Multi-Way, Multi-View Learning

Ilkka Huopaniemi[1] Tommi Suvitaival[1] Janne Nikkilä[1,2]
Matej Orešič[3] Samuel Kaski[1]

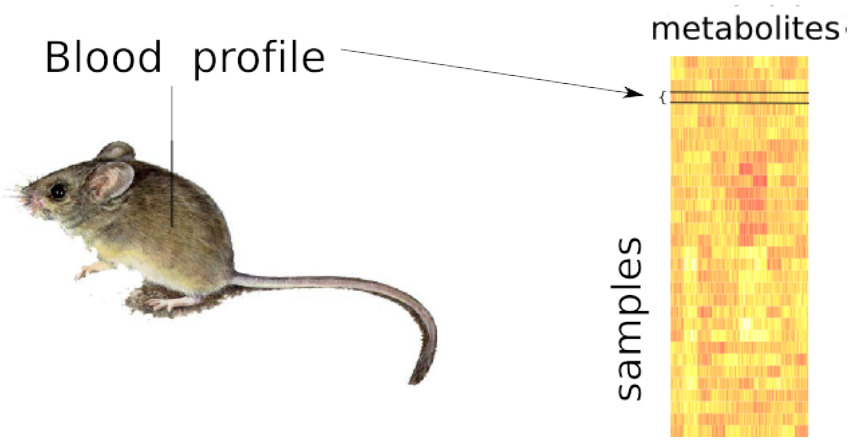
1 Department of Information and Computer Science, Helsinki University of Technology, Finland
2 Department of Basic Veterinary Sciences, (Division of Microbiology and Epidemiology), Faculty of Veterinary Medicine, University of Helsinki, Finland
3 VTT Technical Research Centre of Finland, Espoo, Finland

December 12, 2009

TEKNILLINEN KORKEAKOULU
HELSINKI UNIVERSITY OF TECHNOLOGY

Metabolomics data

- Metabolomics: Diseases affect the concentrations of metabolites = chemical compounds
- Concentrations of > 100 metabolites (Mass Spectrometry)
- Covariates, find biomarkers for disease



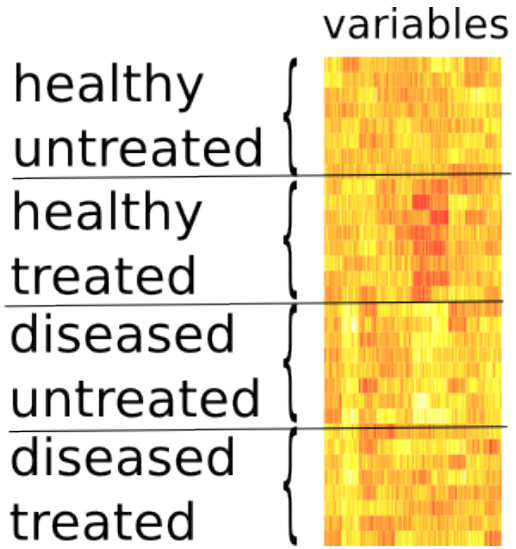
- Find features with differential concentration, confidence intervals

Multi-way experimental setups

- Current approach **One-way methods**: diseased-healthy comparison, only find biomarkers for disease
- **Additional covariates**: Medical treatment, gender, time-series, **multi-way**
  - ANOVA, linear model, MANOVA

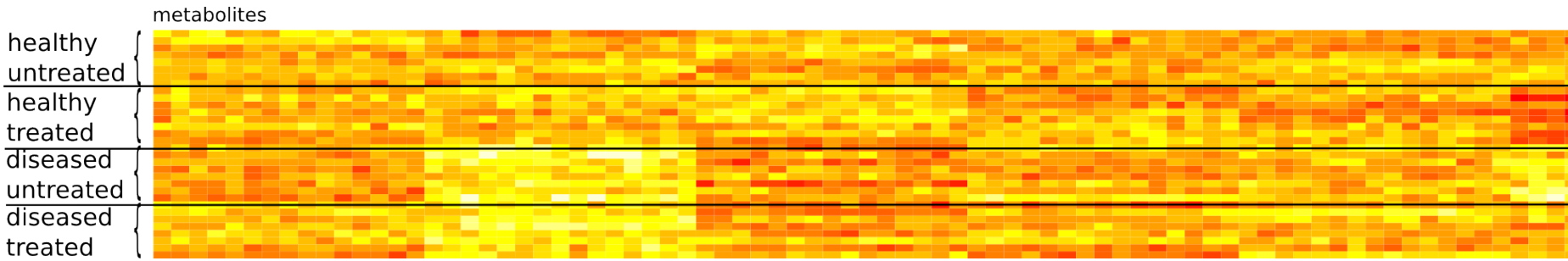
$$\mathbf{x} = \mu^x + \alpha_a^x + \beta_b^x + (\alpha\beta)_{ab}^x + \text{noise}.$$

