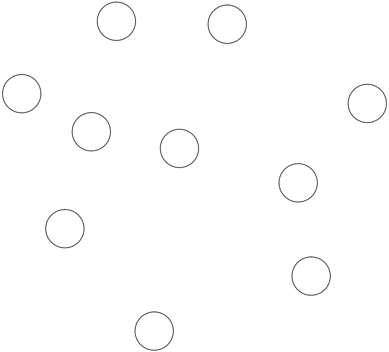
Time-Varying Mixed Graphical Models

Jonas Haslbeck and Lourens Waldorp

Psychological Methods
University of Amsterdam, the Netherlands

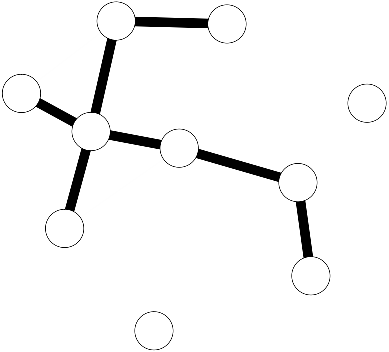
IOPS Winter Conference 2016

Multivariate System: Patient



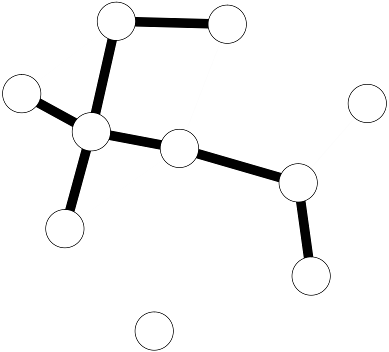
Variables are symptoms and contextual variables

Multivariate System: Patient



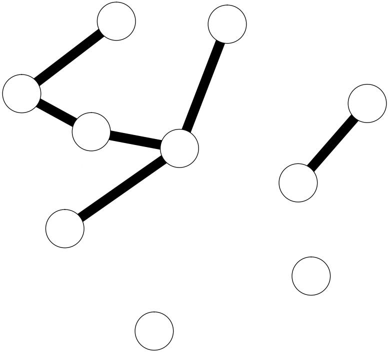
Possibly very complicated interactions

Multivariate System: Patient



All parameters can change over time

Goal: Approximate Multivariate System



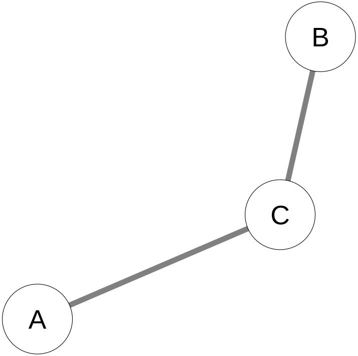
... using time-varying Mixed Graphical Models

What are Graphical Models?

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

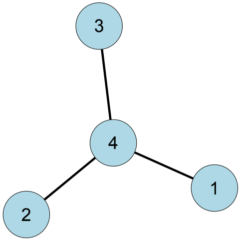
$$X_A \not\perp\!\!\!\perp X_C | X_B \qquad \Longleftrightarrow$$

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



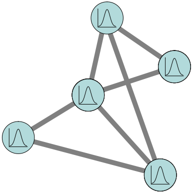
Example: Gaussian Graphical Model

$$\Sigma^{-1} = \begin{array}{c c} & \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{array}{c} X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} & \begin{pmatrix} 3.45 & 0 & 0 & 3.18 \\ 0 & 2.14 & 0 & 0.82 \\ 0 & 0 & 3.21 & 1.05 \\ 3.18 & 0.82 & 1.05 & 8.77 \end{pmatrix} \end{array} \Longleftrightarrow$$



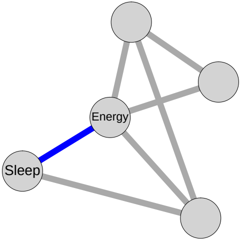
$$P(X_1, \dots, X_p) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\}$$

