$$\widehat{\beta} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \end{array} \begin{array}{c} \begin{array}{ccc} X_1 & X_2 & X_3 \end{array} \\ \begin{pmatrix} NA & 0 & 4.78 \\ 0 & NA & 0.12 \\ 5.11 & 0 & NA \end{pmatrix} \end{array}$$

How to estimate a Mixed Graphical Model?



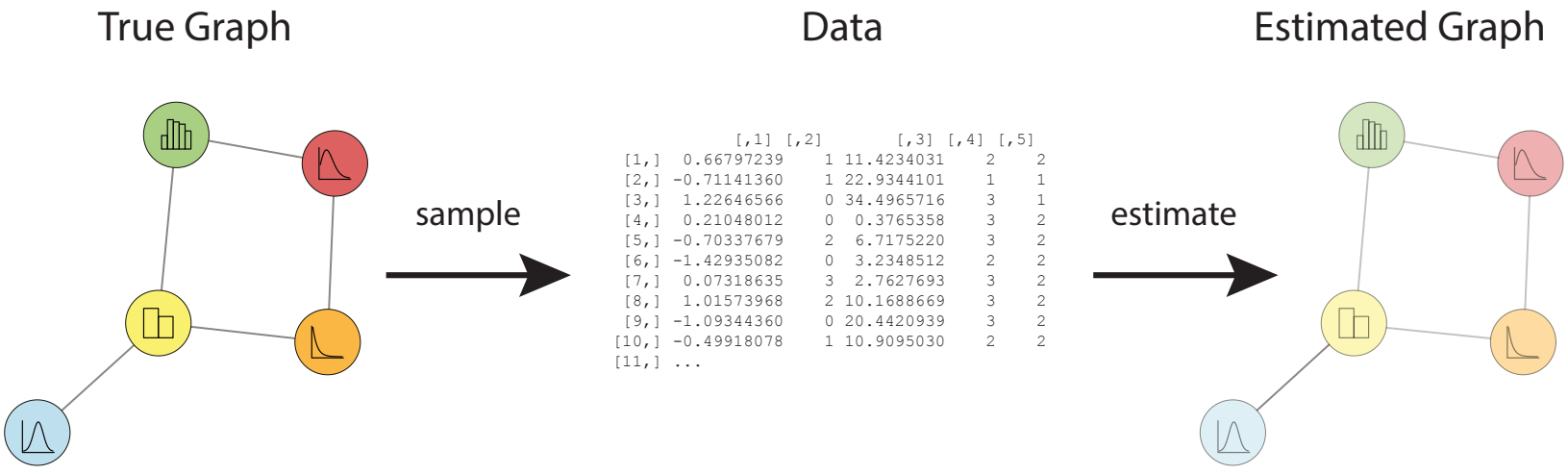
+

$$\Gamma = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{pmatrix} 4.04 & 1.05 & 0 & 3.18 & 0 \\ 1.05 & 1.44 & 2.75 & 0.82 & 0.79 \\ 0 & 2.75 & 5.52 & 1.05 & 4.22 \\ 3.18 & 0.82 & 1.05 & 4.50 & 3.88 \\ 0 & 0.79 & 4.22 & 3.88 & 3.18 \end{pmatrix} \end{array}$$