  - μ^x grand mean
  - α_a^x, β_b^x main effects
  - $(\alpha\beta)_{ab}^x$ **Interaction effect** of covariates



Small sample-size problem

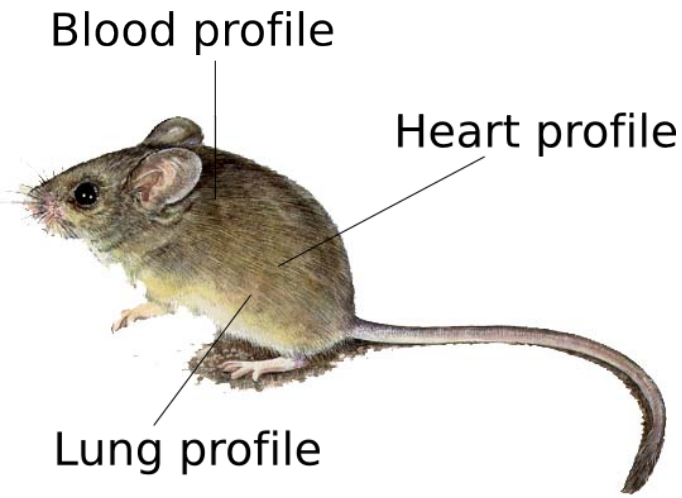
- High dimensionality
- Small number of samples (economical & ethical cost)
- Standard approaches do not work
- Uncertainty estimation crucial



- Strong dimensionality reduction
- Use prior knowledge: Metabolites form similarly behaving groups

Biological multi-source experiments

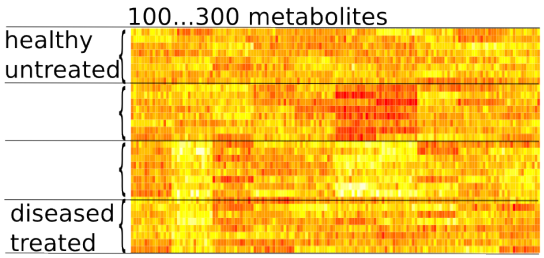
Metabolic profiles from several tissues



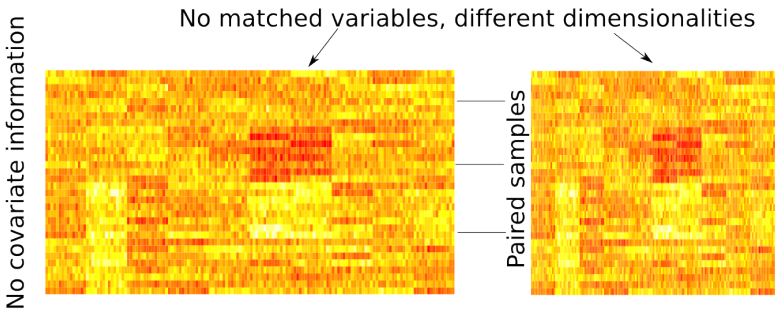
- Different metabolites in different tissues
- Same individual (pairing)
- Are there disease-related dependencies between the tissues?

Multi-Way, Multi-Source setups

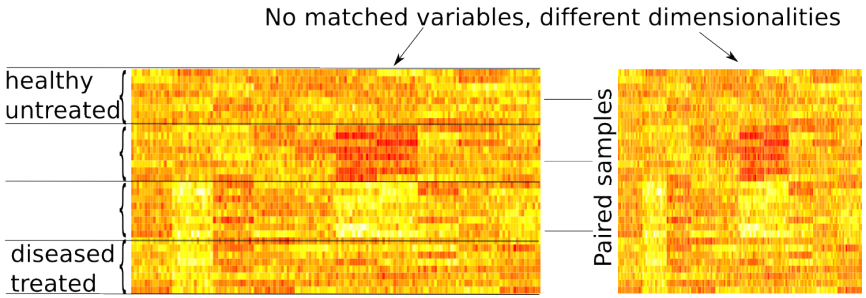
Multi-way analysis (ANOVA)