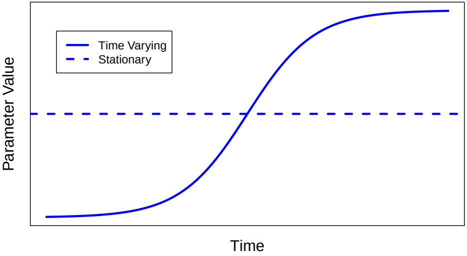
Time-**in**variant Graphical Model



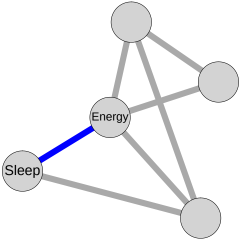
$$
\begin{array}{c c}
\textit{Time} & \begin{array}{ccccc} X_1 & X_2 & X_3 & X_4 & X_5 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ \vdots \\ T-1 \\ T \end{array} &
\begin{pmatrix}
3.45 & 1 & 0.98 & 3 & 1 \\
1.11 & 3 & 0.82 & 3 & 2 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0.12 & 2 & 0.71 & 2 & 2 \\
-0.78 & 1 & 0.18 & 1 & 1
\end{pmatrix}
\end{array}
$$

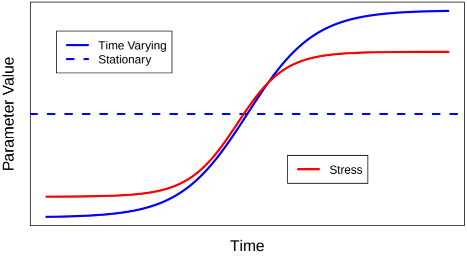
Parameters may change over time!



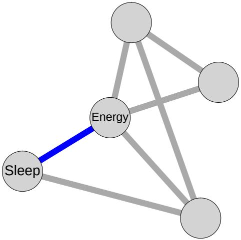


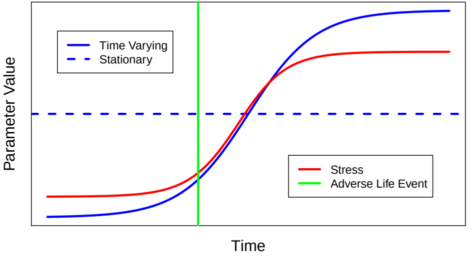
Parameters may change over time!





Parameters may change over time!



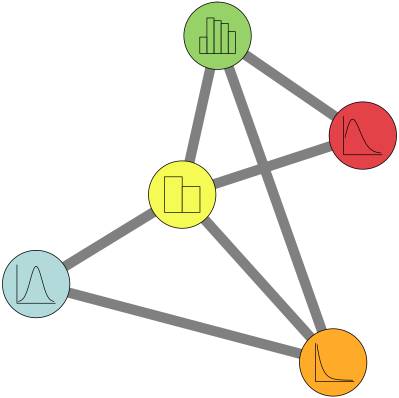


Introducing time-varying Mixed Graphical Models

1) Stationary Mixed Graphical Models

2) Extension to the time-varying case

Mixed Exponential Graphical Model



Construction of Mixed Exponential Graphical Model

Conditional univariate members of the exponential family

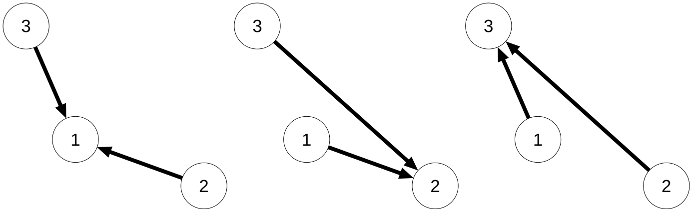
$$P(X_s|X_{\setminus s}) = \exp\left\{E_s(X_{\setminus s})\phi_s(X_s) + C_s(X_s) - \Phi(X_{\setminus s})\right\},$$