Probability distribution over mixed data

Feasible estimation procedure

Simulation

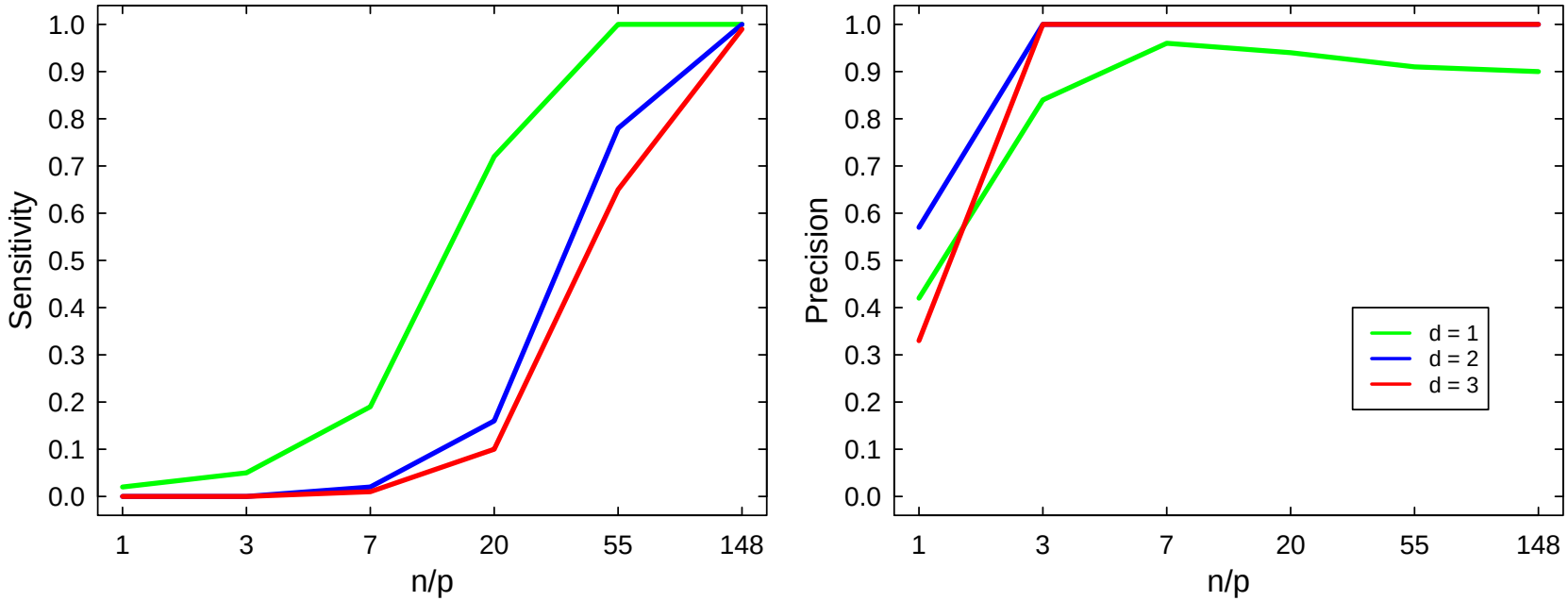


Performance measures:

Sensitivity $:= \frac{|\hat{E} \cap E_0|}{|E_0|}$ (true positive rate)

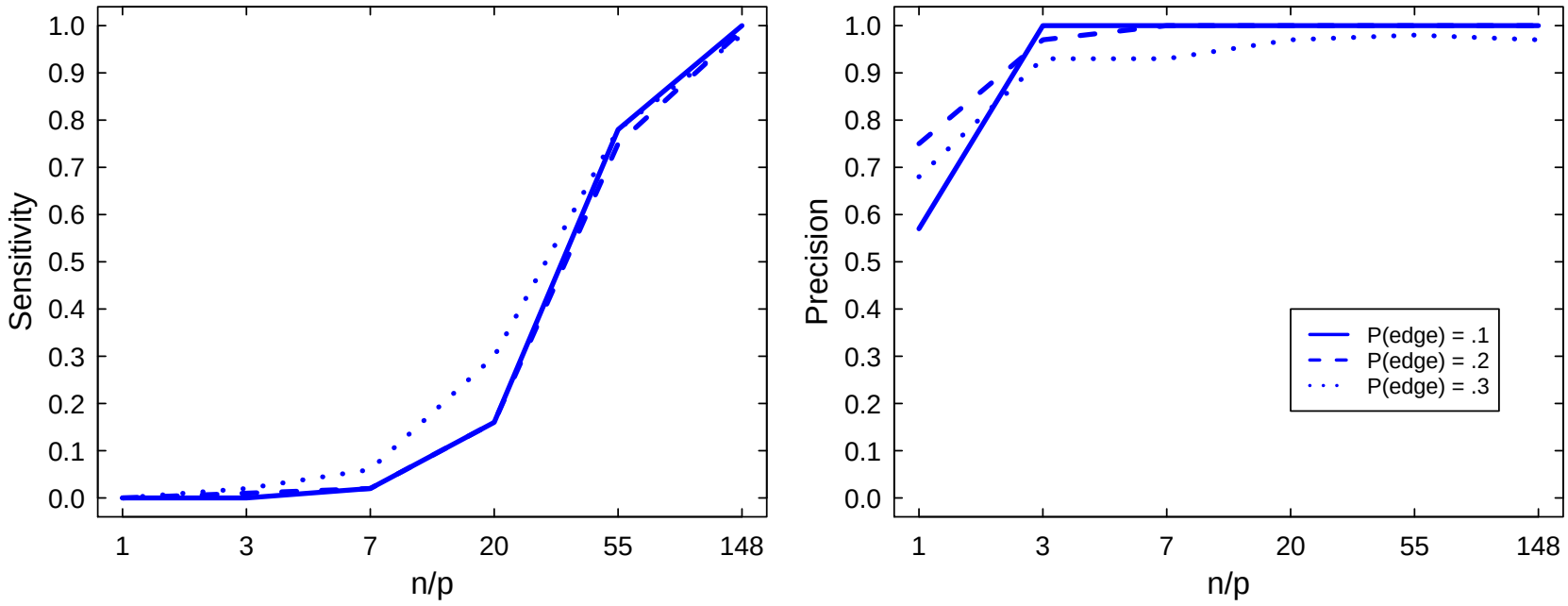
Precision $:= \frac{|\hat{E} \cap E_0|}{|\hat{E}|}$ (positive predictive value)