Multi-source analysis (CCA)



Multi-source, multi-way analysis



Multi-View Learning with multiple **independent** covariates

Decomposition into **shared** and **source-specific** covariate effects

IDEA: write a linear ANOVA-model where view is one covariate

$$\mathbf{x}_d = \boldsymbol{\mu}_d + \boldsymbol{\alpha}_a + \boldsymbol{\beta}_b + (\boldsymbol{\alpha\beta})_{ab} + \boldsymbol{\gamma}_d + (\boldsymbol{\alpha\gamma})_{ad} + (\boldsymbol{\beta\gamma})_{bd} + (\boldsymbol{\alpha\beta\gamma})_{abd} + \text{noise},$$

a and b are traditional covariates, disease and treatment

d denotes the view.

- Problem: sources $\mathbf{x}_d$ have different domains
- Possible when using dependencies between datasets (paired samples)
- f^x projects the effects to different domains by hierarchical latent varible modelling

$$\begin{aligned}\mathbf{x} &= \mu + f^x(\alpha_a + \beta_b + (\alpha\beta)_{ab}) + f^x(\alpha_a^x + \beta_b^x + (\alpha\beta)_{ab}^x) + \text{noise},\\ \mathbf{y} &= \mu + f^y(\alpha_a + \beta_b + (\alpha\beta)_{ab}) + f^y(\alpha_a^y + \beta_b^y + (\alpha\beta)_{ab}^y) + \text{noise}, \quad (1)\end{aligned}$$

IDEA: write a linear ANOVA-model where view is one covariate

$$\mathbf{x}_d = \mu_d + \alpha_a + \beta_b + (\alpha\beta)_{ab} + \gamma_d + (\alpha\gamma)_{ad} + (\beta\gamma)_{bd} + (\alpha\beta\gamma)_{abd} + \text{noise},$$

a and b are traditional covariates, disease and treatment

d denotes the view.

- Problem: sources $\mathbf{x}_d$ have different domains

- Possible when using dependencies between datasets (paired samples)
- f^x projects the effects to different domains by hierarchical latent varible modelling

$$\begin{aligned}\mathbf{x} &= \boldsymbol{\mu} + f^x(\boldsymbol{\alpha}_a + \boldsymbol{\beta}_b + (\boldsymbol{\alpha\beta})_{ab}) + f^x(\boldsymbol{\alpha}^x_a + \boldsymbol{\beta}^x_b + (\boldsymbol{\alpha\beta})^x_{ab}) + \text{noise},\\ \mathbf{y} &= \boldsymbol{\mu} + f^y(\boldsymbol{\alpha}_a + \boldsymbol{\beta}_b + (\boldsymbol{\alpha\beta})_{ab}) + f^y(\boldsymbol{\alpha}^y_a + \boldsymbol{\beta}^y_b + (\boldsymbol{\alpha\beta})^y_{ab}) + \text{noise}, \quad (1)\end{aligned}$$

