Sparse Graphs
(from noisy data, obviously)

Ernst Wit
e.c.wit@rug.nl

Joint work with
Ivan Vujacic, Fentaw Abegaz and Antonio Abbruzzo

Johann Bernoulli Institute
University of Groningen

February 1, 2013

God is the answer. But what was the question?

A common design:

Many items are screened across few units

- possibly: redundant items
- sometimes: automatically
- possibly: through time.

Possible question

Sparse dependency structure among the items,
... possibly with longitudinal structure.

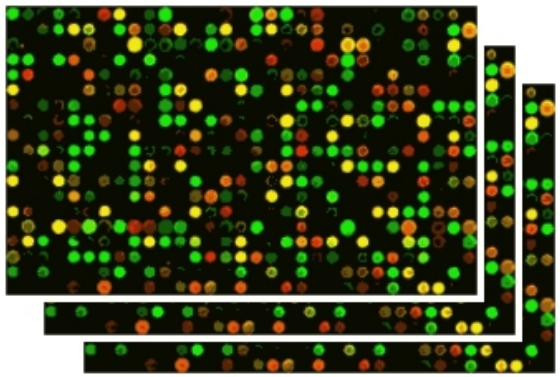
Sparsity:

- Identifiability
- Computational efficiency
- Ontological plausibility
- Epistemic expediency?

rijksuniversiteit groningen

Motivation: Dynamic Genomic Networks

- Transcription: snap-shot of gene activity in time and space (tissue).
- High-throughput microarray and RNA-seq data measure gene activity.



Running example: T-cell time-series dataset.

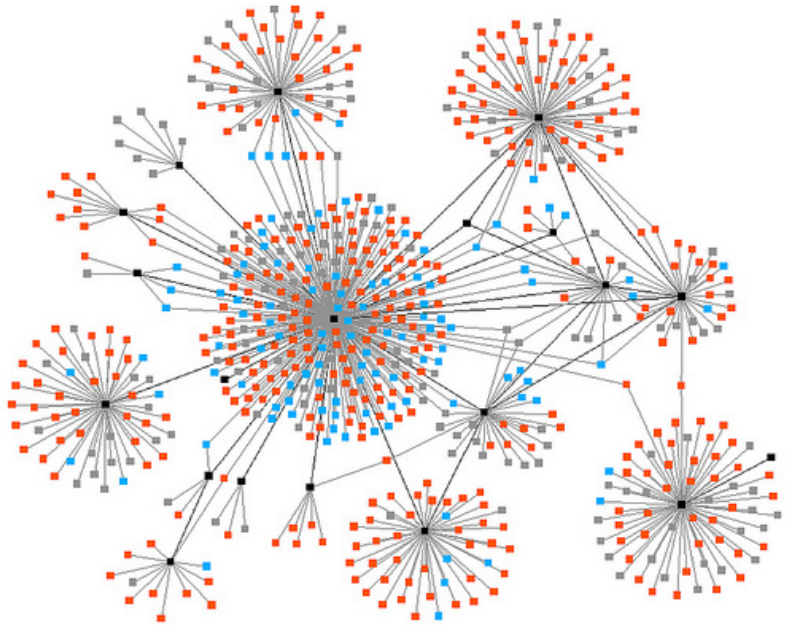
- Temporal expression of 58 genes for 10 unequally spaced time points.
- At each time point there are 44 separate measurements.

Definition (Aim)

Determine dynamic genomic graph G on basis of $\{Y_{gt}^{(i)}\}_{gti}$.

Motivation: Features of a Dynamic Genomic Process

- "**p** >> **n**":
  - Number of observations smaller than number of variables.
  - Thousands of variables and hundreds of observations.
- **Structure:**
  - Highly complex and structured phenomenon.
  - Possibly with additional topographical structure (small world).
- **Sparsity:** only small number of links between nodes.



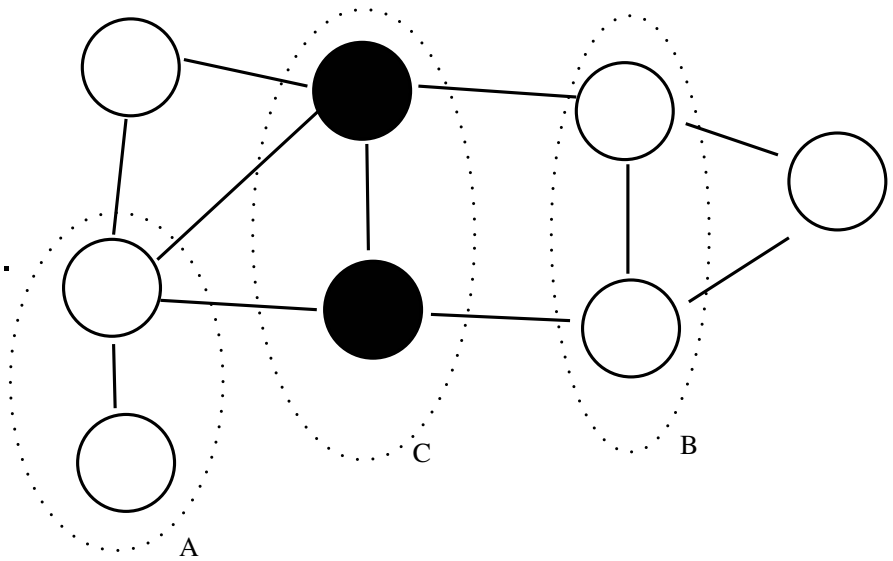
rijksuniversiteit groningen

Undirected Graphical Models

- **Graph**: $G = (V, E)$, $V = \{v_1, \ldots, v_p\}$ and $E \subseteq V \times V$.
- Can be interpreted as a **conditional independence graph**.

Markov Properties:

- Global: $Y_A \perp Y_B | Y_C$.
- Local: $Y_i \perp Y_{V \setminus \{ne(i) \cup i\}} | Y_{ne(i)}$.
- Pairwise $Y_i \perp Y_j | Y_{V \setminus \{i,j\}}$.



Definition (Graphical model)

A couple $M = (G, P)$, where G a graph and P a *probability measure* that satisfies a *Markov property*.

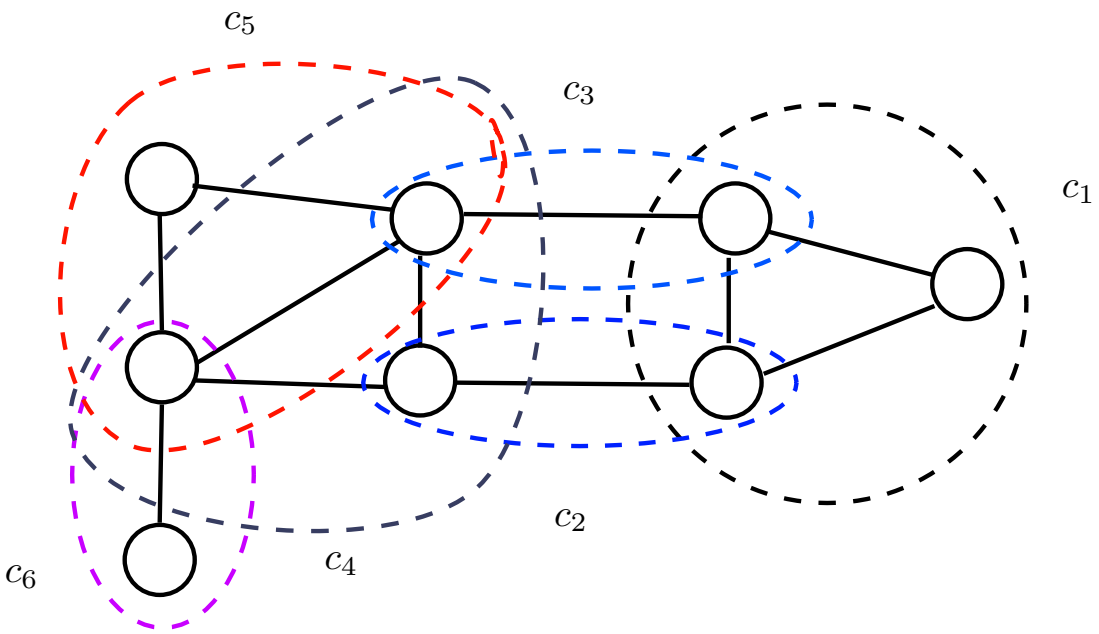
Factorization and Markov properties

Theorem (Hammersley-Clifford)

If the density p=dP is positive and continuous then

$$\textit{Factorization} \iff \textit{Markov}(\textit{Global}/\textit{Local}/\textit{Pairwise})$$

Factorization: $f(\mathbf{y}) = \frac{1}{z}\prod_{c\in C} h_c(\mathbf{y}_c)$, where c are cliques of G.