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:

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

(Yang et al., 2014)

Neighborhood Regression Method



(Meinshausen & Buehlmann, 2006)

Estimation Algorithm

For each node s :

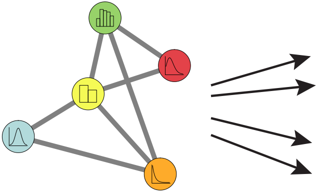
1. Regress $X_{\setminus s}$ on X_s
   - $\min_{(\theta_0,\theta)\in\mathbb{R}^p} \left[\frac{1}{N} \sum_{i=1}^{N} (y_i - \theta_0 - X_{\setminus s;i}^T \theta)^2 + \lambda_n ||\theta||_1 \right]$
   - Select λ_n using EBIC
2. Threshold Parameter Estimates
   - $\tau_n \asymp \sqrt{d}||\theta||_2 \sqrt{\frac{\log p}{n}}$

Combine Estimates from both regressions

- AND-rule: Edge present if both parameters are nonzero
- OR-rule: Edge present if at least one parameter is nonzero

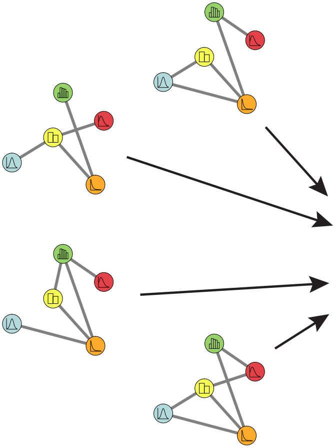
(Loh & Wainwright, 2013)

Recab: Time-**in**variant mixed Graphical Model



$$
\begin{array}{l c}
\textit{Time} & \begin{array}{ccccc} X_1 & X_2 & X_3 & X_4 & X_5 \end{array} \\
\begin{array}{l} 1 \\ 2 \\ \vdots \\ T-1 \\ T \end{array} &
\begin{pmatrix}
3.45 & 1 & 0.98 & 3 & 1 \\
1.11 & 3 & 0.82 & 3 & 2 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0.12 & 2 & 0.71 & 2 & 2 \\
-0.78 & 1 & 0.18 & 1 & 1
\end{pmatrix}
\end{array}
$$

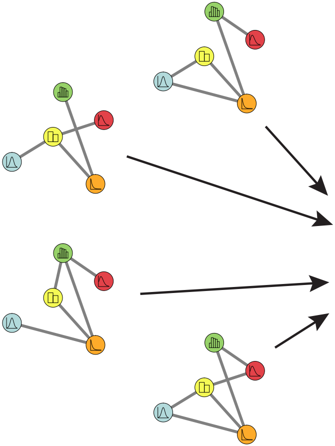
Time-varying mixed Graphical Model



$$
\begin{array}{ll}
 & \begin{array}{ccccc} X_1 & X_2 & X_3 & X_4 & X_5 \end{array} \\
\begin{array}{l} 1 \\ 2 \\ \vdots \\ T-1 \\ T \end{array} &
\begin{pmatrix}
3.45 & 1 & 0.98 & 3 & 1 \\
1.11 & 3 & 0.82 & 3 & 2 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0.12 & 2 & 0.71 & 2 & 2 \\
-0.78 & 1 & 0.18 & 1 & 1
\end{pmatrix}
\end{array}
$$

The row labels above are headed *Time*.

