

Abstracting Complex Systems using Mixed Graphical Models

Jonas Haslbeck

Psychosystems lab

University of Amsterdam, the Netherlands

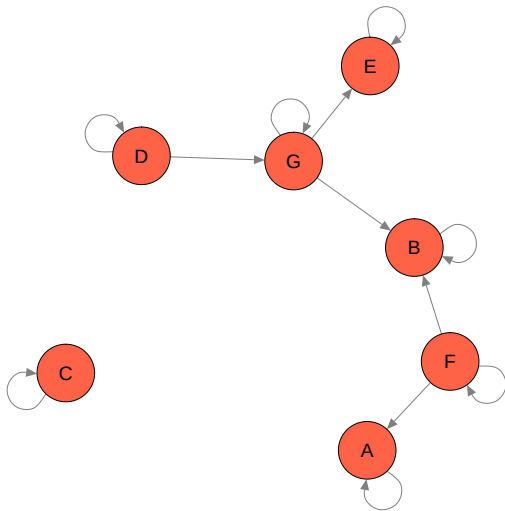
`psychosystems.org`

`jmbh.github.io`

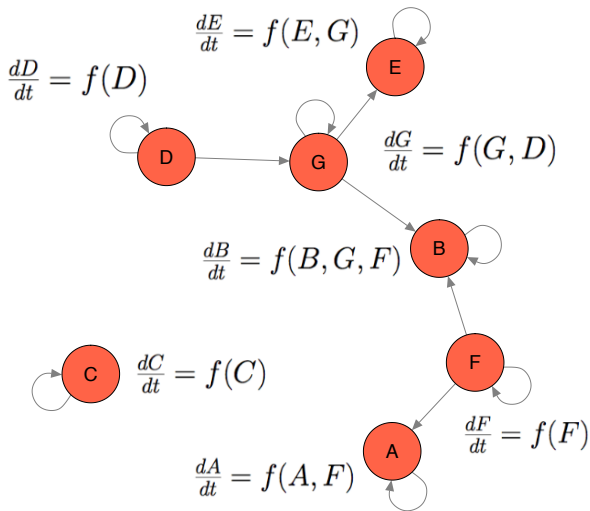
Complexity Laboratory Utrecht (CLUe) Lunch Meeting

Utrecht, October 20th

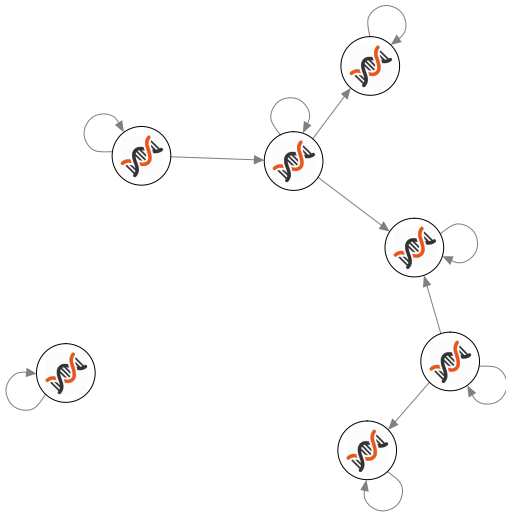
Multivariate System



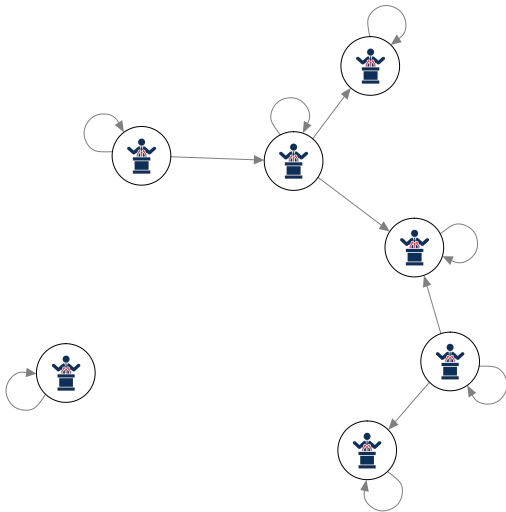
Multivariate System



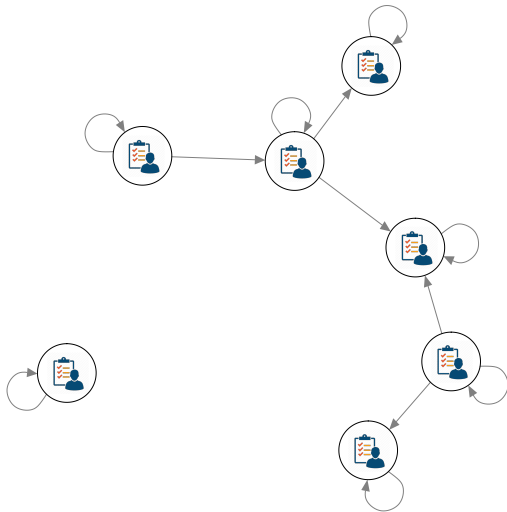
Gene Expressions



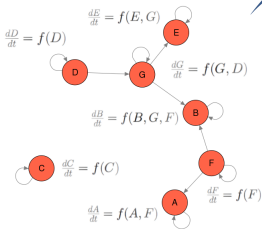
Voting Behavior of Members of Parliament