The Bayesian Multi-Way, Multi-View model

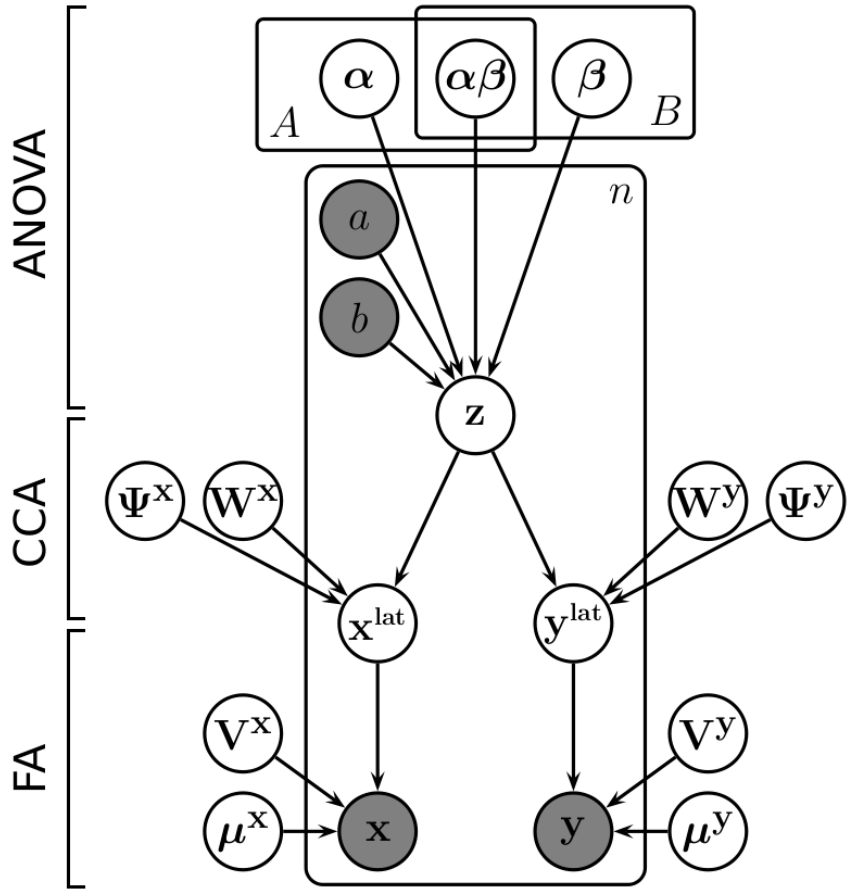
Linear model on shared components

Integrating the sources (bCCA)

Decomposition into shared and source-specific components

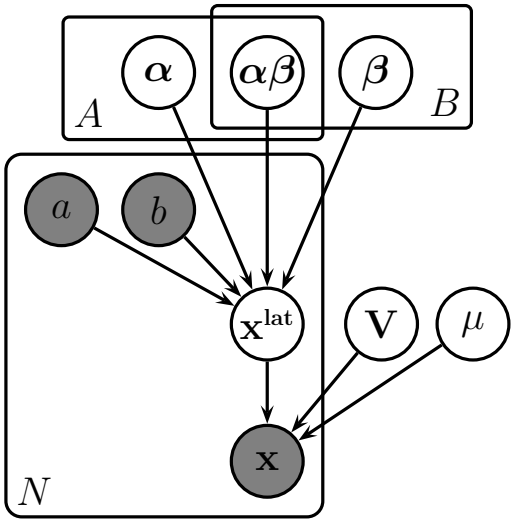
Dimension reduction by regularized clustering Factor analysis

Solved by Gibbs sampling



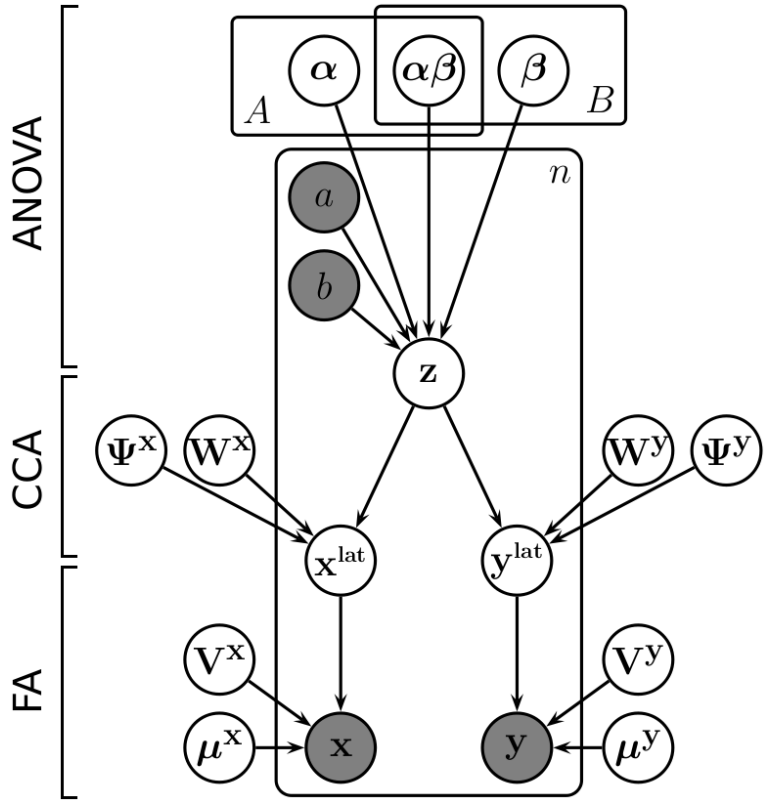
Single-source vs. Multi-source

Single source, multi-way model



[Huopaniemi et al., Two-way analysis of high-dimensional collinear data, ECML09 special issue of Data Mining and Knowledge Discovery: Volume 19, Issue 2 (2009), Page 261]

