Abstracting Complex Systems using Mixed Graphical Models

Jonas Haslbeck

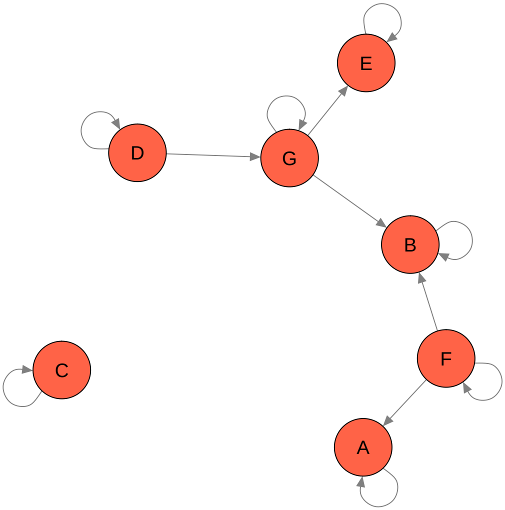
Psychosystems lab
University of Amsterdam, the Netherlands

psychosystems.org
jmbh.github.io

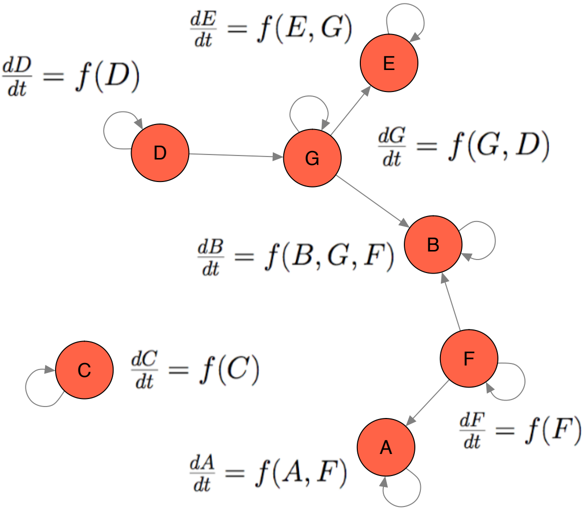
Complexity Laboratory Utrecht (CLUe) Lunch Meeting