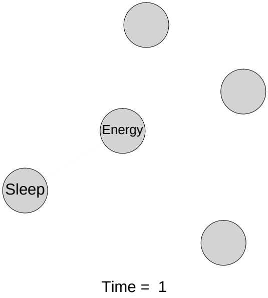
Time-varying mixed Graphical Model

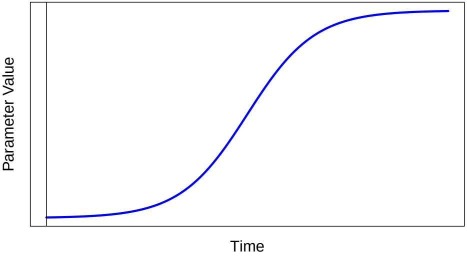


$$
\begin{array}{l c}
\textit{Time} & \begin{array}{ccccc} \quad X_1 & \quad X_2 & \quad X_3 & \quad X_4 & \quad X_5 \end{array} \\
\begin{array}{l} 1 \\ 2 \\ \vdots \\ T-1 \\ T \end{array} &
\begin{pmatrix}
3.45 & 1 & 0.98 & 3 & 1 \\
1.11 & 3 & 0.82 & 3 & 2 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0.12 & 2 & 0.71 & 2 & 2 \\
-0.78 & 1 & 0.18 & 1 & 1
\end{pmatrix}
\end{array}
$$

But: we have the scaling $\tau_n \asymp \sqrt{d}||\theta||_2\sqrt{\frac{\log p}{n}}$

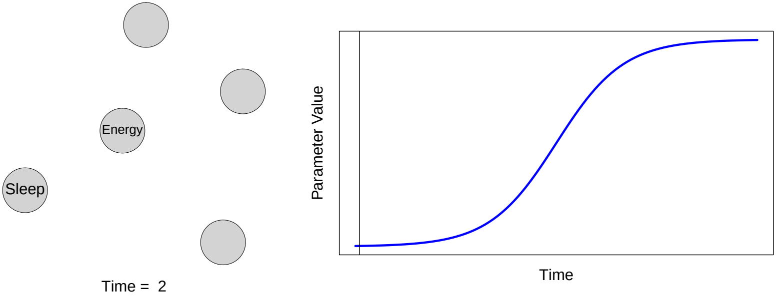
Local Stationarity





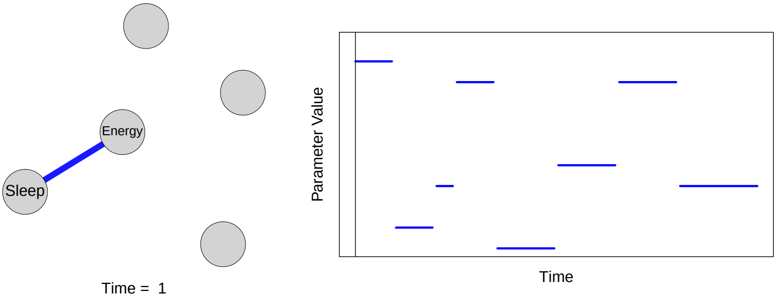
Assumption: Edge parameters are a smooth function of time

Local Stationarity



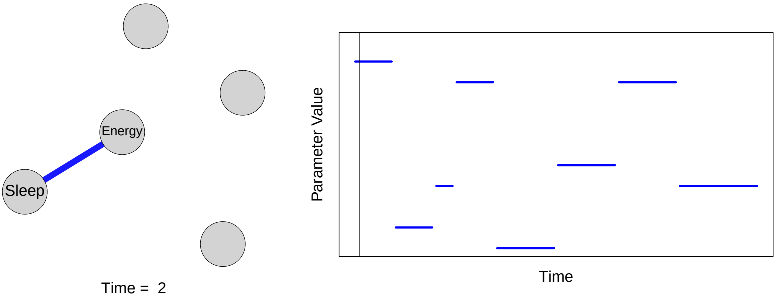
Assumption: Edge parameters are a smooth function of time

Local Stationarity Violated!