Simulation: Potts model (m = 3)



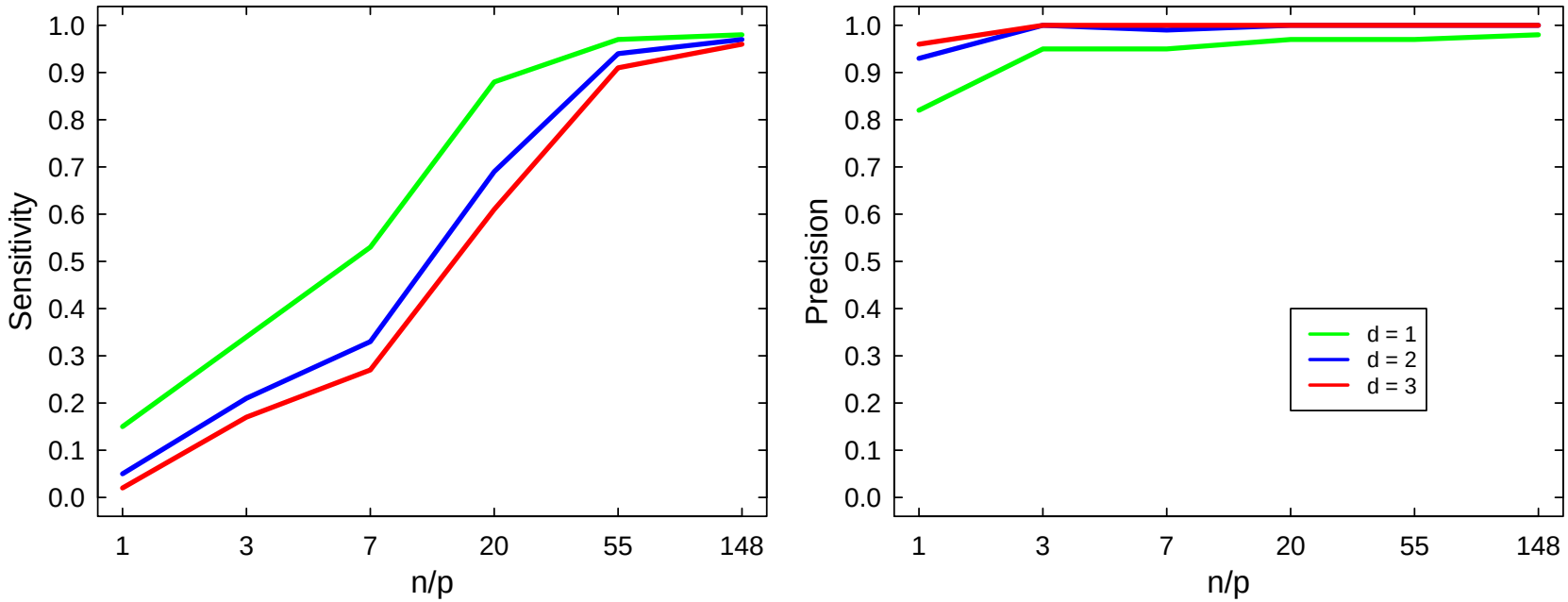
Random graph; $P(Edge) = .1$

Simulation: Potts model (m = 3)