Utrecht, October 20th

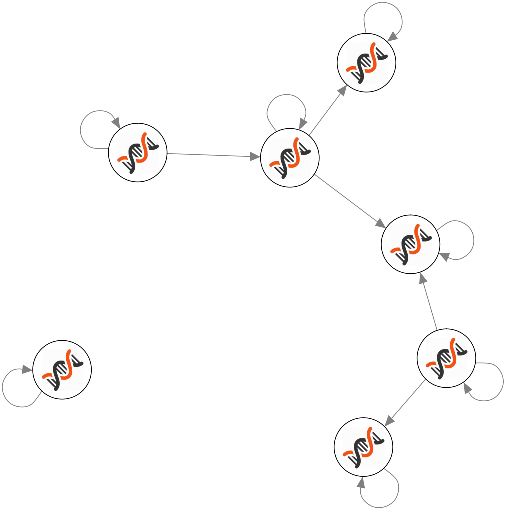
Multivariate System



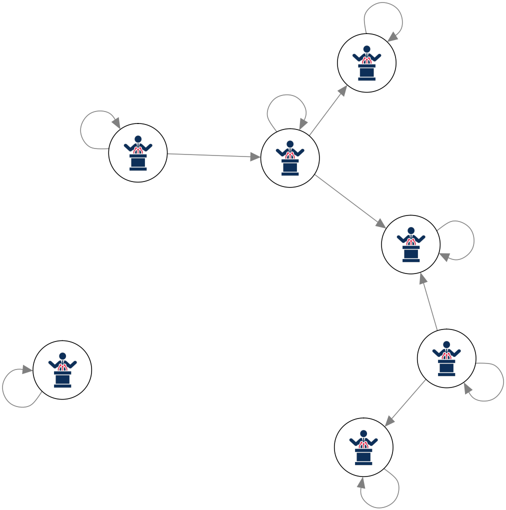
Multivariate System



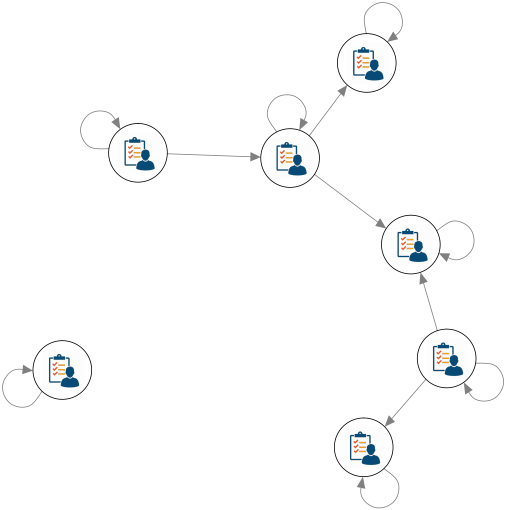
Gene Expressions

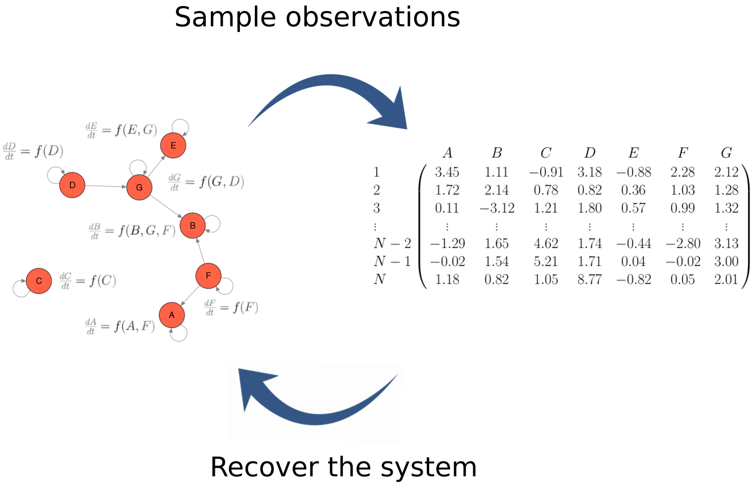


Voting Behavior of Members of Parliament



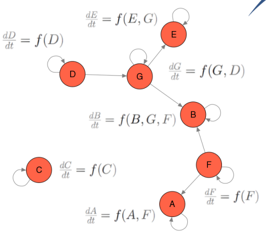
Symptoms of Mental Disorders





Sample observations
$\frac{dE}{dt} = f(E, G)$
$\frac{dD}{dt} = f(D)$
$\frac{dG}{dt} = f(G, D)$
$\frac{dB}{dt} = f(B, G, F)$
$\frac{dC}{dt} = f(C)$
$\frac{dF}{dt} = f(F)$
$\frac{dA}{dt} = f(A, F)$
E
D
G
B
C
F
A
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ N-2 \\ N-1 \\ N \end{array} \begin{array}{c} \begin{array}{ccccccc} A & B & C & D & E & F & G \end{array} \\ \begin{pmatrix} 3.45 & 1.11 & -0.91 & 3.18 & -0.88 & 2.28 & 2.12 \\ 1.72 & 2.14 & 0.78 & 0.82 & 0.36 & 1.03 & 1.28 \\ 0.11 & -3.12 & 1.21 & 1.80 & 0.57 & 0.99 & 1.32 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1.29 & 1.65 & 4.62 & 1.74 & -0.44 & -2.80 & 3.13 \\ -0.02 & 1.54 & 5.21 & 1.71 & 0.04 & -0.02 & 3.00 \\ 1.18 & 0.82 & 1.05 & 8.77 & -0.82 & 0.05 & 2.01 \end{pmatrix} \end{array}$$
Recover the system