Multi-source, multi-way

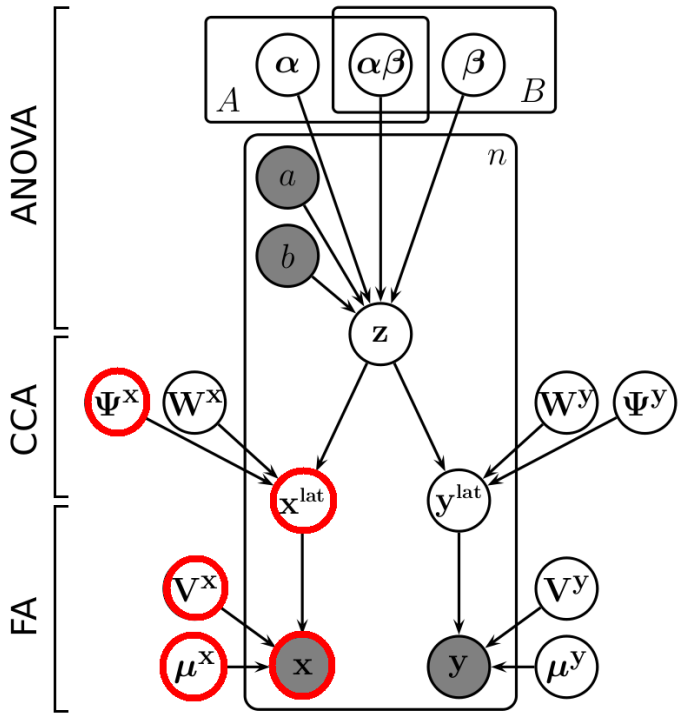


Regularized Factor Analysis

Factor analysis model

$$\mathbf{x}_j^{lat} \sim \mathcal{N}(0, \boldsymbol{\Psi}^x)$$
$$\mathbf{x}_j \sim \mathcal{N}(\boldsymbol{\mu}^x + \mathbf{V}^x \mathbf{x}_j^{lat}, \boldsymbol{\Lambda}^x) \,. \qquad (2)$$

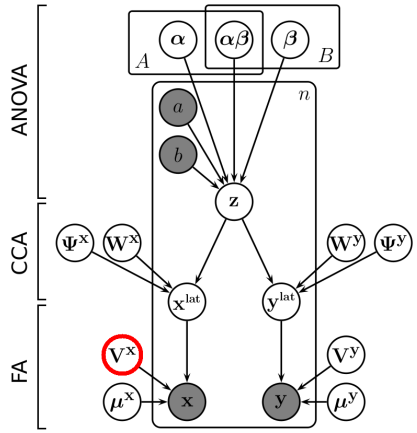
- $\boldsymbol{\Psi}^x$ comes from CCA
- Projection matrix $\mathbf{V}^x$ not invertible, regularize by forming it by clustering variables.
- Each variable comes from one factor(cluster) only
- Interpretation of ANOVA-effects for clusters of similarly behaving variables



Regularizing the projection matrix by clustering

- Each variable comes from exactly one factor
- Cluster variables to different factors by a mixture model
- Multiply by variable-specific scales λ_i

$$\mathbf{V}^x = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 \\ \vdots & \vdots & \vdots \\ 0 & \lambda_j & 0 \\ 0 & \lambda_{j+1} & 0 \\ \vdots & \vdots & \vdots \end{bmatrix}$$