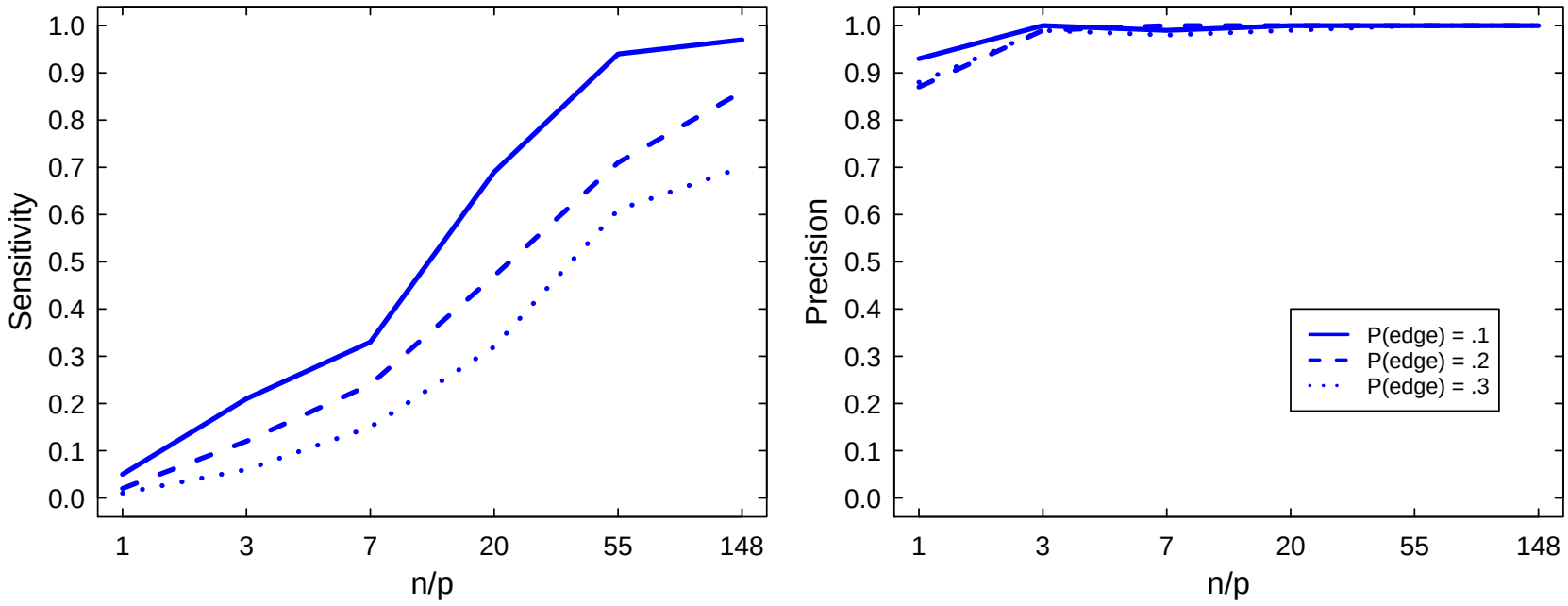
Random graph; $d = 2$

Simulation: Ising-Gaussian



Random graph; $P(Edge) = .1$

Simulation: Ising-Gaussian

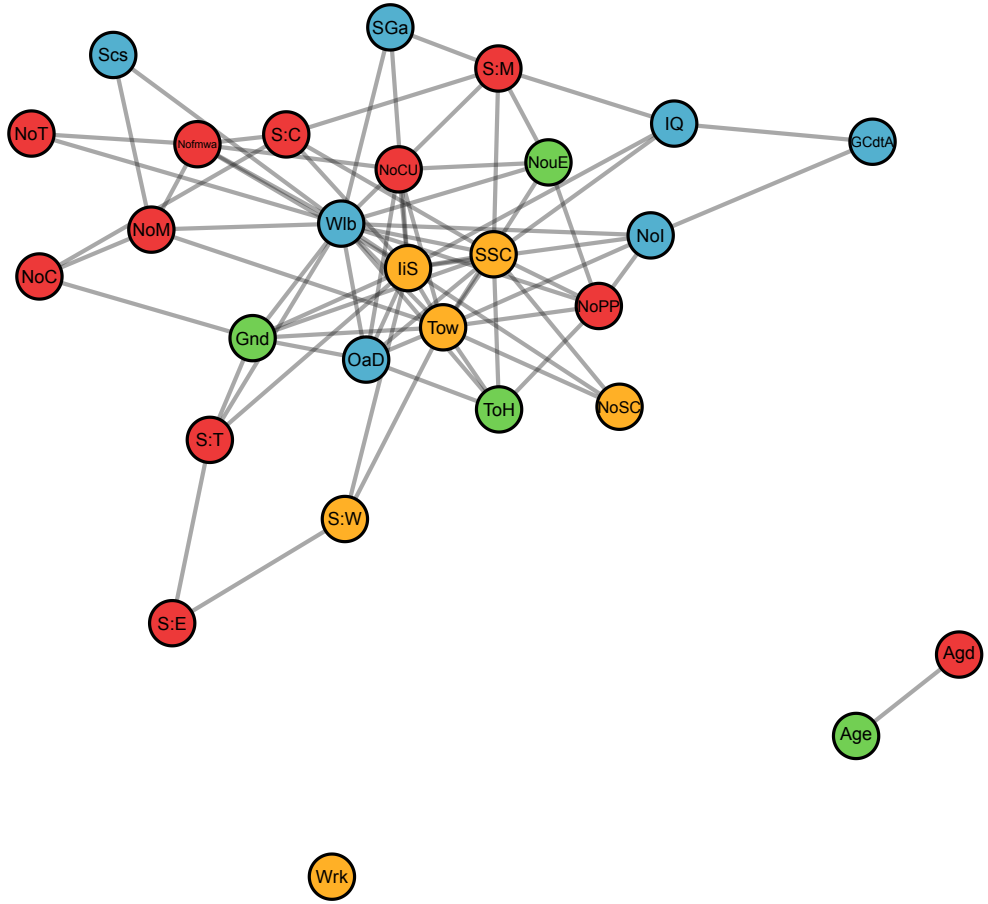


Random graph; $d = 2$

Exploring Autism-dataset

- 27 Variables describing the life of individuals diagnosed with Autism Spectrum Disorder (ASD) in the Netherlands ($N = 3521$)
- Variables: Workinghours, Type of Work, Type of housing, Satisfaction with Work, Openness about diagnosis, Education, Interests, IQ, Integration in Society, ...