Symptoms of Mental Disorders



Sample Observations

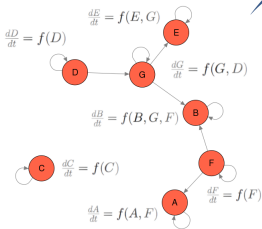



	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>
1	3.45	1.11	-0.91	3.18	-0.88	2.28	2.12
2	1.72	2.14	0.78	0.82	0.36	1.03	1.28
3	0.11	-3.12	1.21	1.80	0.57	0.99	1.32
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>N</i> - 2	-1.29	1.65	4.62	1.74	-0.44	-2.80	3.13
<i>N</i> - 1	-0.02	1.54	5.21	1.71	0.04	-0.02	3.00
<i>N</i>	1.18	0.82	1.05	8.77	-0.82	0.05	2.01



Recover the system

Sample Observations



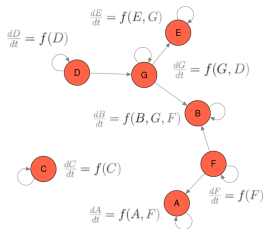

	A	B	C	D	E	F	G
1	3.45	1.11	-0.91	3.18	-0.88	2.28	2.12
2	1.72	2.14	0.78	0.82	0.36	1.03	1.28
3	0.11	-3.12	1.21	1.80	0.57	0.99	1.32
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$N - 2$	-1.29	1.65	4.62	1.74	-0.44	-2.80	3.13
$N - 1$	-0.02	1.54	5.21	1.71	0.04	-0.02	3.00
N	1.18	0.82	1.05	8.77	-0.82	0.05	2.01

Recover the **structure** of the system

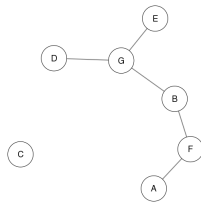


True Model, Probability Distribution, Network Model

True Model



Conditional independence network



Approximate

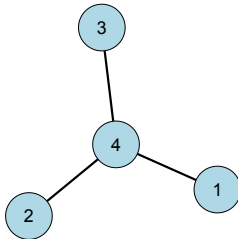
Summarize

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

Probability Distribution: Mixed Graphical Model

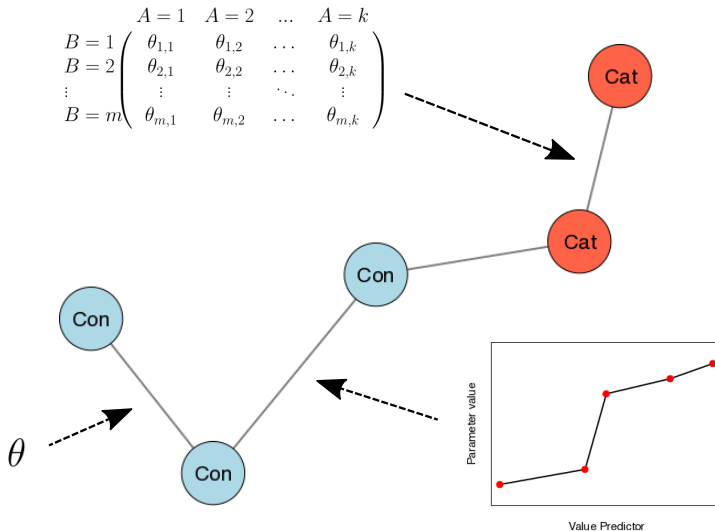
Simple Example: Gaussian Graphical Model

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

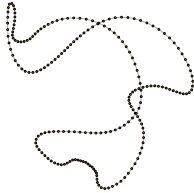


$$P(X_1, \dots, X_p) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

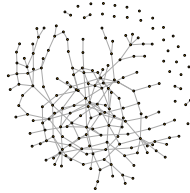
General Graphical Models



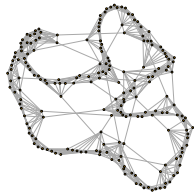
Study multivariate distribution as network