Generative model of CCA for integrating the sources

Generative model of Bayesian CCA

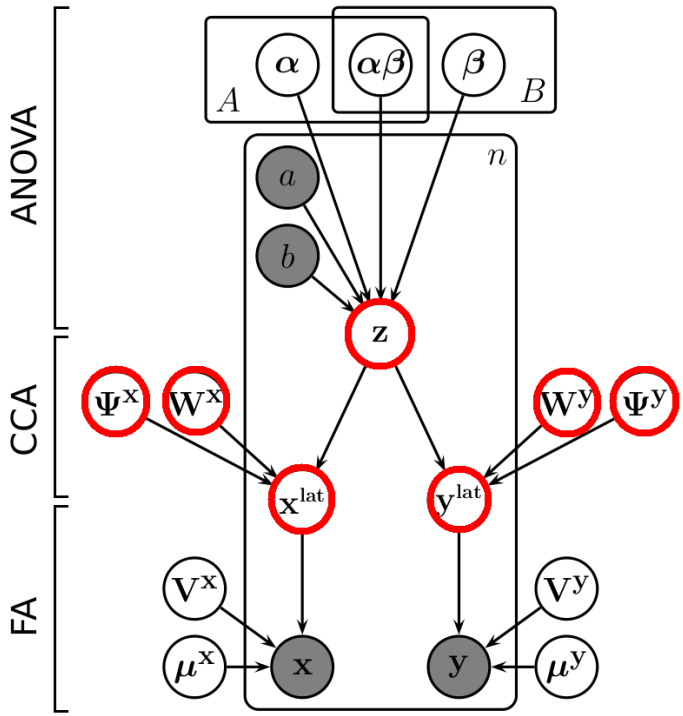
$$\mathbf{z}_j \sim N(0, \mathbf{I}),$$
$$\mathbf{x}_j^{lat} \sim N(\mathbf{W}^x \mathbf{z}_j, \boldsymbol{\Psi}^x), \qquad (3)$$

Projection matrix $\mathbf{W}^x$ has a hierarchical ARD (Automatic relevance determination)-prior

Now, the CCA generates a latent factor $\mathbf{x}_j^{lat}$

In normal bCCA $\mathbf{z}_j$ has a zero-mean prior

However, now the shared variable $\mathbf{z}_j$ will be given a linear model as prior



Bach and Jordan 2005, Klami and Kaski 2007

Covariate information as a linear model on the shared variable

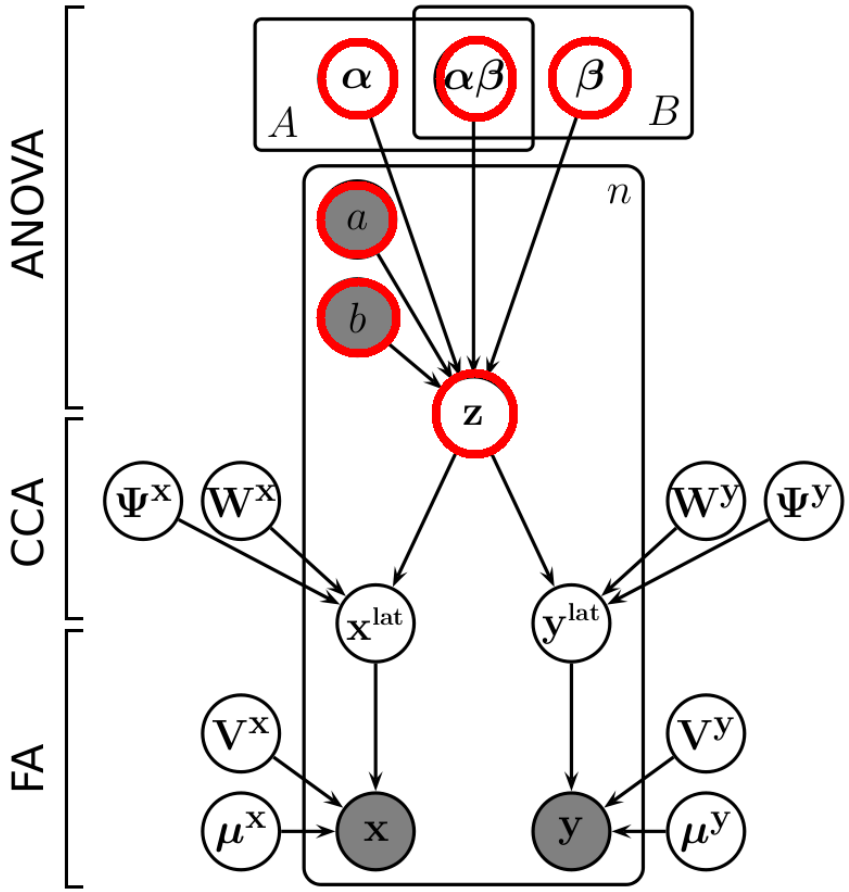
$$\mathbf{z}_j = \boldsymbol{\alpha}_a + \boldsymbol{\beta}_b + (\boldsymbol{\alpha}\boldsymbol{\beta})_{ab} + \boldsymbol{\epsilon}_j, \quad (4)$$