Sample observations

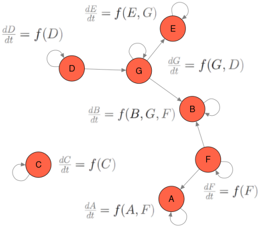


$$
\begin{array}{l}
 \\ 1 \\ 2 \\ 3 \\ \vdots \\ N-2 \\ N-1 \\ N
\end{array}
\begin{array}{c}
\begin{array}{ccccccc} A & B & C & D & E & F & G \end{array} \\
\begin{pmatrix}
3.45 & 1.11 & -0.91 & 3.18 & -0.88 & 2.28 & 2.12 \\
1.72 & 2.14 & 0.78 & 0.82 & 0.36 & 1.03 & 1.28 \\
0.11 & -3.12 & 1.21 & 1.80 & 0.57 & 0.99 & 1.32 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
-1.29 & 1.65 & 4.62 & 1.74 & -0.44 & -2.80 & 3.13 \\
-0.02 & 1.54 & 5.21 & 1.71 & 0.04 & -0.02 & 3.00 \\
1.18 & 0.82 & 1.05 & 8.77 & -0.82 & 0.05 & 2.01
\end{pmatrix}
\end{array}
$$

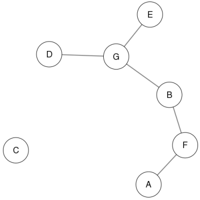
Approximate the system

True Model, Probability Distribution, Network Model

True Model



Conditional Independence Network



Approximate

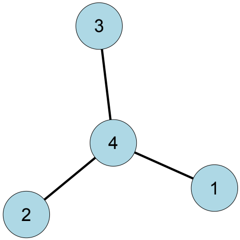
Summarize

$$P(X_1, \ldots, X_p, \theta)$$

Multivariate Probability Distribution

Simple Example: Gaussian Graphical Model

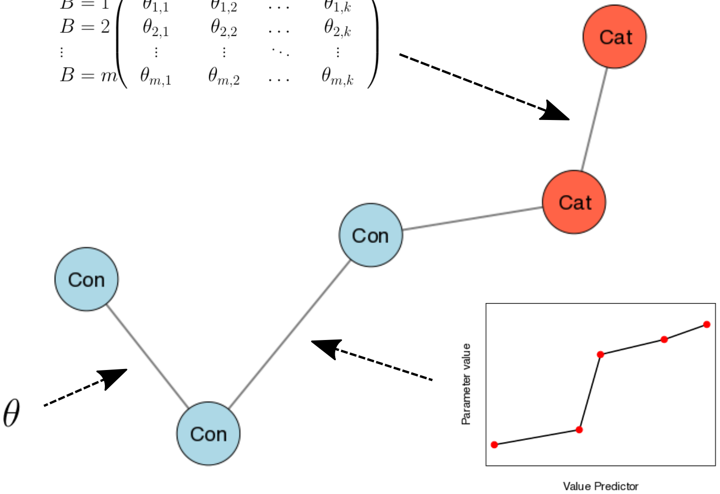
$$\Sigma^{-1} = \begin{array}{c|c} & \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{array}{c} X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} & \begin{pmatrix} 3.45 & 0 & 0 & 3.18 \\ 0 & 2.14 & 0 & 0.82 \\ 0 & 0 & 3.21 & 1.05 \\ 3.18 & 0.82 & 1.05 & 8.77 \end{pmatrix} \end{array} \iff$$



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

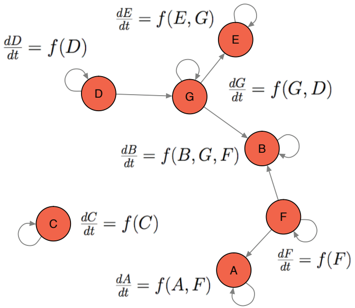
General Graphical Models

$$
\begin{array}{c}
 \\
B=1 \\
B=2 \\
\vdots \\
B=m
\end{array}
\begin{array}{c}
\begin{array}{cccc} A=1 & A=2 & \dots & A=k \end{array} \\
\begin{pmatrix}
\theta_{1,1} & \theta_{1,2} & \cdots & \theta_{1,k} \\
\theta_{2,1} & \theta_{2,2} & \cdots & \theta_{2,k} \\
\vdots & \vdots & \ddots & \vdots \\
\theta_{m,1} & \theta_{m,2} & \cdots & \theta_{m,k}
\end{pmatrix}
\end{array}
$$