Example:
Graph (left) contains 6 cliques (complete subgraphs), and so

$$p(\mathbf{y}) = \prod_{i=1}^{6} h_i(y_{c_i}).$$

rijksuniversiteit groningen

Gaussian graphical models

Definition (Gaussian graphical model)

GGM is a graphical model (G, P), where P is determined by:

$$\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Theta}^{-1}).$$

Note:

$$\begin{aligned} p(\mathbf{y}) &\propto \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\Theta(\mathbf{y}-\boldsymbol{\mu})\right) \\ &= \prod_{i=1}^{p}\prod_{j=1}^{p} e^{-(y_i-\mu_i)(y_j-\mu_j)\theta_{ij}/2} \\ &= \prod_{c\in C} h_c(\mathbf{y}_c) \end{aligned}$$

So, if $\theta_{ij} = 0$, then certainly nodes i and j are not in **same** clique:

$$\theta_{ij} = 0 \Leftrightarrow Y_i \perp Y_j | \mathbf{Y}_{V\setminus\{i,j\}}.$$

rijksuniversiteit groningen

Traditional: Hypothesis Test

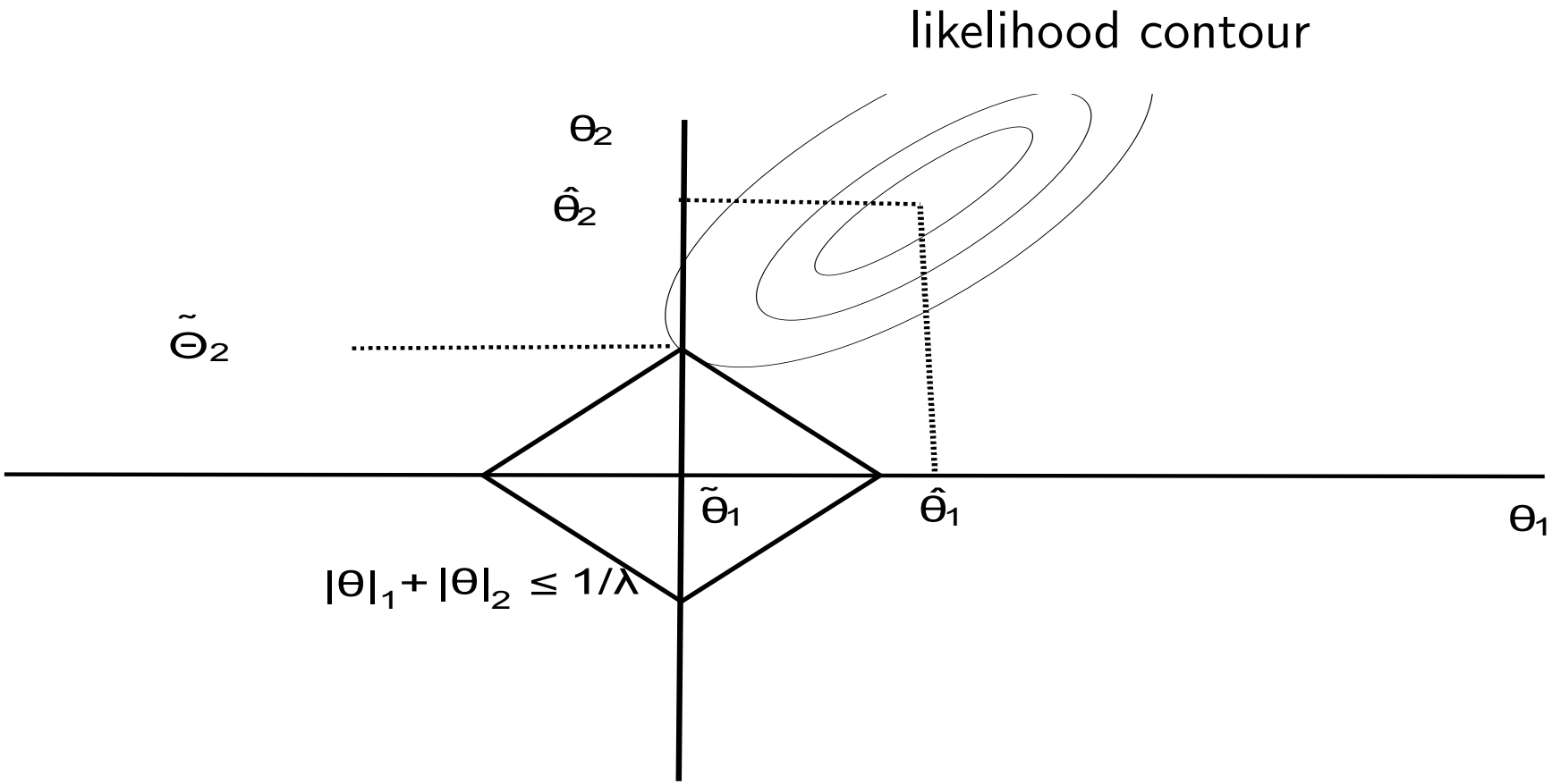
$$H_0 : \theta_{ij} = 0 \iff \text{Edge } (i,j) \text{ is absent}$$
$$H_1 : \theta_{ij} \neq 0 \iff \text{Edge } (i,j) \text{ is present}$$

- Hypothesis tests have poor model selection properties.
- Where do we start to test?
- Multiple testing problem.
- Space of models is of size $2^{\binom{p}{2}}$.
- If $p > n$, then $\mathbf{S}$ is singular.

Penalized Inference

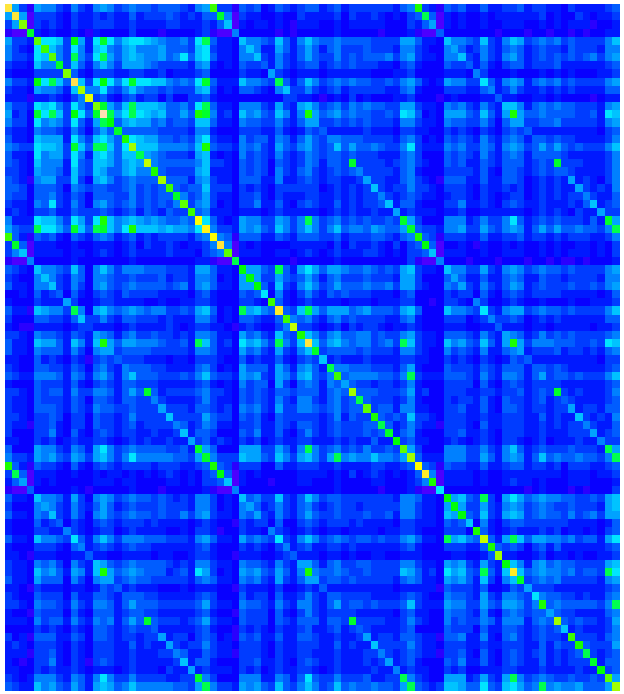
Idea: Maximize the likelihood $l(\mu, \Theta|y)$ **subject to constraint**

$$||\mathbf{\Theta}||_1 = \sum_{i<j} |\theta_{i,j}| \leq 1/\lambda$$



Netherlands Study on Depression and Anxiety (NESDA)

In 3 "waves" 1708 depression patients were scored on 28 items in an IDS questionnaire.

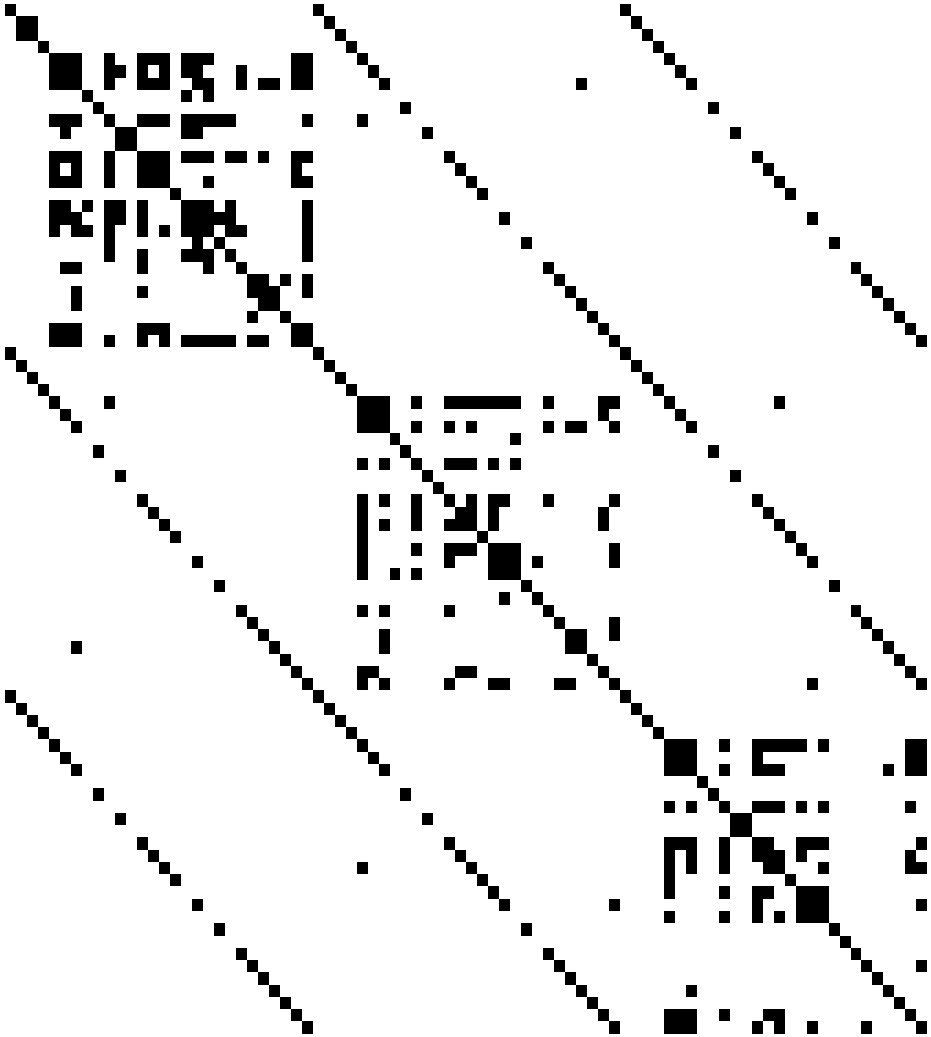


84 $\times$ 84 empirical covariance matrix: notice 3 waves of 28 items

Psychiatrists want to know:

- What is a typical depression "signature"? Are there more than one?
- What is the relationship between the items?