a and b are the observed covariates

Note that the baseline (grand mean) $\boldsymbol{\mu}_x$ is on the lower level

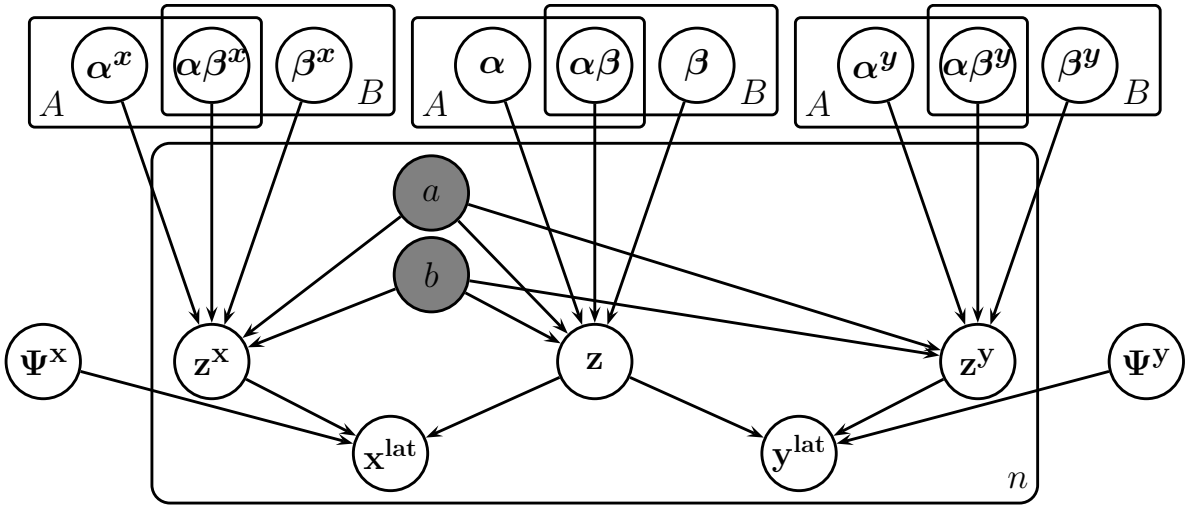
Centering $\boldsymbol{\mu}_x$ to the control group (healthy, untreated)

$\boldsymbol{\alpha}_a$ models the effect of disease
$\boldsymbol{\beta}_b$ treatment
$(\boldsymbol{\alpha}\boldsymbol{\beta})_{ab}$ interaction



Decomposition into shared and source-specific parts

Estimate the shared effect, and the source-specific effects



$$\mathbf{z}_j = \boldsymbol{\alpha}_a + \boldsymbol{\beta}_b + (\boldsymbol{\alpha\beta})_{ab} + \boldsymbol{\epsilon}_j, \tag{5}$$

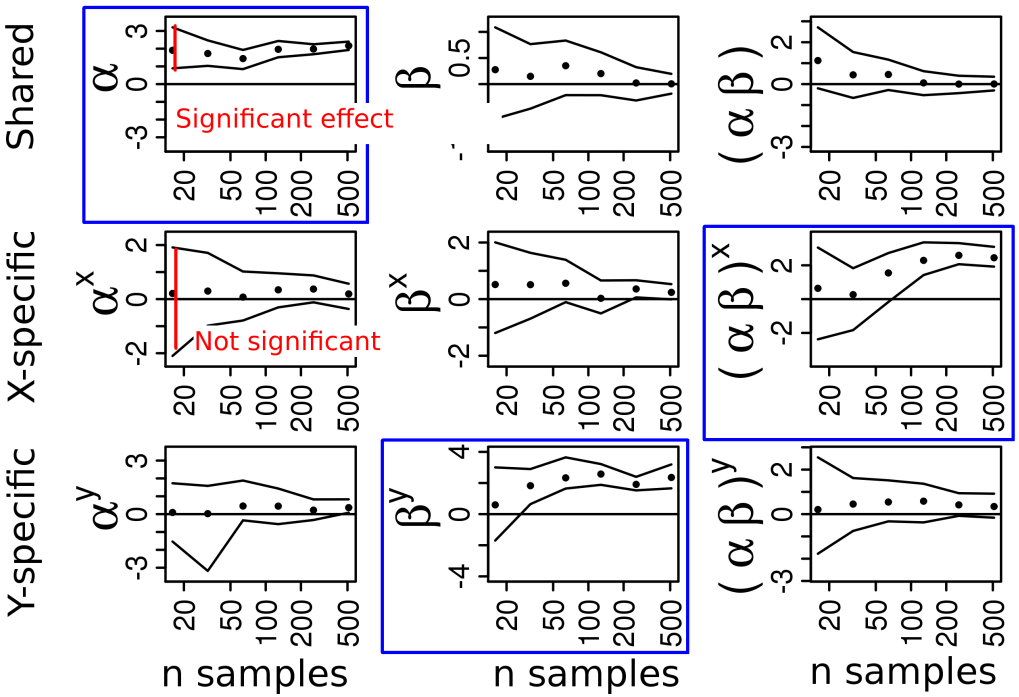
$$\mathbf{z}^x_j = \boldsymbol{\alpha}^x_a + \boldsymbol{\beta}^x_b + (\boldsymbol{\alpha\beta})^x_{ab} + \boldsymbol{\epsilon}^x_j, \tag{6}$$