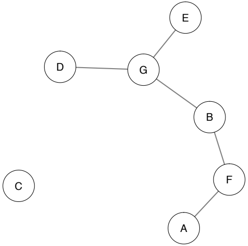


Goal:

Abstract structure of true system in simpler MGM model class

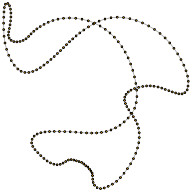


True System

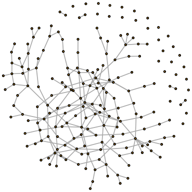


Abstraction in simpler MGM model class

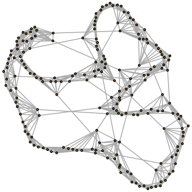
Study multivariate distribution as network



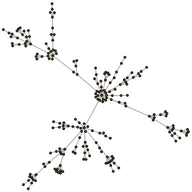
Ring Network



Random Network

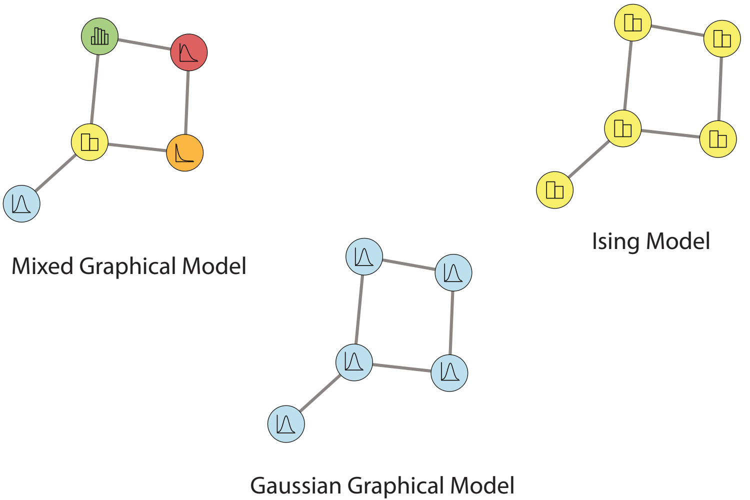


Small World Network



Scale−free Network

Mixed Graphical Models



Constructing MGMs

Each node/variable is a univariate exponential family distribution conditional on all other variables

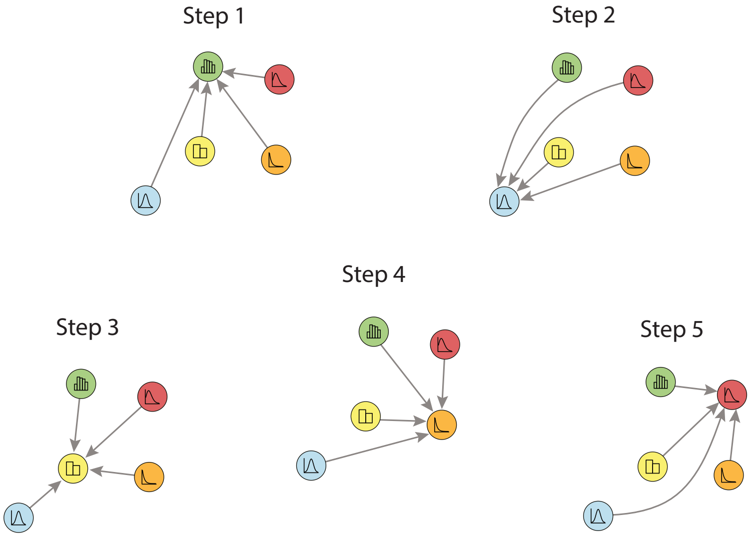
$$P(X, \beta) = \exp\left\{\eta(\theta)B(X) + C(X) - A(\theta)\right\}$$

and the natural parameter θ is a linear combination of all other variables:

$$\theta_{t,i} = \beta_{0,i} + \begin{bmatrix}\beta_{i,1} & \dots & \beta_{i,p}\end{bmatrix} \begin{bmatrix} X_2 \\ \vdots \\ X_p \end{bmatrix}$$