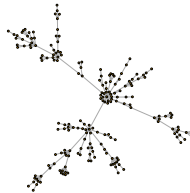
Ring Network



Random Network



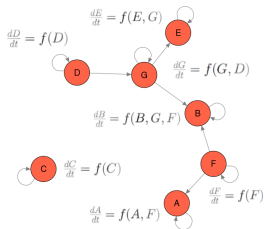
Small World Network



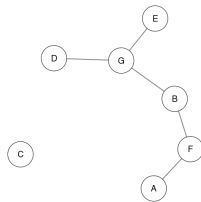
Scale-free Network

True Model, Probability Distribution, Network Model

True Model



Conditional independence network



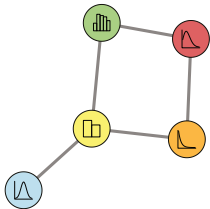
Approximate

Summarize

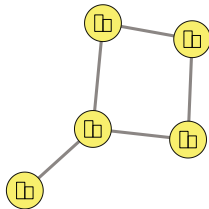
$$P(X_1, \dots, X_p, \theta)$$

Probability Distribution: Mixed Graphical Model

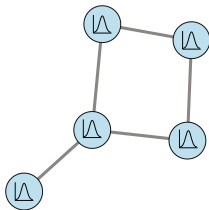
Exponential Family of Distributions



Mixed Graphical Model



Ising Model



Gaussian Graphical Model

Constructing MGMs

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

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

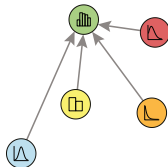
and the natural parameter θ is a linear combination of all other variables:

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

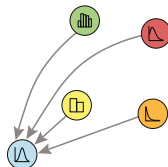
(Yang et al., 2014; Chen, Witten & Shojaie, 2015; Haslbeck & Waldorp, 2017)

Estimating Mixed Graphical Models

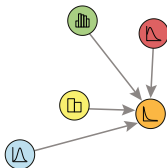
Step 1



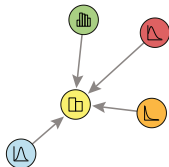
Step 2



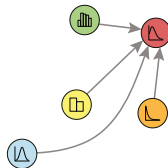
Step 4



Step 3



Step 5



(Meinshausen & Buehlmann, 2006)

ℓ_1 -regularized Estimation

We minimize the negative log-likelihood $F(\mathbf{X}_j, \boldsymbol{\theta}_i)$ together with the ℓ_1 -norm over all parameters:

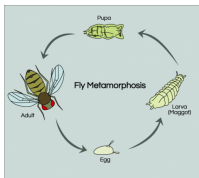
$$\arg_{\boldsymbol{\theta}_i} \min \left\{ \frac{1}{n} \sum_{j=1}^n F(\mathbf{X}_j, \boldsymbol{\theta}_i) + \lambda_i \|\boldsymbol{\theta}_i\|_1 \right\}$$

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

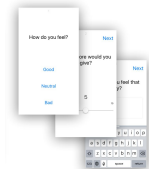
(Hastie, Tibshirani & Wainwright, 2015)

Back to Applications



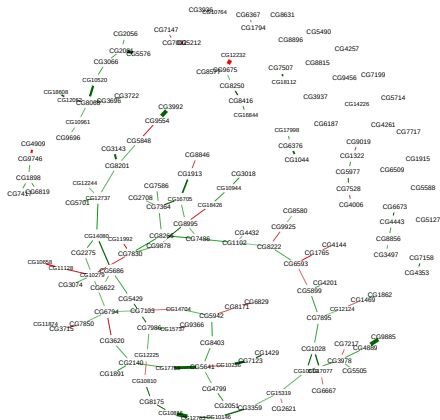
67 measurements of 150 gene expressions related to the immune system of *Drosophila melanogaster* (fruit fly) over its full life cycle

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



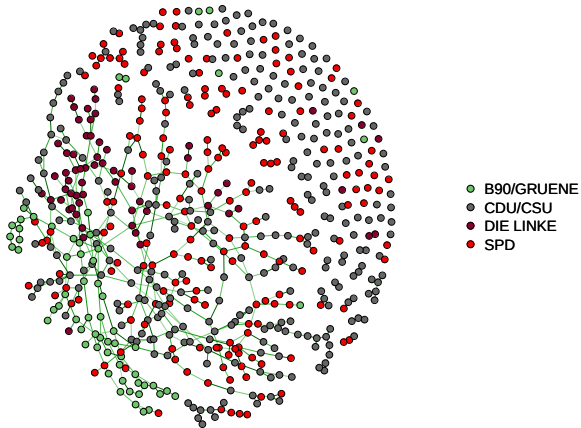
1476 measurements of 14 variables related to mood, activity and social context of one individual over 238 consecutive days