Standard Graphical Lasso: NESDA depression study



Inferred graph for the NESDA study, represented via adjacency matrix

rijksuniversiteit groningen

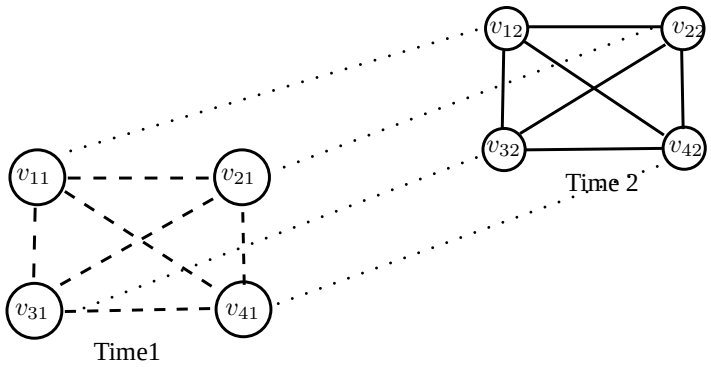
Dynamic Genomic Graphs

- Γ be a set of "genes".
- T be a set of ordered "time points".

Definition (Dynamic genomic graph)

A dynamic genomic graph is a pair $G = (V, E)$.

- Vertices: $V = \{v_{ij}\}_{i \in \Gamma, j \in T}$, where Γ and T are finite sets.
- Links: ordered pair of elements $E \subseteq V \times V$.



rijksuniversiteit groningen

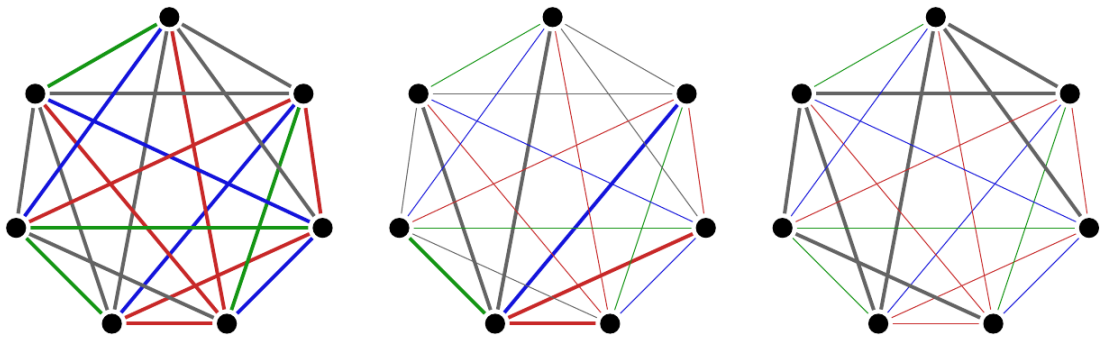
Coloured graphs

Definition (Coloured graph)

A coloured graph is a triplet $G_F = (V, E, F)$, where $G = (V, E)$ is a graph and F is a mapping on the links, i.e.:

$$F : E \to C,$$

where C is a finite set of colours.

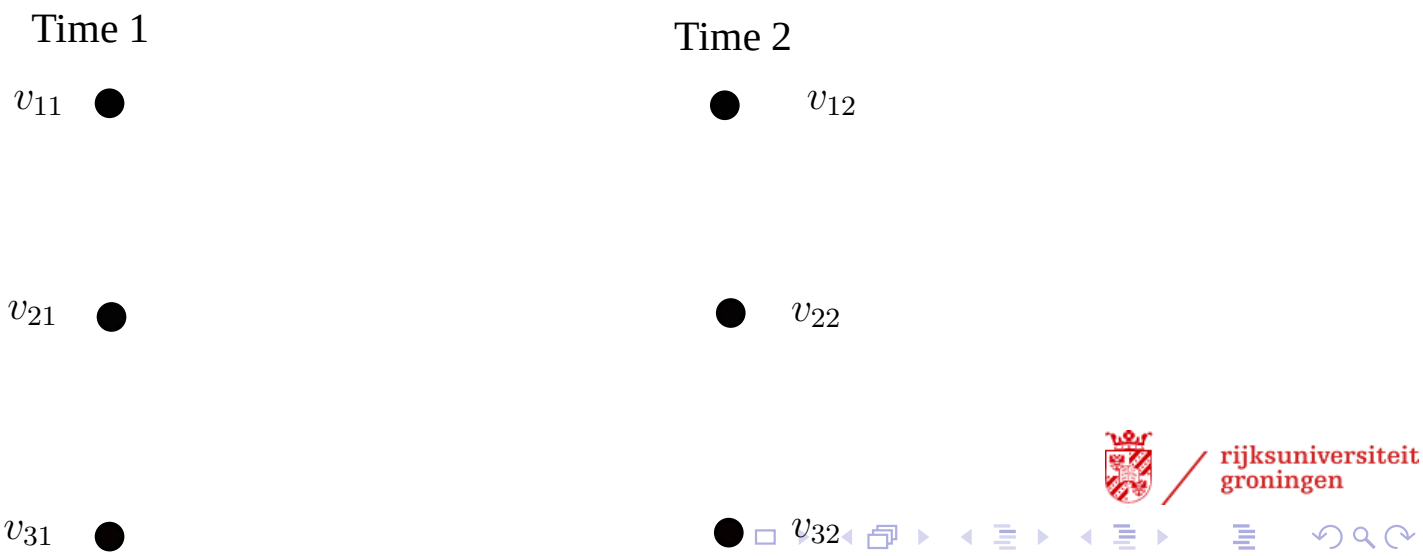


Special kind of coloured graphs: Factorial Graphs

Denote mapping $F : E \to C$ by $E \prec F$.

In analogy with ANOVA, we define the following colouring:

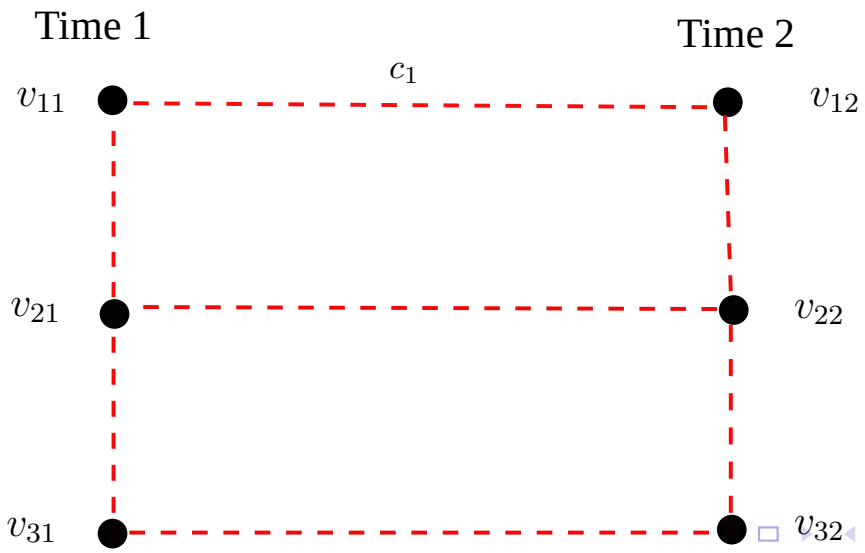
- $\boxed{E \prec 0 \Rightarrow \text{an empty graph.}}$
- $E \prec F_1$.
- $E \prec F_T$.
- $E \prec F_\Gamma$.
- $E \prec F_{\Gamma T}$.



Special kind of coloured graphs: Factorial Graphs

Denote mapping $F : E \to C$ by $E \prec F$.
In analogy with ANOVA, we define the following colouring:

- $E \prec 0 \Rightarrow$ an empty graph.
- $\boxed{E \prec F_1\text{: same colour for all links}}$
- $E \prec F_T$.
- $E \prec F_\Gamma$.
- $E \prec F_{\Gamma T}$.