(Yang et al., 2014; Chen, Witten & Shojaie, 2015; Haslbeck & Waldorp, 2017)

Estimating Mixed Graphical Models



(Meinshausen & Buehlmann, 2006)

ℓ_1-regularized Estimation

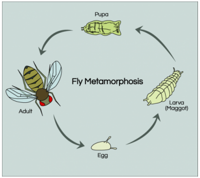
We minimize the negative log-likelihood $F(\boldsymbol{X_j}, \boldsymbol{\theta_t})$ together with the ℓ_1-norm over all parameters:

$$\arg_{\boldsymbol{\theta_t}} \min \left\{ \frac{1}{n} \sum_{j=1}^{n} w_{j,t_e} F(\boldsymbol{X_j}, \boldsymbol{\theta_t}) + \lambda_i ||\boldsymbol{\theta_t}||_1 \right\}$$

This has two consequences:

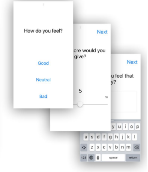
1. We can control the bias (model too simple) vs. variance (model too complicated) trade-off with tuning parameter λ_i
2. Small parameters are set to exactly zero

Back to Applications



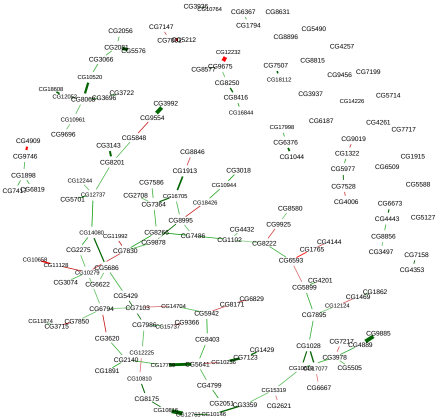
67 measurements of 150 gene expressions related to the immune system of Drosophila melanogaster (fruit fly) over its full life cycle

Votes of 623 members of the German parliament on 136 bills from Nov 2013 - April 2015



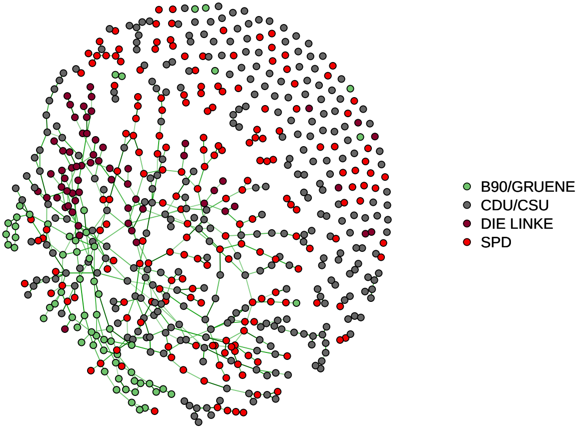
1476 measurements of 16 mood related variables of one individual over 238 consecutive days

Gene Expressions of Fruit Fly



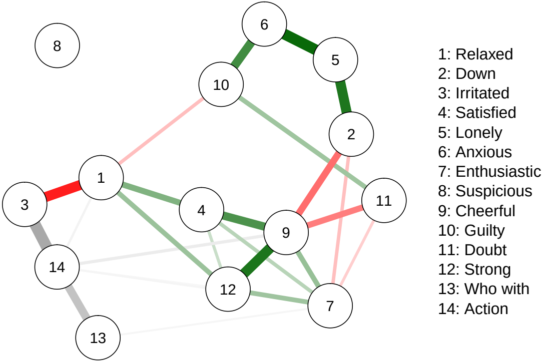
67 Measurements of 150 genes expressions related to immune system of the fruit fly (Lebre et al., 2010)

Voting Behavior of Members of German Parliament



136 public votes, 623 members of parliament of 4 parties

Symptoms of Mental Disorder



1476 measurements of 14 variables related to mood, activity and social context of one individual over 238 consecutive days (Kossakowski et al., 2017)

Practical:

Estimate MGM on Symptom Data

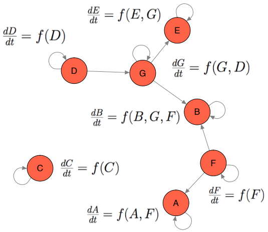
RStudio Server:

http://clue.science.uu.nl:8787

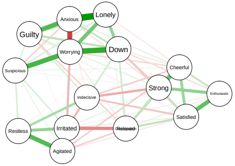
Login: Your UU Solis-ID & password

Direction of Influence & Interactions as function of time

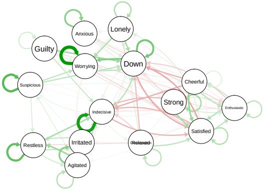
True Structure:



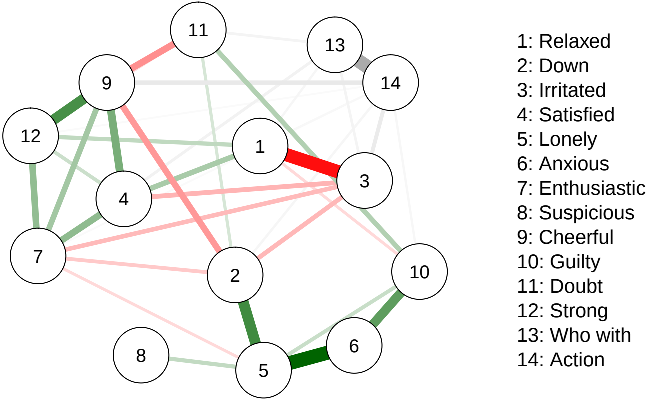
Instantaneous Influence



Influence over time (1h)

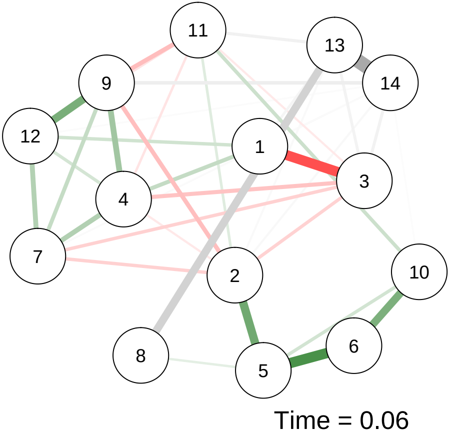


Does the system under investigation change over time?



Time = 0

Does the system under investigation change over time?

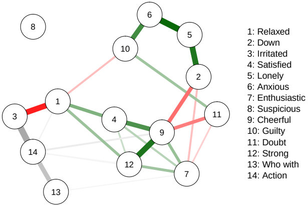


1: Relaxed
2: Down
3: Irritated
4: Satisfied
5: Lonely
6: Anxious
7: Enthusiastic
8: Suspicious
9: Cheerful
10: Guilty
11: Doubt
12: Strong
13: Who with
14: Action

mgm: Summary

mgm package implements:

- Mixed Graphical Models (MGMs)
- Time-varying MGMs
- mixed Vector Autoregressive (mVAR) models
- Time-varying mVARs



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

References

- Haslbeck, J., & Waldorp, L. J. (2017). Estimating mixed graphical models in high-dimensional data. arXiv preprint arXiv:1510.05677.
- Yang, E., Baker, Y., Ravikumar, P., Allen, G., & Liu, Z. (2014, April). Mixed graphical models via exponential families. In Artificial Intelligence and Statistics (pp. 1042-1050).
- Chen, S., Witten, D. M., & Shojaie, A. (2014). Selection and estimation for mixed graphical models. Biometrika, 102(1), 47-64.
- Kossakowski, J., Groot, P., Haslbeck, J., Borsboom, D., & Wichers, M. (2017). Data from Critical Slowing Down as a Personalized Early Warning Signal for Depression. Journal of Open Psychology Data, 5(1).
- Lebre, S., Becq, J., Devaux, F., Stumpf, M. P., & Lelandais, G. (2010). Statistical inference of the time-varying structure of gene-regulation networks. BMC systems biology, 4(1), 130.
- Meinshausen, N., & Bhlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. The Annals of Statistics, 1436-1462.