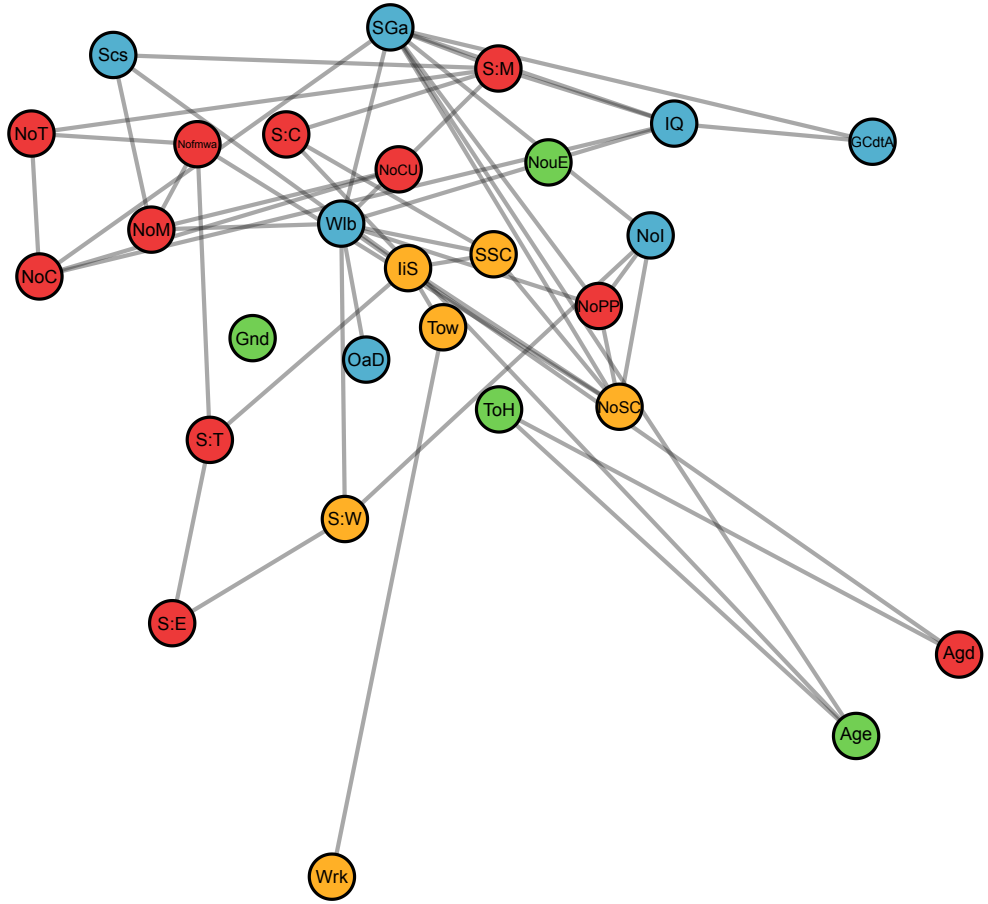
Exploring Autism-dataset: Graph-visualization



Gnd: Gender
IQ: IQ
Agd: Age diagnosis
OaD: Openness about Diagnosis
Scs: Success selfrating
Wlb: Well being
IiS: Integration in Society
Nofmwa: No of family members with autism
NoC: No of Comorbidities
NoPP: No of Physical Problems
NoT: No of Treatments
NoM: No of Medications
NoCU: No of Care Units
ToH: Type of Housing
NouE: No of unfinished Educations
Tow: Type of work
Wrk: Workinghours
NoI: No of Interests
NoSC: No of Social Contacts
GCdtA: Good Characteristics due to Autism
SGa: Satisfaction: Given advice
S:T: Satisfaction: Treatment
S:M: Satisfaction: Medication
S:C: Satisfaction: Care
S:E: Satisfaction: Education
S:W: Satisfaction: Work
SSC: Satisfaction: Social Contacts
Age: Age