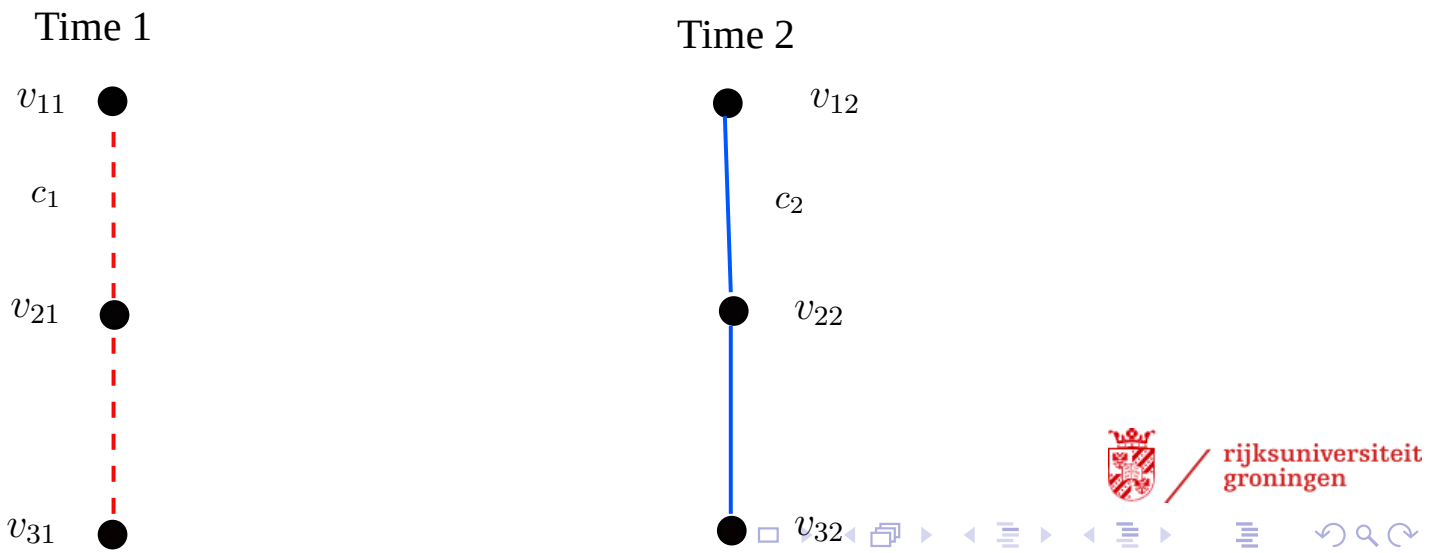
rijksuniversiteit groningen

Special kind of coloured graphs: Factorial Graphs

Denote mapping $F : E \to C$ by $E \prec F$.

In analogy with ANOVA, we define the following colouring:

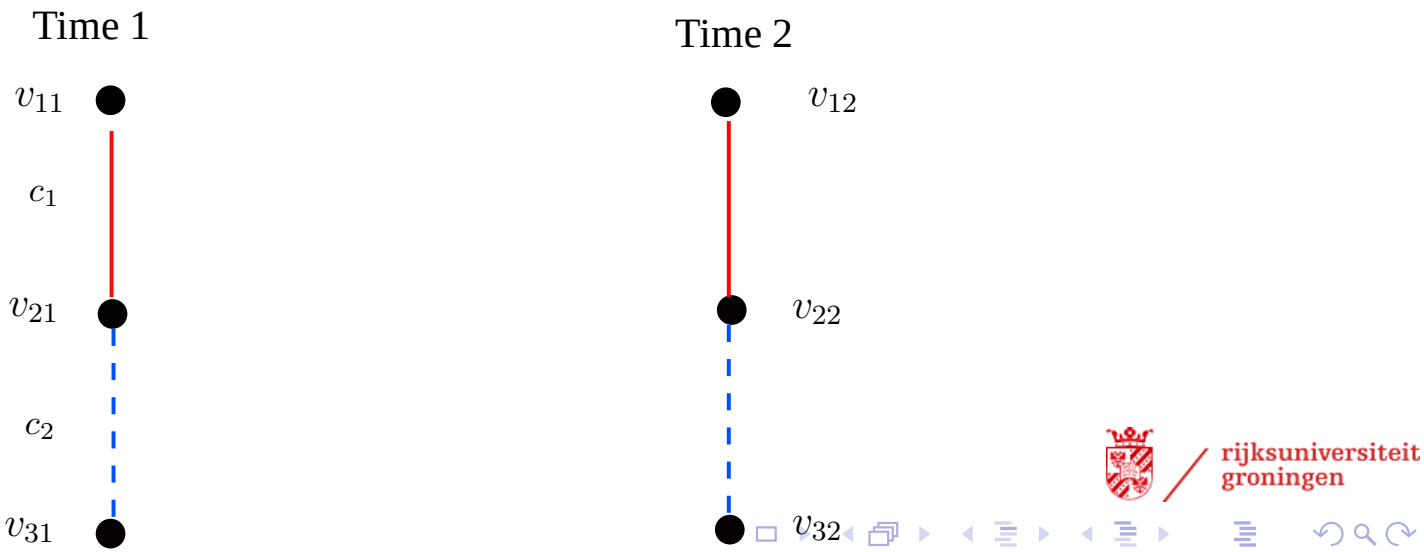
- $E \prec 0 \Rightarrow$ an empty graph.
- $E \prec F_1$.
- $\boxed{E \prec F_T\text{: same colour across all genes}}$
- $E \prec F_\Gamma$.
- $E \prec F_{\Gamma T}$.



Special kind of coloured graphs: Factorial Graphs

Denote mapping $F : E \to C$ by $E \prec F$.
In analogy with ANOVA, we define the following colouring:

- $E \prec 0 \Rightarrow$ an empty graph.
- $E \prec F_1$.
- $E \prec F_T$.
- $\boxed{E \prec F_\Gamma\text{: same colour across all times}}$
- $E \prec F_{\Gamma T}$.



Special kind of coloured graphs: Factorial Graphs

Denote mapping $F : E \to C$ by $E \prec F$.
In analogy with ANOVA, we define the following colouring:

- $E \prec 0 \Rightarrow$ an empty graph.
- $E \prec F_1$.
- $E \prec F_T$.
- $E \prec F_\Gamma$.
- $\boxed{E \prec F_{\Gamma T}\text{: all different colours}}$

Natural Partitions

Definition (Natural partition)

Let $E = \{S_i, N_i\}_{i=0}^{n_T-1}$ be subsets of links where S_i, N_i are defined as follows:

$$S_i = \{\{(v_{jt}, v_{j,t+i}), (v_{j,t+i}, v_{jt})\} | j \in \Gamma, t = 1, \ldots, n_T - i\},$$

and

$$N_i = \{\{(v_{jt}, v_{k,t+i}), (v_{k,t+i}, v_{jt})\} | \forall j \neq k \in \Gamma, t = 1, \ldots, n_T - i\}.$$

The natural partitions imply subgraphs of G and imply partions of $\mathbf{\Theta}$ for GGMs:

$$\mathbf{\Theta} = \left[\begin{array}{cc|cc|cc|cc} S_0 & N_0 & S_1 & N_1 & S_2 & N_2 & \ldots & \ldots \\ & S_0 & N_1 & S_1 & N_2 & S_2 & \ldots & \ldots \\ \hline & & S_0 & N_0 & S_1 & N_1 & S_2 & N_2 \\ & & & S_0 & N_1 & S_1 & N_2 & S_2 \\ \hline & & & & S_0 & N_0 & S_1 & N_1 \\ & & & & & S_0 & N_1 & S_1 \end{array}\right]$$

Factorial Graphical Models

Definition (Gaussian graphical models for factorially coloured graphs)

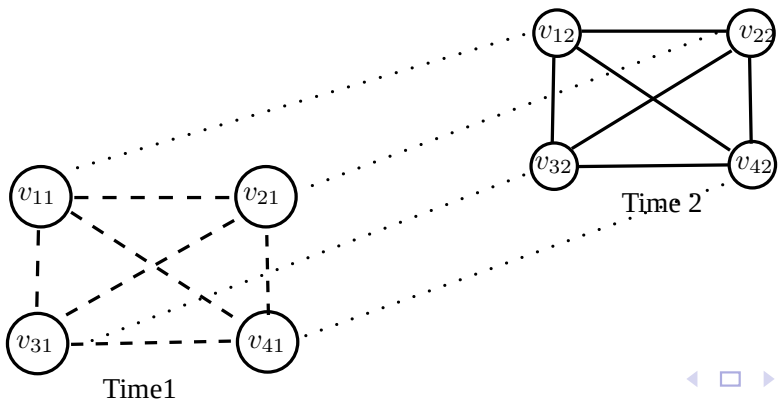
A factorial Gaussian graphical model is a graphical model defined on:

- a dynamic factorial graph $G = (V, E, F)$, where
- a factorial colouring F is applied *separately* to natural partitions

$$S_i \prec F_{S_i}, \quad N_i \prec F_{N_i}, \quad i = 0, \ldots, n_T - 1$$

- which determines Θ in

$$\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Theta}^{-1}).$$



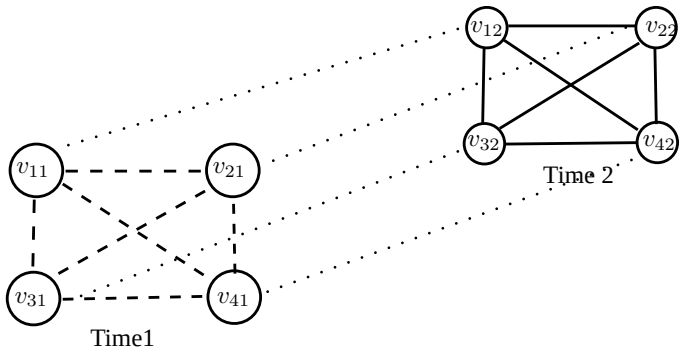
rijksuniversiteit groningen

Example: Factorial Gaussian Graphical Model

Model:

$$(S_0 \prec 1), \quad N_0 \prec F_T, \quad S_1 \prec 1, \quad N_1 \prec 0.$$

Factorial coloured graph:



Precision Matrix:

$$\boldsymbol{\Theta} = \left[\begin{array}{cccc|cccc} \theta_1 & \theta_2 & \theta_2 & \theta_2 & \theta_4 & 0 & 0 & 0 \\ & \theta_1 & \theta_2 & \theta_2 & 0 & \theta_4 & 0 & 0 \\ & & \theta_1 & \theta_2 & 0 & 0 & \theta_4 & 0 \\ & & & \theta_1 & 0 & 0 & 0 & \theta_4 \\ \hline & & & & \theta_1 & \theta_3 & \theta_3 & \theta_3 \\ & & & & & \theta_1 & \theta_3 & \theta_3 \\ & & & & & & \theta_1 & \theta_3 \\ & & & & & & & \theta_1 \end{array}\right]$$

rijksuniversiteit groningen

Penalized likelihood for GGMs

- Consider an experiment: $|\Gamma|$ genes measured across $|T|$ time points.
- Assume n iid samples $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)}$, where $\mathbf{y}^{(i)} = (y_1^{(i)}, \ldots, y_{\Gamma T}^{(i)})$.
- Assume $\mathbf{Y}^{(i)} \sim N(\mathbf{0}, \boldsymbol{\Theta}^{-1})$, then

Likelihood:

$$l(\boldsymbol{\Theta}|\mathbf{y}) \propto \log(\boldsymbol{\Theta}) - \mathrm{tr}(\mathbf{S}\boldsymbol{\Theta}).$$

AIM: Optimization of penalized likelihood:

$$\hat{\boldsymbol{\Theta}} := \mathrm{argmax}_{\boldsymbol{\Theta}}\{l(\boldsymbol{\Theta}|\mathbf{y})\}$$

subject to

- $\boldsymbol{\Theta} \succeq 0$;
- $||\boldsymbol{\Theta}||_1 \leq 1/\lambda$;
- some factorial colouring F.

Inferring penalized factorial Gaussian graphical models

LogdetPPA. Newton-CG primal proximal point algorithm (*Wang et al., 2010*, including Kim Toh and Defeng Sun) is used to solve optimization:

$$\hat{\boldsymbol{\Theta}} := \underset{\boldsymbol{\Theta}}{\operatorname{argmin}} -\{\log|\boldsymbol{\Theta}| - \operatorname{tr}(\boldsymbol{\Theta}\mathbf{S}) + \boldsymbol{\lambda}'\boldsymbol{\theta}^{+} + \boldsymbol{\lambda}'\boldsymbol{\theta}^{-} : A(\boldsymbol{\Theta}) = \mathbf{0},$$

$$B(\boldsymbol{\Theta}) - \boldsymbol{\theta}^{+} + \boldsymbol{\theta}^{-} = \mathbf{0}, \boldsymbol{\Theta} \succeq 0, \boldsymbol{\theta}^{+}, \boldsymbol{\theta}^{-} \geq 0\}$$

- $A(\boldsymbol{\Theta})$: linear constraints which depend on coloured graph.
- $B(\boldsymbol{\Theta})$: ℓ_1-norm penalty on elements of precision matrix.
- $\boldsymbol{\theta}^{+}$ and $\boldsymbol{\theta}^{-}$ are additional variables (slack variables).
- $\boldsymbol{\Theta} \succeq 0$: semi-positive definite constraint.

Solves $\hat{\Theta}$ up to 2000×2000 .

rijksuniversiteit groningen

T-cell data

Aim. Use large time-course experiment to characterize response of human T-cell line (Jurkat) to PMA and ionomycin treatment.

T-cell time-series dataset

- Temporal expression of 58 genes for 10 unequally spaced time points.
- At each time point there are 44 separate measurements.
- See Rangel et al. (2004) for more details.

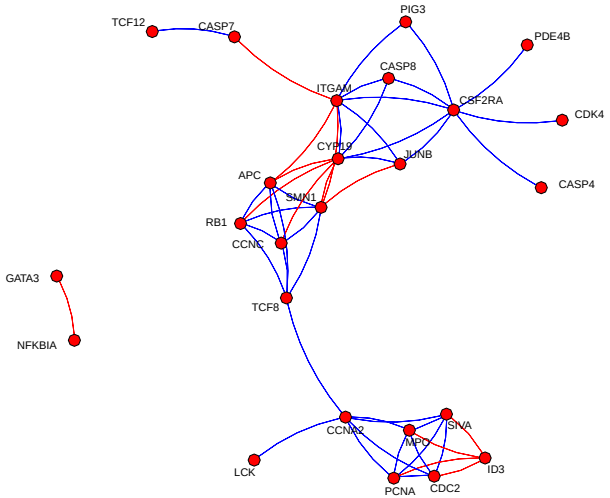
Application to T-cell data

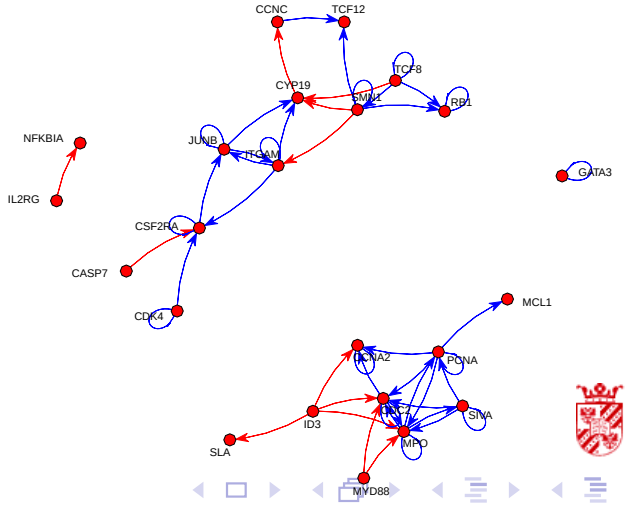
$$S_0 \prec F_1, \mathbf{N_0} \prec \mathbf{F_\Gamma}, S_1 \prec F_T, \mathbf{N_1} \prec \mathbf{F_\Gamma}$$

$$\mathbf{\Theta} = \left[\begin{array}{cc|cc|cc|cc} S_0^1 & N_0^1 & S_1^1 & N_1^1 & 0 & 0 & \dots & \dots \\ & S_0^1 & N_1^1 & S_1^1 & 0 & 0 & \dots & \dots \\ \hline & & S_0^2 & N_0^2 & S_1^2 & N_1^2 & 0 & 0 \\ & & & S_0^2 & N_1^2 & S_1^2 & 0 & 0 \\ \hline & & & & S_0^3 & N_0^3 & S_1^3 & N_1^3 \\ & & & & & S_0^3 & N_1^3 & S_1^3 \end{array}\right]$$

$N_0^1 = N_0^2 = \ldots = N_0^{10}$

$N_1^1 = N_1^2 = \ldots = N_1^{10}$





rijksuniversiteit groningen

What have we achieved so far? And problems!

Summary:

- Penalized Gaussian graphical models
- Coloured graphs
- Natural partitions
- Factorially coloured Gaussian graphical models

Problems:

1. Factorial colouring not particularly flexible in modeling time dynamics.
2. Gaussian assumption may be too restrictive for realistic genomic data.
3. How to choose the appropriate level of penalization?

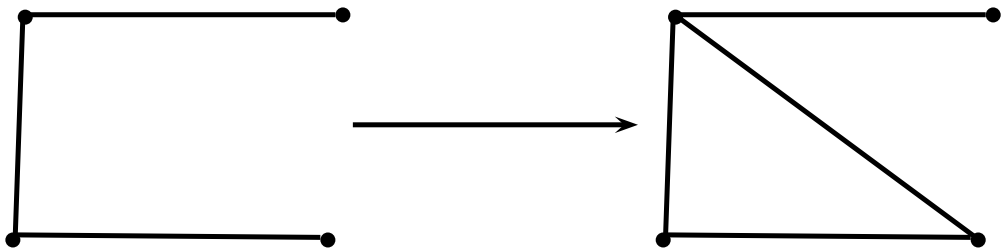
rijksuniversiteit groningen

1. Extension: Slowly changing graphical models

- Problem: Estimate changes in the dynamic of the network.