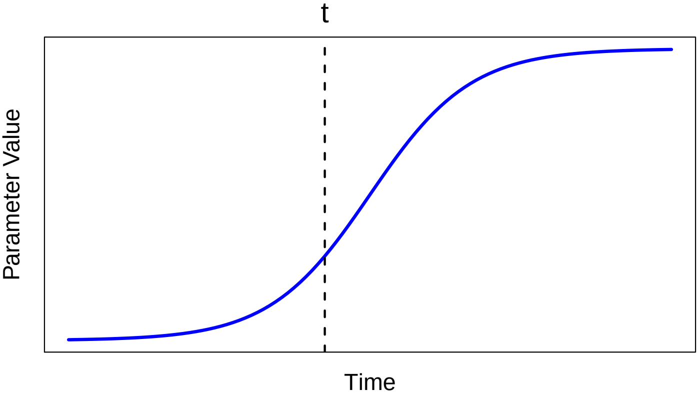
Edge parameters are *not* a smooth function of time!

Local Stationarity Violated!

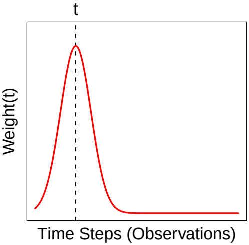


Edge parameters are *not* a smooth function of time!

Again: Local Stationarity



Time-varying Graphs via Weighted Regression

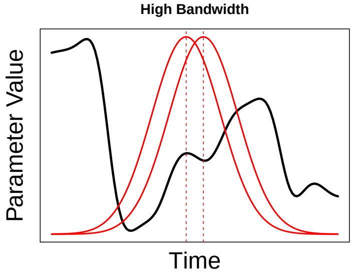


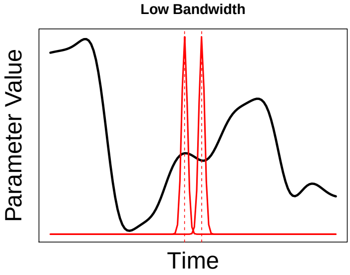
$$
\begin{array}{c}
\textit{Time} \\
1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ \vdots \\ T
\end{array}
\begin{array}{c}
\begin{array}{ccccc}
X_1 & \ldots & X_5 & \textit{Weight}(t)
\end{array} \\
\begin{pmatrix}
0.92 & \ldots & -1.47 & \color{red}{0} \\
0.78 & \ldots & -0.48 & \color{red}{0.13} \\
0.07 & \ldots & 0.42 & \color{red}{0.60} \\
-1.99 & \ldots & 1.36 & \color{red}{1.00} \\
0.62 & \ldots & -0.10 & \color{red}{0.60} \\
-0.06 & \ldots & 0.39 & \color{red}{0.13} \\
-0.21 & \ldots & 0.99 & \color{red}{0} \\
\vdots & \vdots & \vdots & \color{red}{\vdots} \\
-0.16 & \ldots & 0.18 & \color{red}{0}
\end{pmatrix}
\end{array}
$$

Weighted cost function:

$$\min_{(\theta_0,\theta)\in\mathbb{R}^p} \left[\frac{1}{N} \sum_{i=1}^{N} \color{red}{w_{i;t}} (y_{i;t} - \theta_{0;t} - X_{\backslash s;i}^{T} \theta_t)^2 + \lambda_n ||\theta_t||_1 \right]$$

What is the right bandwidth?