Mixed distribution

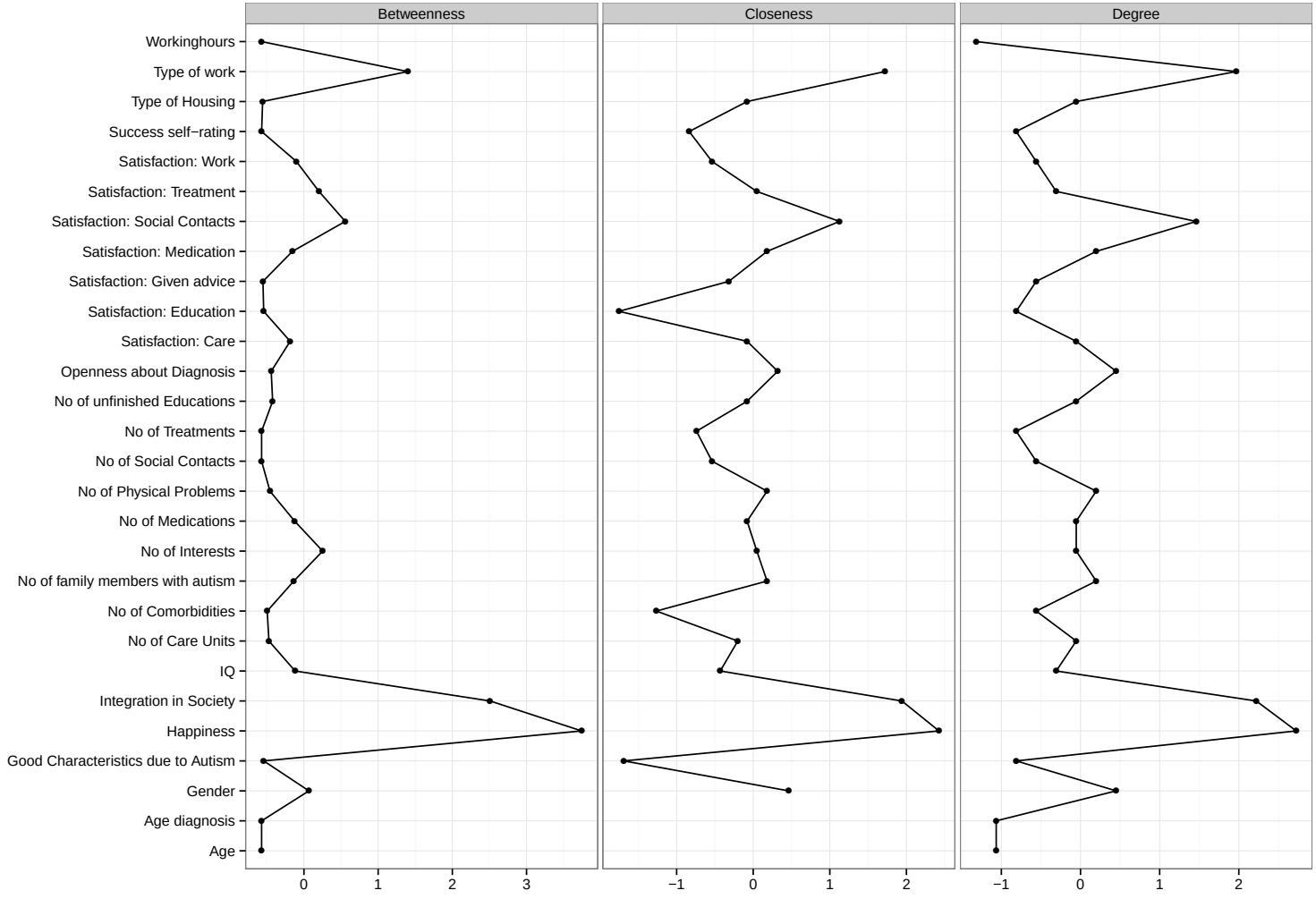
Exploring Autism-dataset: Graph-visualization



Gnd: Gender
IQ: IQ
Agd: Age diagnosis
OaD: Openness about Diagnosis
Scs: Success selfrating
Wlb: Well being
IiS: Integration in Society
Nofmwa: No of family members with autism
NoC: No of Comorbidities
NoPP: No of Physical Problems
NoT: No of Treatments
NoM: No of Medications
NoCU: No of Care Units
ToH: Type of Housing
NouE: No of unfinished Educations
Tow: Type of work
Wrk: Workinghours
NoI: No of Interests
NoSC: No of Social Contacts
GCdtA: Good Characteristics due to Autism
SGa: Satisfaction: Given advice
S:T: Satisfaction: Treatment
S:M: Satisfaction: Medication
S:C: Satisfaction: Care
S:E: Satisfaction: Education
S:W: Satisfaction: Work
SSC: Satisfaction: Social Contacts
Age: Age

Gaussian distribution

Exploring Autism-dataset: Centrality-measures



R-package **mgm**

Install:

```
library(devtools)
install_github('jmbh/mgm') # development version
install.packages('mgm') # CRAN version
library(mgm)
```