Main Idea: Penalize changes between graphs at different time points

$$||\Delta\Theta||_1 = \sum_{s=0}^{t-1}\sum_{k=0}^{t-1} ||N_s^k - N_s^{k+1}||_1.$$



- Solution: Penalized maximum likelihood subject to constraints.

rijksuniversiteit groningen

Implement Slowly Changing via logdetPPA

$$\hat{\boldsymbol{\Theta}} := \underset{\boldsymbol{\Theta}}{\operatorname{argmin}} \{ -\log |\boldsymbol{\Theta}| + \operatorname{tr}(\mathbf{S}\boldsymbol{\Theta}) + \lambda_1 \mathbf{x}^+ + \lambda_1 \mathbf{x}^- + \lambda_2 \mathbf{y}^+ + \lambda_2 \mathbf{y}^- \}$$

$$\begin{aligned} \text{subject to} \quad & \mathbf{A}(\boldsymbol{\Theta}) - \mathbf{y}^+ + \mathbf{y}^- = \mathbf{0} \\ & \mathbf{B}(\boldsymbol{\Theta}) - \mathbf{x}^+ + \mathbf{x}^- = \mathbf{0} \\ & \mathbf{C}(\boldsymbol{\Theta}) = \mathbf{0} \\ & \boldsymbol{\Theta} \succeq 0, \mathbf{x}^+, \mathbf{x}^-, \mathbf{y}^+, \mathbf{y}^- \geq \mathbf{0}. \end{aligned}$$

- $A(\boldsymbol{\Theta})$: set of linear constraints which depends on the $||\Delta\boldsymbol{\Theta}||_1$.
- $B(\boldsymbol{\Theta})$ ℓ_1-norm penalty on element of precision matrix.
- $C(\boldsymbol{\Theta})$ zero elements a priori.
- $\mathbf{x}^+$, $\mathbf{x}^-$, $\mathbf{y}^-$ and $\mathbf{y}^+$ are slack variables.

rijksuniversiteit groningen

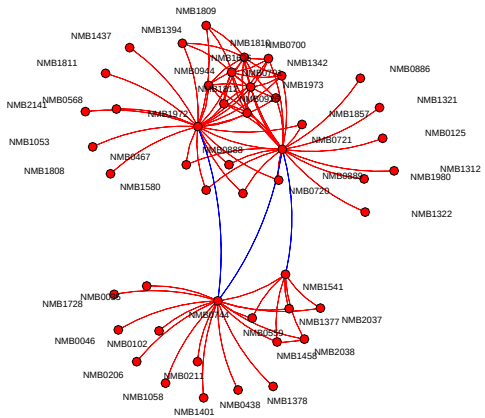
Application to a time-course dataset

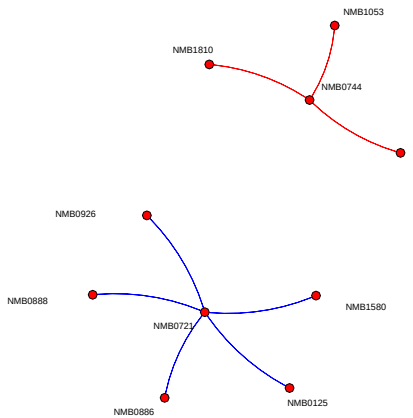
$$\mathbf{S_0} \prec F_{\Gamma T}, \mathbf{N_0} \prec 1, \mathbf{N_1} \prec 1, \mathbf{N_2} \prec 0$$

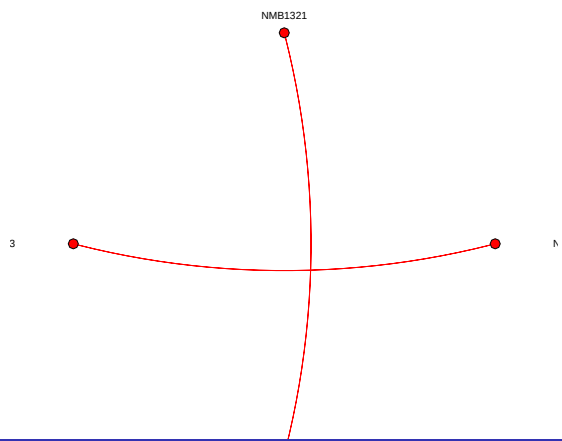
$$\Theta = \left[\begin{array}{cc|cc|cc|cc} S_0^1 & N_0^1 & S_1^1 & N_1^1 & 0 & 0 & \cdots & \cdots \\ & S_0^1 & N_1^1 & S_1^1 & 0 & 0 & \cdots & \cdots \\ \hline & & S_0^2 & N_0^2 & S_1^2 & N_1^2 & 0 & 0 \\ & & & S_0^2 & N_1^2 & S_1^2 & 0 & 0 \\ \hline & & & & S_0^3 & N_0^3 & S_1^3 & N_1^3 \\ & & & & & S_0^3 & N_1^3 & S_1^3 \end{array}\right]$$

N_0^1 $\qquad$ $|N_0^1| - |N_0^2|$ $\qquad$ $|N_0^2| - |N_0^3|$

N0 at time 1 $\qquad$ 1 $\qquad$ 7

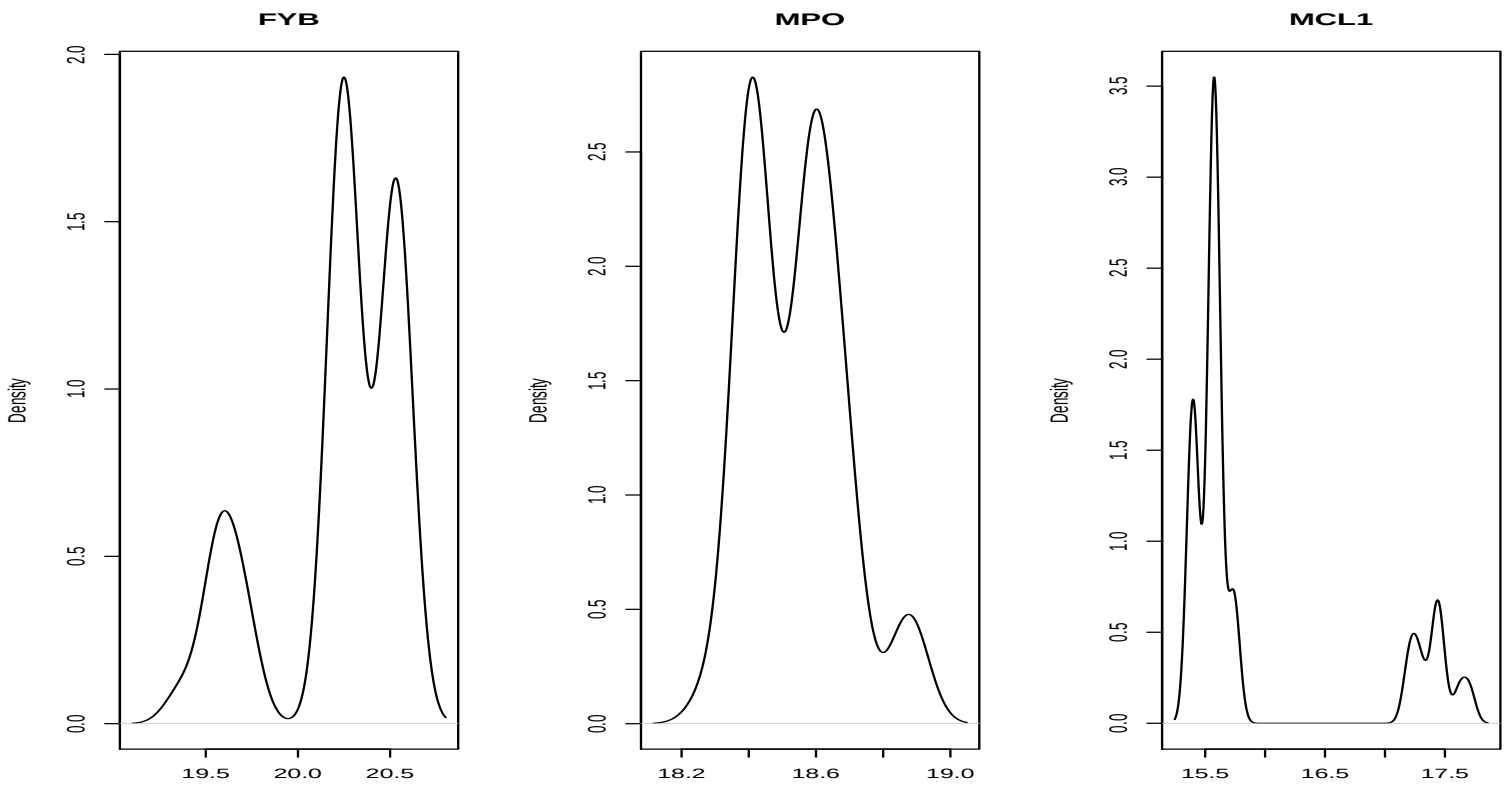






2. Extension: Non-Normality

- Problem: Non-Normality of the data (e.g. T-cell).