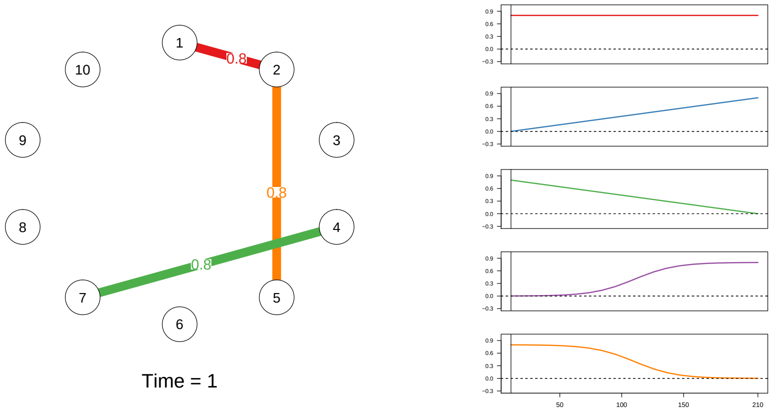


More Information for estimation

Higher sensitivity to change

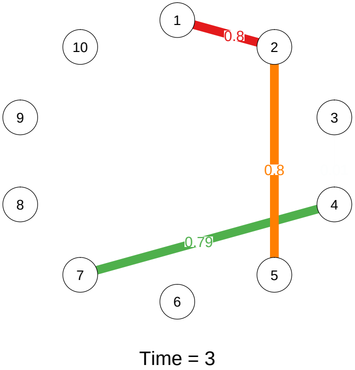
$$\text{Scaling: } \tau_n \asymp \sqrt{d}||\theta||_2\sqrt{\frac{\log p}{n}}$$

