Carnegie Mellon

Sparse Gaussian Chain Graph Models

Calvin McCarter & Seyoung Kim

MACHINE LEARNING DEPARTMENT

RAY AND STEPHANIE LANE Center for Computational Biology

Background on Gaussian Chain Graph Models

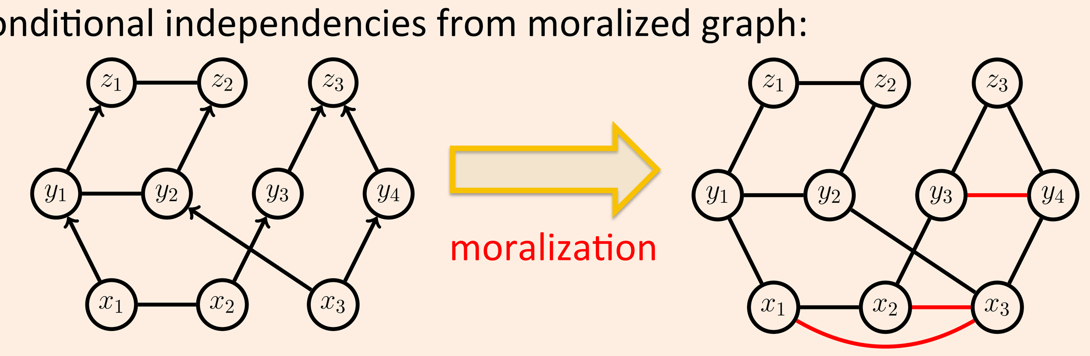
Chain graph models

- Given partition $\{\mathbf{x}_1, \ldots, \mathbf{x}_C\}$, where $\mathbf{x}_\tau \in \mathbb{R}^{|\tau|}$

$$p(\mathbf{x}) = \prod_{\tau=1}^{C} \boxed{p(\mathbf{x}_\tau | \mathbf{x}_{\mathrm{pa}(\tau)})}$$

CRF as chain component model:
- Directed edges: $\mathbf{x}_{\mathrm{pa}(\tau)} \rightarrow \mathbf{x}_\tau$
- Undirected edges: $\mathbf{x}_\tau$

- Conditional independencies from moralized graph:



$y_1 \perp z_2 \mid y_2, z_1$

$x_1 \perp y_3 \mid x_2, x_3$

Non-adjacent variables in moralized graph are conditionally independent given all other variables.

Chain Component Models

Multivariate Linear Regression (Abegaz & Wit, 2013) **Model:**

$$N(\underbrace{\mathbf{B}_\tau \mathbf{x}_{\mathrm{pa}(\tau)}}_{\text{directed edges}}, \underbrace{\Theta_\tau^{-1}}_{\text{undirected edges}})$$

- Markov properties for chain graph models with CRF components do not hold.

Conditional Gaussian Graphical Model (Lauritzen & Wermuth, 1989) **(CGGM):**

$$\exp\left(-\frac{1}{2}\underbrace{\mathbf{x}_\tau^T \Theta_\tau \mathbf{x}_\tau}_{\text{undirected edges}} - \underbrace{\mathbf{x}_\tau^T \Theta_{\tau,\mathrm{pa}(\tau)} \mathbf{x}_{\mathrm{pa}(\tau)}}_{\text{directed edges}}\right) / A(\mathbf{x}_{\mathrm{pa}(\tau)})$$

$$= N\left(-\Theta_\tau^{-1}\Theta_{\tau,\mathrm{pa}(\tau)}\mathbf{x}_{\mathrm{pa}(\tau)}, \Theta_\tau^{-1}\right) = N\left(\mathbf{B}_\tau \mathbf{x}_{\mathrm{pa}(\tau)}, \Theta_\tau^{-1}\right) \quad \boxed{\text{inference}}$$

- Markov properties for chain graph models with CRF components hold.

Almost no work on structure learning for Gaussian chain graph models

Learning the Structure of Gaussian Chain Graph Models

Optimization for linear regression chain component models:

$$\min \sum_{\tau=1}^{C} \mathrm{tr}\left((\mathbf{X}_\tau - \mathbf{X}_{\mathrm{pa}(\tau)}\mathbf{B}_\tau^T)\Theta_\tau(\mathbf{X}_\tau - \mathbf{X}_{\mathrm{pa}(\tau)}\mathbf{B}_\tau^T)^T\right) - N\log|\Theta_\tau| + \lambda\sum_{\tau=1}^{C}\|\mathbf{B}_\tau\|_1 + \gamma\sum_{\tau=1}^{C}\|\Theta_\tau\|_1$$

- ☐ Bi-convex – multiple local optima (Rothman et al., 2010)
- ☐ Slow optimization algorithms

Optimization for CGGM chain component models:

$$\min -\mathcal{L}(\mathbf{X};\Theta) + \lambda\sum_{\tau=1}^{C}\|\Theta_{\tau,\mathrm{pa}(\tau)}\|_1 + \gamma\sum_{\tau=1}^{C}\|\Theta_\tau\|_1$$