- Solution: Copula Gaussian graphical models

rijksuniversiteit groningen

Non-canonical Gaussian copula

- **Sklar's theorem.** Every joint multivariate distribution can be represented by a copula and its marginals.
- We introduced a non-canonical Gaussian copula.

Definition (Non-canonical Gaussian copula)

A non-canonical Gaussian copula with matrix $\mathbf{\Theta}$ and vector $\boldsymbol{\sigma}$ is:

$$C_{\mathbf{\Theta}}(u_1, \ldots, u_p) = \mathbf{\Phi}_{\mathbf{\Theta}}(\Phi_1^{-1}(u_1), \ldots, \Phi_p^{-1}(u_p)),$$

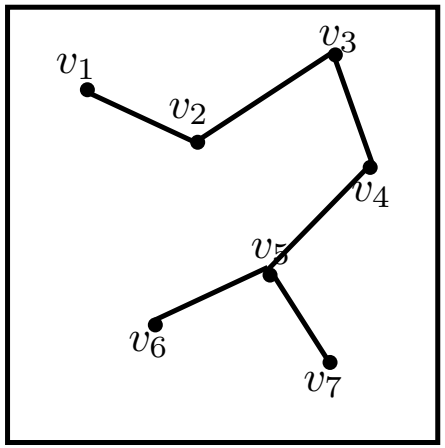
where,

- $\Phi_i(\cdot)$ represents cdf of univariate $N(0, \sigma_i^2)$ distribution,
- $\mathbf{\Phi}_{\mathbf{\Theta}}(\cdot)$ is cdf of multivariate $N(\mathbf{0}, \mathbf{\Theta}^{-1})$ with $\forall i : \theta_{ii} = 1$.

Copula Gaussian graphical models

IDEA:

- Graph exists on a hidden Gaussian variable $\mathbf{Z} \sim N(0, \Theta^{-1})$,
- $\mathbf{Z}$ gives rise to observed non-Gaussian data $\mathbf{Y}$.



Latent Variable: $\mathbf{Z} \sim N(\mu, \mathbf{\Theta^{-1}})$

$$F_1, F_2, \vdots, F_7$$

Marginal Distribution: $\mathbf{F}$

$$Y_1 = F_1^{-1}(\phi(Z_1))$$
$$Y_2 = F_2^{-1}(\phi(Z_2))$$
$$\vdots$$
$$Y_7 = F_p^{-1}(\phi(Z_7))$$

Observed: $\mathbf{Y}$

- We consider the F_i as nuisance parameters.
- For continuous variables: 1-to-1 relationship between Z and Y.
  For discrete variables, relationship is more complicated!

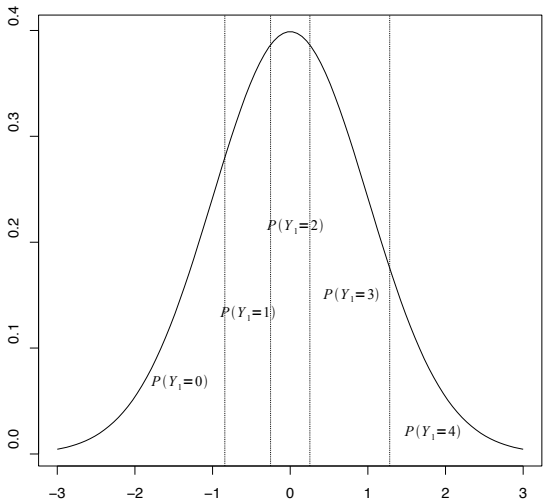
Gaussian Copula for Ordinal Variables

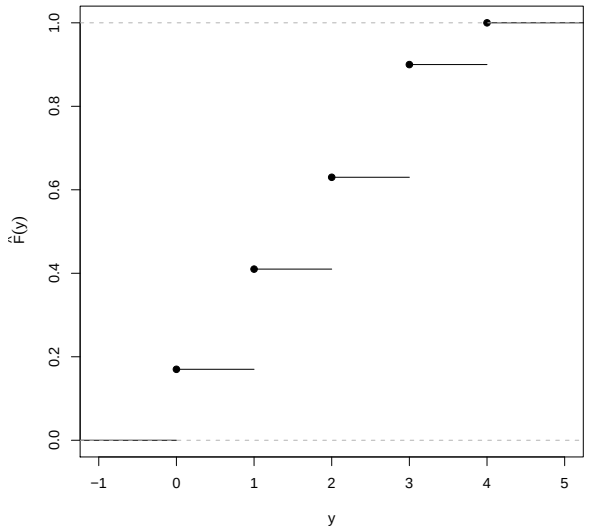
In NESDA depression example $Y = (Y_1, \ldots, Y_p)$ consist of ordinal variables:

$$Y_i \in \{1, 2, 3, 4\}, \quad \text{severity score for item } i$$

Assume existence latent variable $\mathbf{Z} = (Z_1, \ldots, Z_p) \sim N(\mathbf{0}, \mathbf{\Theta}^{-1})$, such that

$$Y_i^{(k)} < Y_i^{(l)} \iff Z_i^{(k)} < Z_i^{(l)}.$$





Values of z_i that correspond to y_i,

$$\mathbf{D}(y_i) = \{\mathbf{z} \in \mathbb{R} \mid F_i^{-1}(\Phi_i(z)) = y_i\}$$

rijksuniversiteit groningen

Maximum Likelihood Estimation

- Gaussian copula density function is:

$$\phi_{\mathbf{\Theta}}(\mathbf{u}) = |\mathbf{\Theta}|^{1/2} exp(-\frac{1}{2}\mathbf{\Phi}^{-1}(\mathbf{u})'(\mathbf{\Theta} - \mathbf{I})\mathbf{\Phi}^{-1}(\mathbf{u})),$$

$$\mathbf{u} = \mathbf{F}(\mathbf{z}) = (F_1(z_1), \ldots, F_p(z_p))$$

- The profile marginal likelihood is:

$$L(\mathbf{\Theta}|\mathbf{D}, \hat{F}_1, \ldots, \hat{F}_p) = \int_{\mathbf{z}\in\mathbf{D}} \phi_{\mathbf{\Theta}}(\mathbf{z}) d\mathbf{z}$$

Problem: Marginal likelihood is hard.

Solution: The iterative EM Algorithm

- **E-Step**:
  Compute the quantity $Q(\mathbf{\Theta}|\mathbf{\Theta}^{(m)}) = \mathbb{E}_{\mathbf{Z}}[l(\mathbf{\Theta})|\mathbf{Y}, \mathbf{\Theta}^{(m)}]$

$$Q(\mathbf{\Theta} \mid \mathbf{\Theta}^{(m)}) \quad = \quad \frac{n}{2}\left\{\log|\mathbf{\Theta}| - \mathrm{tr}\left(\mathbf{\Theta}\bar{\mathbf{R}}\right) + \mathrm{tr}\left(\bar{\mathbf{R}}\right)\right\},$$