$$\mathbf{x}^{lat}_j \sim \mathcal{N}(\mathbf{W}^x_{\text{shared}}\mathbf{z}_j + \mathbf{W}^x_{\text{specific}}\mathbf{z}^x_j, \Psi^x). \tag{7}$$

In practice, solve one shared and two specific components by fixing columns of $\mathbf{W}^x$ to zero.

How many samples needed to find effects from data generated by the model?

Generated 200-dim datasets with a multi-way, multi-source setup

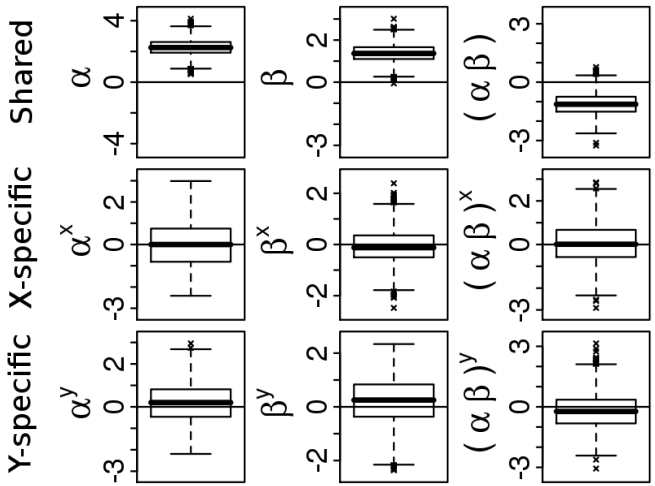


The method finds the generated effects even with small sample size

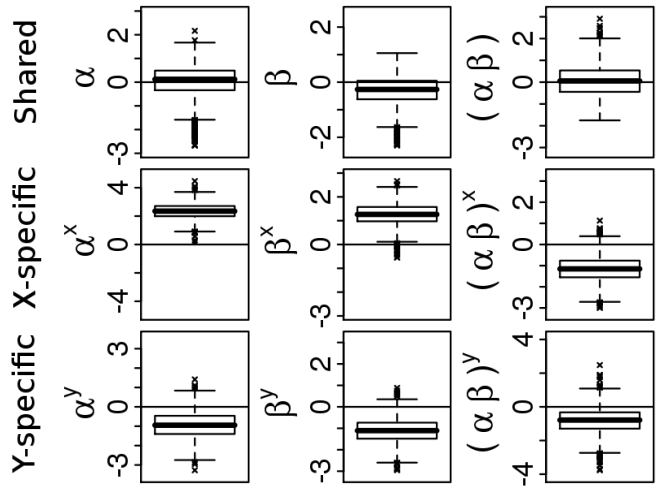
Lung cancer study: Lipidomic profiles, blood, lung, heart

Diseased - healthy, treated - untreated.

α disease effect, β treatment effect, $\alpha\beta$ interaction effect



Experiment 1



Experiment 2

Blood - lung (disease tissue)

Shared disease and treatment effects found

Blood - heart (no disease expected)

Only source-specific effects found

Conclusions

- Biological experiments often have a multi-way, multi-source experimental setup
- Our method finds shared and source-specific covariate effects for similarly behaving groups of variables
- Can deal with the main restriction: $n \ll p$
- Shared effects in different domains can be found in shared-space

Acknowledgements:
Tekes(Finnish Technology Agency) MASI program, AIRC (Adaptive Informatics Research Centre, TKK), HIIT (Helsinki Institute for Information Technology), Graduate School of Computer Science and Engineering, TKK, ComBi graduate School, EU FP7 NoE PASCAL2, ICT 216886