Gene Expressions of Fruit Fly



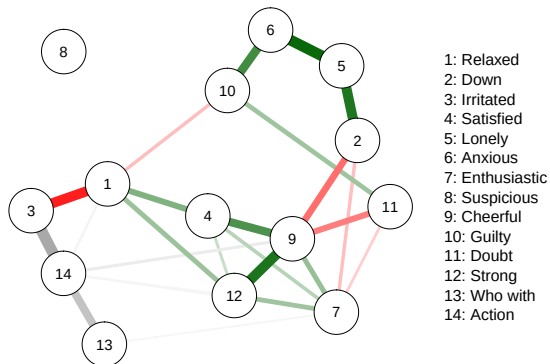
67 Measurements of 150 genes expressions related to immune system of the fruit fly (Lebre et al., 2010)

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:

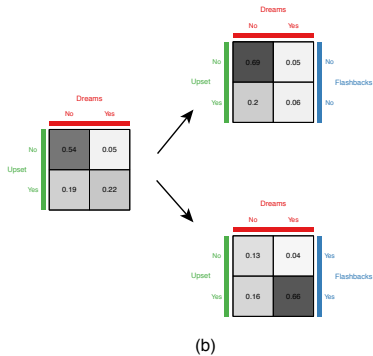
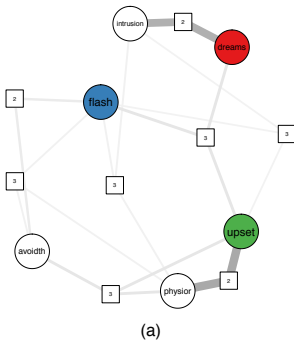
Estimate MGM on Symptom Data

RStudio Server:

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

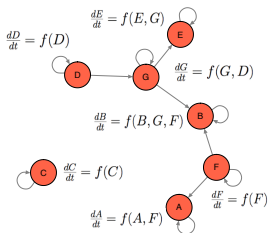
Login: Your UU Solis-ID & password

Interactions of order > 2



Direction of Influence & Interactions as function of time

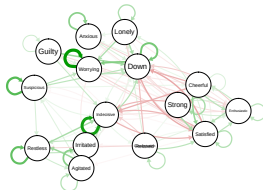
True Structure:



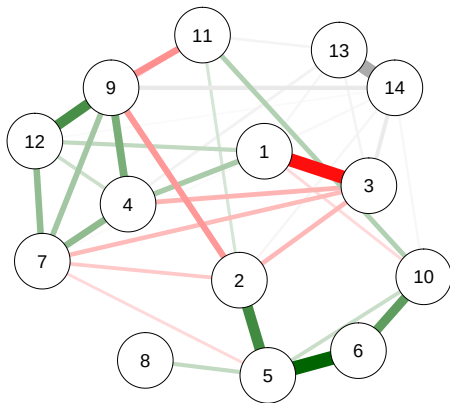
Instantaneous Influence



Influence over time (1h)



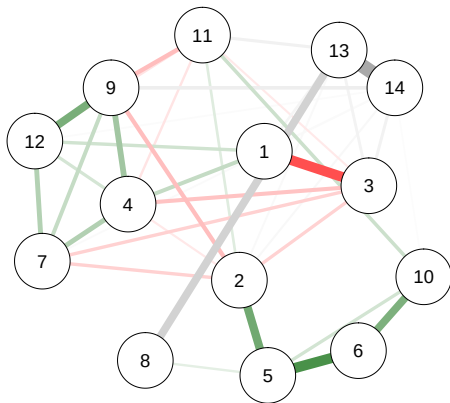
Does the system under investigation change over time?



Time = 0

- 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

Does the system under investigation change over time?



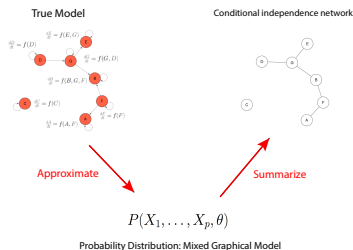
- 1: Relaxed
- 2: Down
- 3: Irritated
- 4: Satisfied
- 5: Lonely
- 6: Anxious
- 7: Enthusiastic
- 8: Suspicious
- 9: Cheerful
- 10: Guilty
- 11: Doubt
- 12: Strong
- 13: Who with
- 14: Action

Time = 0.06

mgm: Summary

mgm package implements:

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



Website: jmbh.github.io (paper and tutorials)

Email: jonashaslbeck@gmail.com

References

- ▶ Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity: the lasso and generalizations. CRC press.
- ▶ 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.