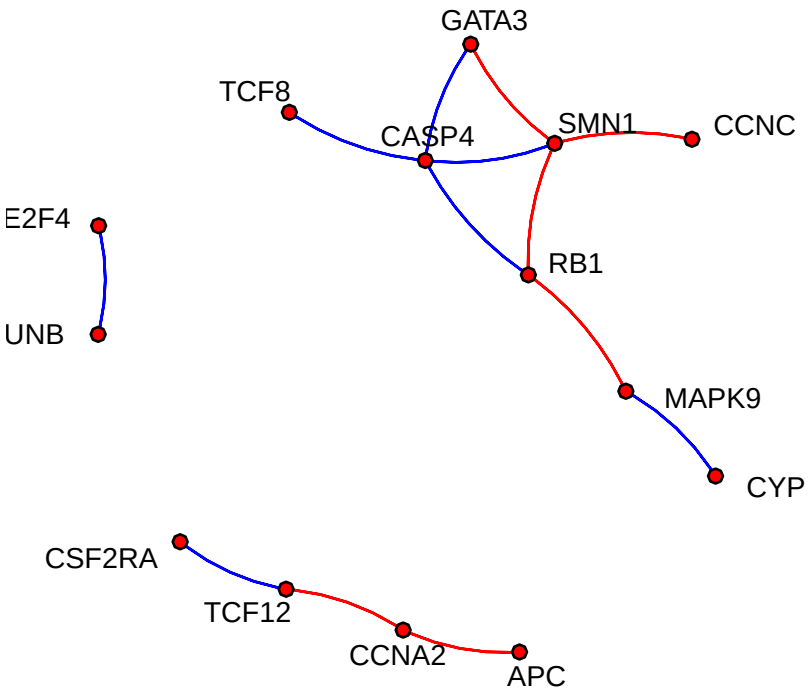
  where

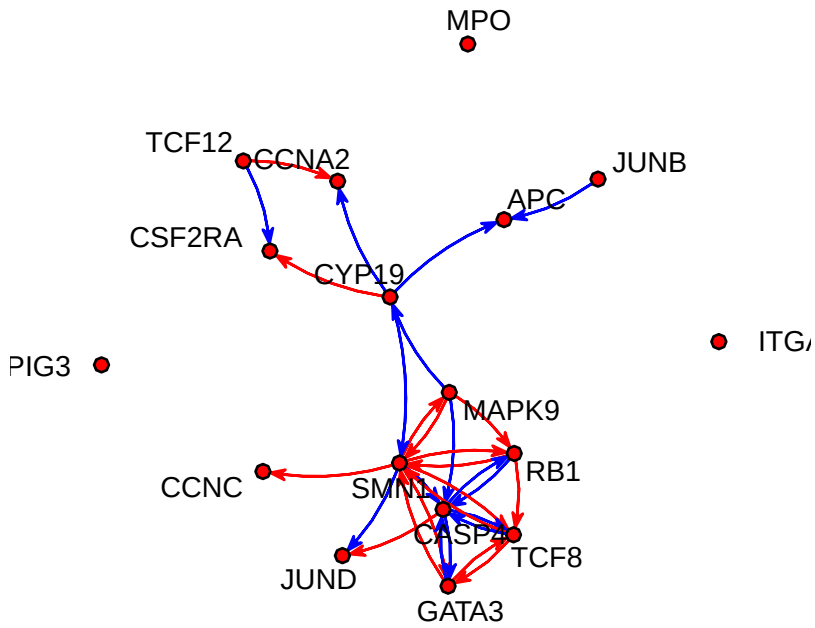
$$\bar{\mathbf{R}} = \mathbb{E}\left[\mathbf{Z}_i\mathbf{Z}_i' \mid \mathbf{Z}_i \in \mathbf{D}(Y_i), \mathbf{\Theta}^{(m)}\right]$$

  can be found explicitly via result in Wilhem. and Manjunath (2010).
- **M-Step**:
  Maximize $Q(\mathbf{\Theta}|\mathbf{\Theta}^{(m)})$ subject to $||\mathbf{\Theta}||_1 \leq 1/\lambda$ and colouring F.

Application of Gaussian copula graphical models to T-cell

$$S_0 \prec F_1, \mathbf{N_0} \prec \mathbf{F_\Gamma}, S_1 \prec F_T, \mathbf{N_1} \prec \mathbf{F_\Gamma}$$





$N_0^1 = N_0^2 = \ldots = N_0^{10}$

$N_1^1 = N_1^2 = \ldots N_1^{10}$

3. Appropriate level of penalization

Model selection

The penalty parameter λ determines the amount of sparsity in the model.

- Consistency for $\lambda = O(\sqrt{log(p)/n})$, but what for small n?
- AIC or BIC are not appropriate for *penalized* likelihoods, because

$$i(\hat{\Theta}_\lambda) \neq 0.$$

Steps for finding the appropriate level of sparsity?

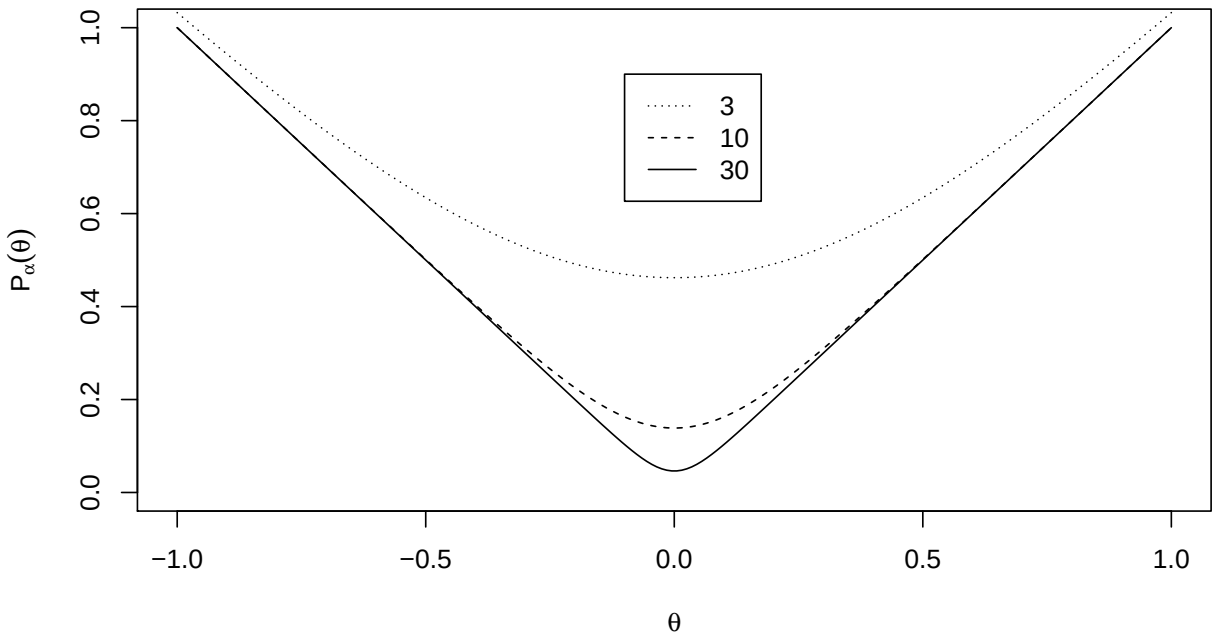
1. We define an approximate problem (AP);
2. We define an information criterion of the AP;
3. We show that limit of solution of AP is meaningful for our problem

Approximated problem: M-estimation

Let $\hat{\Theta}_{\lambda,\alpha}$ be the maximizer of the penalized likelihood

$$l_{\lambda,\alpha}(\Theta) = \sum_{k=1}^{n} l_k(\Theta) - n\lambda P_\alpha(\Theta),$$

where $P_\alpha(\Theta) = \sum_{ij}[\log\{1+\exp(-\alpha\theta_{ij})\} + \log\{1+\exp(\alpha\theta_{ij})\}]/\alpha$.



$\hat{\Theta}_{\lambda,\alpha}$ is an M-estimator of ψ-type: $\sum_{k=1}^{n} l'_{y_k}(\Theta) - \lambda P'_\alpha(\Theta) = 0$.

rijksuniversiteit groningen

GIC for Approximated Problem

Konishi and Kitagawa (2008) defined the GIC for M-estimators:

$$GIC(\hat{\Theta}_{\lambda,\alpha}) = -n\{\log|\hat{\Theta}_{\lambda,\alpha}| - \mathrm{Tr}(\hat{\Theta}_{\lambda,\alpha}S)\} + 2\mathrm{Tr}\{R^{-1}(\alpha)Q\},$$

where $R(\alpha)$ is square matrix of order p^2

$$R(\alpha) = \frac{1}{2}\hat{\Theta}^{-1}_{\lambda,\alpha} \otimes \hat{\Theta}^{-1}_{\lambda,\alpha} + \lambda P''_{\alpha}(\hat{\Theta}_{\lambda,\alpha}), \tag{1}$$

and $P''_{\alpha}(\hat{q}_{\lambda,\alpha})$ diagonal matrix with p^2 elements $(|\hat{\theta}_{11}|''_{\alpha}, \hat{\theta}_{12}|''_{\alpha}, \ldots, |\hat{\theta}_{pp}|''_{\alpha})$ and

$$\begin{aligned} Q(\alpha) &= \frac{1}{4}\mathrm{vect}S\mathrm{vect}S^T - \frac{1}{4n}\sum_{k=1}^{n}\mathrm{vect}_k\mathrm{vect}_k^T \\ &= Q \end{aligned}$$

"GIC" = Limit of AP GIC

L_1 penalized estmimators are no M-estimators, so GIC is not defined.

We define the GIC as α-limit of approximated one.

Definition ("GIC" for L_1 penalized graphical models.)

$$\mathrm{GIC}(\lambda) = -n\{\log|\hat{\Theta}_\lambda| - \mathrm{Tr}(\hat{\Theta}_\lambda S)\} + 2\mathrm{Tr}(R^{-1}Q),$$

Matrix Q is square matrix of order p^2 (as above) and

$$R[I, I] = \hat{\Theta}_\lambda \otimes \hat{\Theta}_\lambda[I, I]^{-1},$$

where I are the non-zero elements in $\mathrm{vect}\hat{\Theta}_\lambda$.

rijksuniversiteit groningen

Performance of the GIC: simulation study

		TP	Sens.	TN	Spec.
N = 8	GIC	13.0	0.531	51.0	0.543
	AIC	0.0	0.000	94.0	1.000
	BIC	0.0	0.000	94.0	1.000
N = 12	GIC	16.0	0.667	45.0	0.479
	AIC	2.0	0.077	93.0	0.989
	BIC	0.5	0.019	94.0	1.000

Results p = 16; Present Links = 26; Absent Links=102

However,

$$\lim_{N\to\infty} GIC = \lim_{N\to\infty} AIC,$$

so GIC is not edge-selection consistent.

rijksuniversiteit groningen

Conclusion and Discussion

Summary:

- Factorial graphical models based on coloured graphs.
- Slowly changing graphical models.
- Extension: Copula Gaussian graphical models.
- How do we evaluate model complexity?

Questions:

- How would we adapt model with other factors?
- Could we combine Copulas with slowly changing networks?
- How to introduce latent structures?