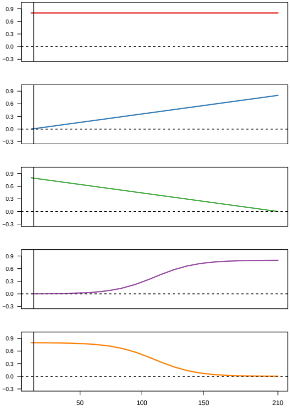
Small Simulation: Typical ESM Data



10 measurements/day $\times$ 3 weeks $=$ 210

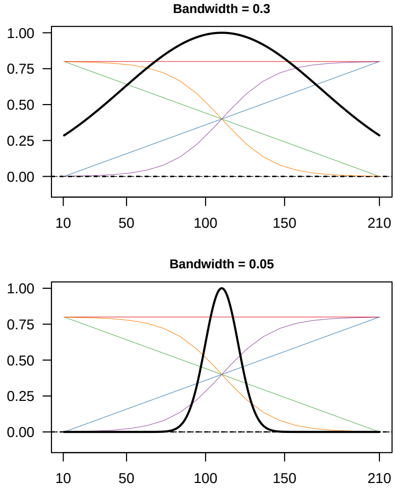
Small Simulation: Typical ESM Data



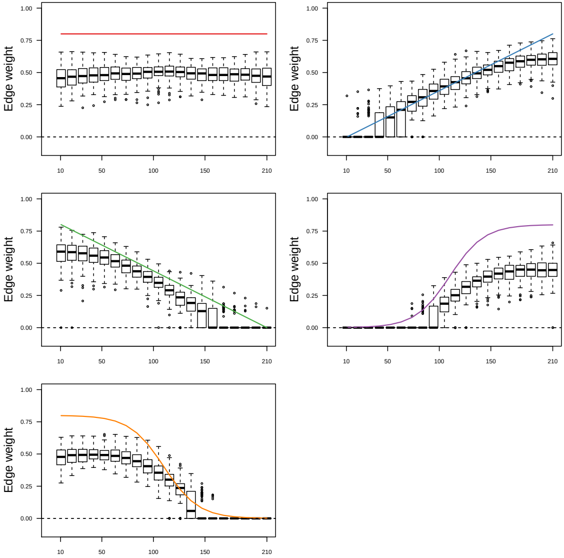


10 measurements/day $\times$ 3 weeks $=$ 210

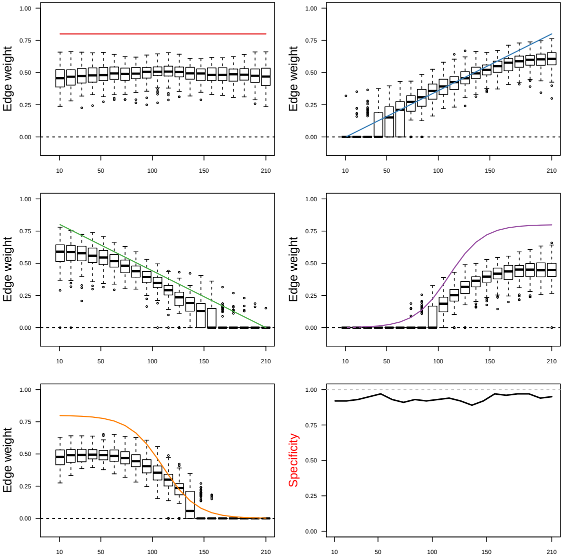
Typical ESM Data: Select Bandwidth



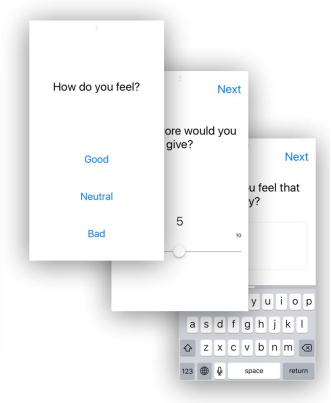
Small Simulation: Results



Small Simulation: Results

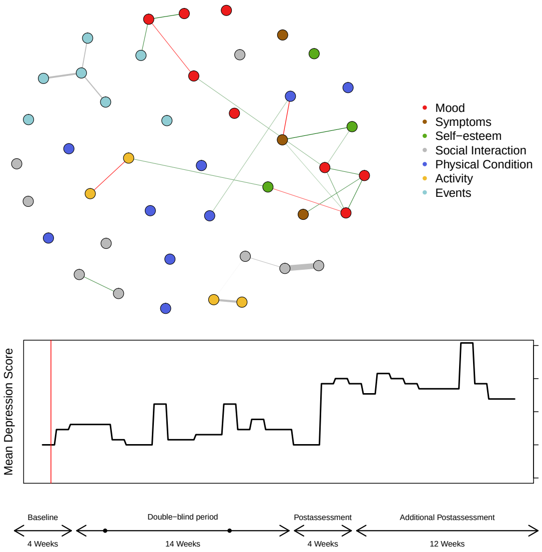


Application to Event Sampling Data



- Time series of 1 person
- 43 variables (continuous & categorical)
- Up to 10 measurements a day
- For 36 weeks

Time-varying Graph of Psychopathology



Time-varying Mixed Graphical Models

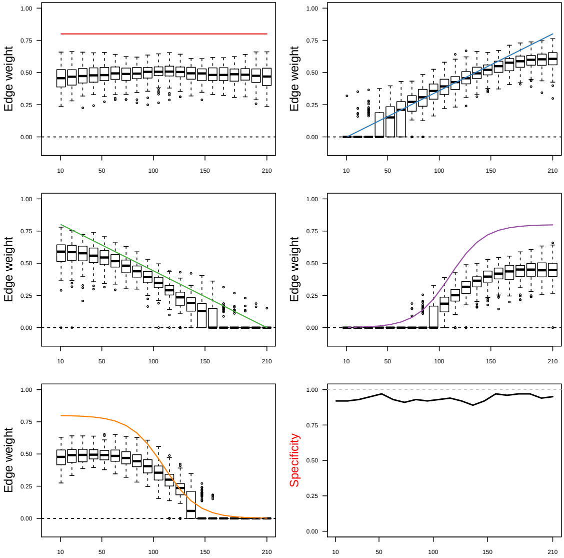
Summary

- Time varying model under assumption of local stationarity
- Allows for mixed variables (e.g. categorical and continuous)
- Scales well for large *p* and allows for $p > n$
- Works in realistic situations
- Also a time-varying mixed VAR version implemented
- Available via R-package *mgm* on CRAN