Fit a mixed graphical model:

```
> round(head(data_mixed2),4)[1:3,]
        [,1] [,2]    [,3] [,4] [,5]
[1,]  0.6680    1 11.4234    2    2
[2,] -0.7114    1 22.9344    1    1
[3,]  1.2265    0 34.4966    3    1

type <- c("g", "p", "e", "c", "c")
levs <- c(1,1,1,3,2)

set.seed(5)
fit <- mgmfit(data = data_example, type = type, lev = levs,
              lambda.sel = "CV", folds = 10, gam = .25,
              d = 2, rule.reg = "AND", rule.cat = "OR")

?mgmfit
```

R-package **mgm**: Output

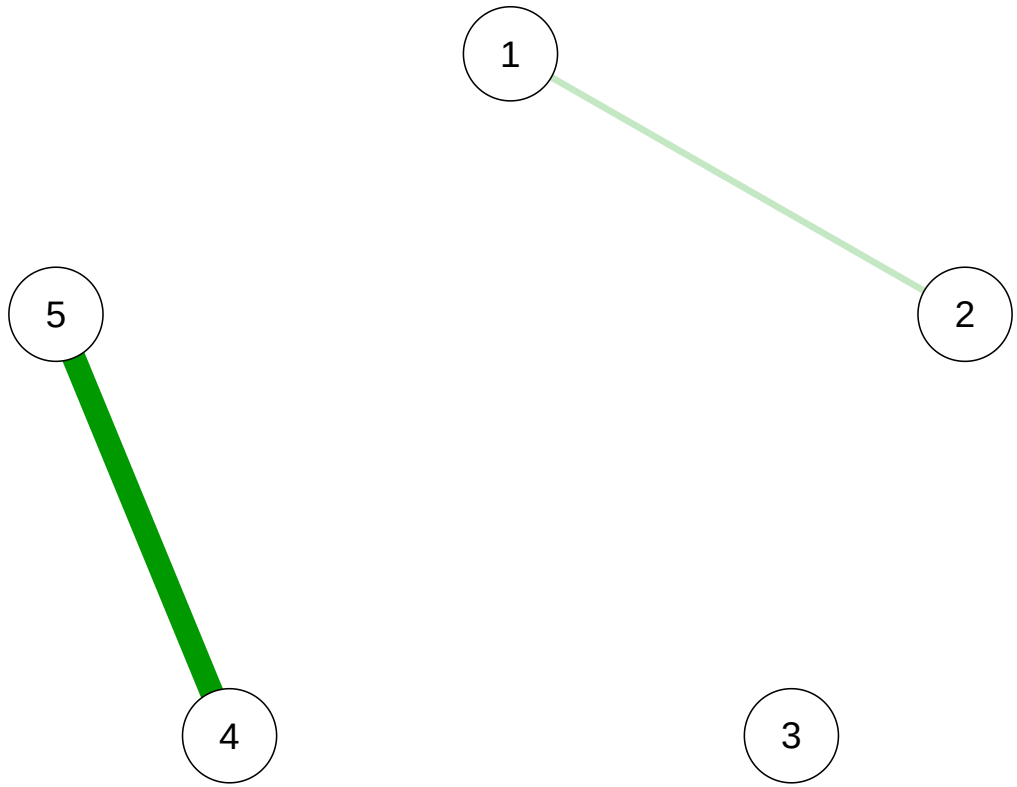
Output:

```
> fit
$adj
     [,1] [,2] [,3] [,4] [,5]
[1,]    0    1    0    0    0
[2,]    1    0    0    0    0
[3,]    0    0    0    0    0
[4,]    0    0    0    0    1
[5,]    0    0    0    1    0

$wadj
          [,1]     [,2] [,3]      [,4]      [,5]
[1,] 0.000000 0.177809    0 0.0000000 0.0000000
[2,] 0.177809 0.000000    0 0.0000000 0.0000000
[3,] 0.000000 0.000000    0 0.0000000 0.0000000
[4,] 0.000000 0.000000    0 0.0000000 0.7687597
[5,] 0.000000 0.000000    0 0.7687597 0.0000000

?mgmfit
```

R-package **mgm**: Visualize

Output:

```
library(qgraph)
qgraph(fit$wadj)
```



Summary

1. Method to estimate Mixed Graphical Models
2. Simulations: works in practical situations
3. R-package implementation: **mgm**

Contact:

jonashaslbeck@gmail.com

http://jmbh.github.io/