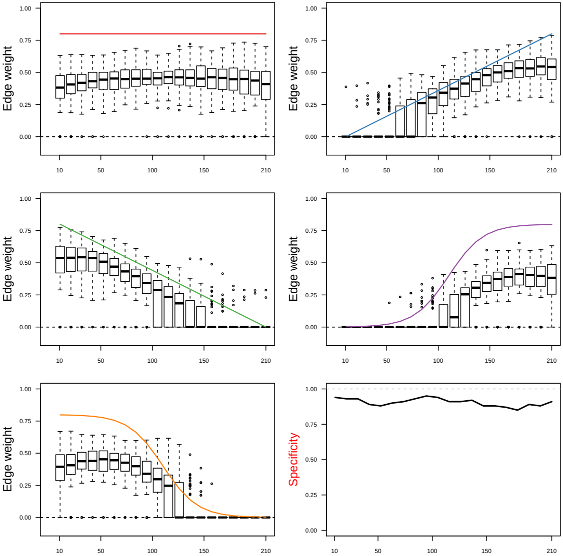
Contact

- Email: `jonashaslbeck@gmail.com`
- Website: `http://jmbh.github.io`

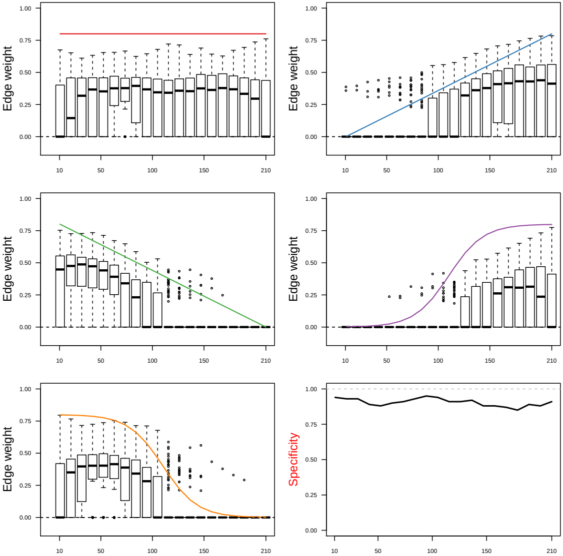
Small Simulation: Same as above N=210



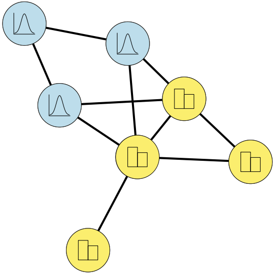
Small Simulation: Same as above but now N=100



Small Simulation: Same as above but now N=50



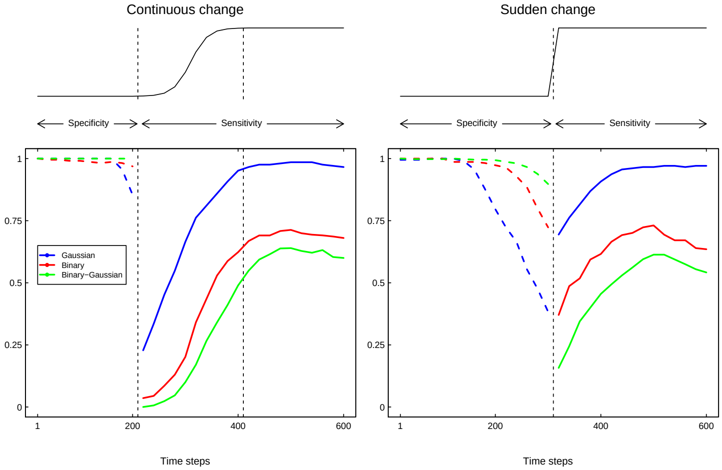
Larger Simulation: Setup



- Binary-Gaussian Graphical Model
- 20 Nodes
- Always 19 edges present
- Of these are always 6 changing
- Type of change: smooth vs. sudden

Simulation: Results

N = 600
Nt = 205.9
Nt/p = 10.3



$$\text{bandwidth} = 0.8/N^{1/3} \approx 0.095$$

Mixed Graphical Model: Conditional Distribution

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:

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

Mixed Graphical Model: Joint Distribution

The joint distribution has the form

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

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

Example Mixed Graphical Model: 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:

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