- ☑ Convex – global optimum (Sohn & Kim, 2012)
- ☑ Fast optimization algorithms (Wytock & Kolter, 2013)

Advantages of CGGMs as Chain Component Models

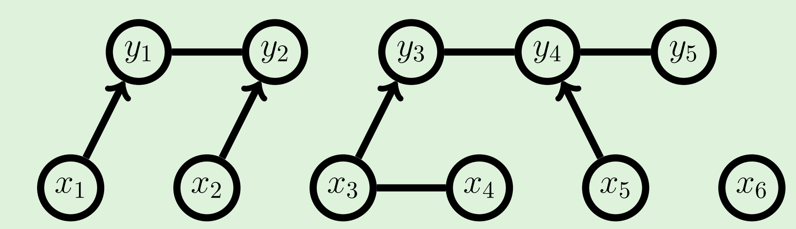
Chain Component Model	Sparse Multivariate Linear Regression	Sparse CGGM
Optimization	Bi-convex	Convex
Computation time	Slow	Fast
Structured sparsity	No	Yes
Leverage model structure for semi-supervised learning?	No	Yes

Sparse Two-Layer Gaussian Chain Graph Models

$$p(\mathbf{y},\mathbf{x}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

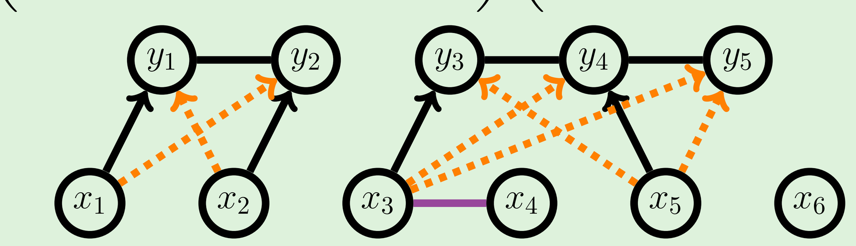
sparse CGGM   sparse GGM

$$= \left(\exp(-\frac{1}{2}\mathbf{y}^T\Theta_{\mathbf{yy}}\mathbf{y} - \mathbf{x}^T\Theta_{\mathbf{xy}}\mathbf{y}/A_1(\mathbf{x}))\right)\left(\exp(-\frac{1}{2}\mathbf{x}^T\Theta_{\mathbf{xx}}\mathbf{x})/A_2\right)$$



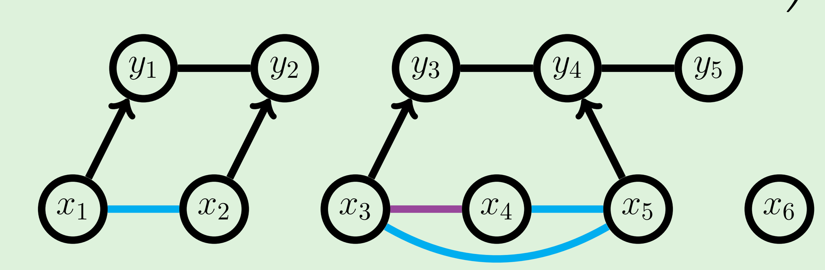
inference

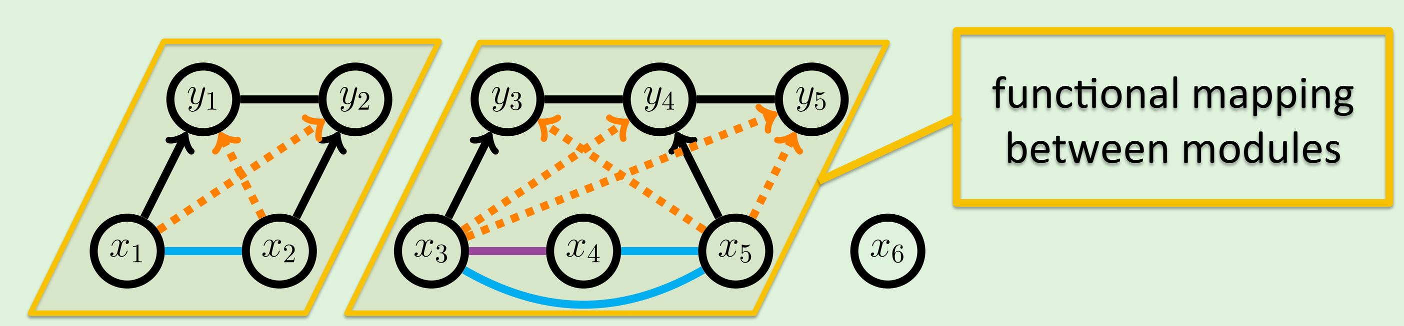
$$= N\left(-\Theta_{\mathbf{yy}}^{-1}\Theta_{\mathbf{xy}}\mathbf{x}, \Theta_{\mathbf{yy}}^{-1}\right)\left(\exp(-\frac{1}{2}\mathbf{x}^T\Theta_{\mathbf{xx}}\mathbf{x})/A_2\right)$$



moralization

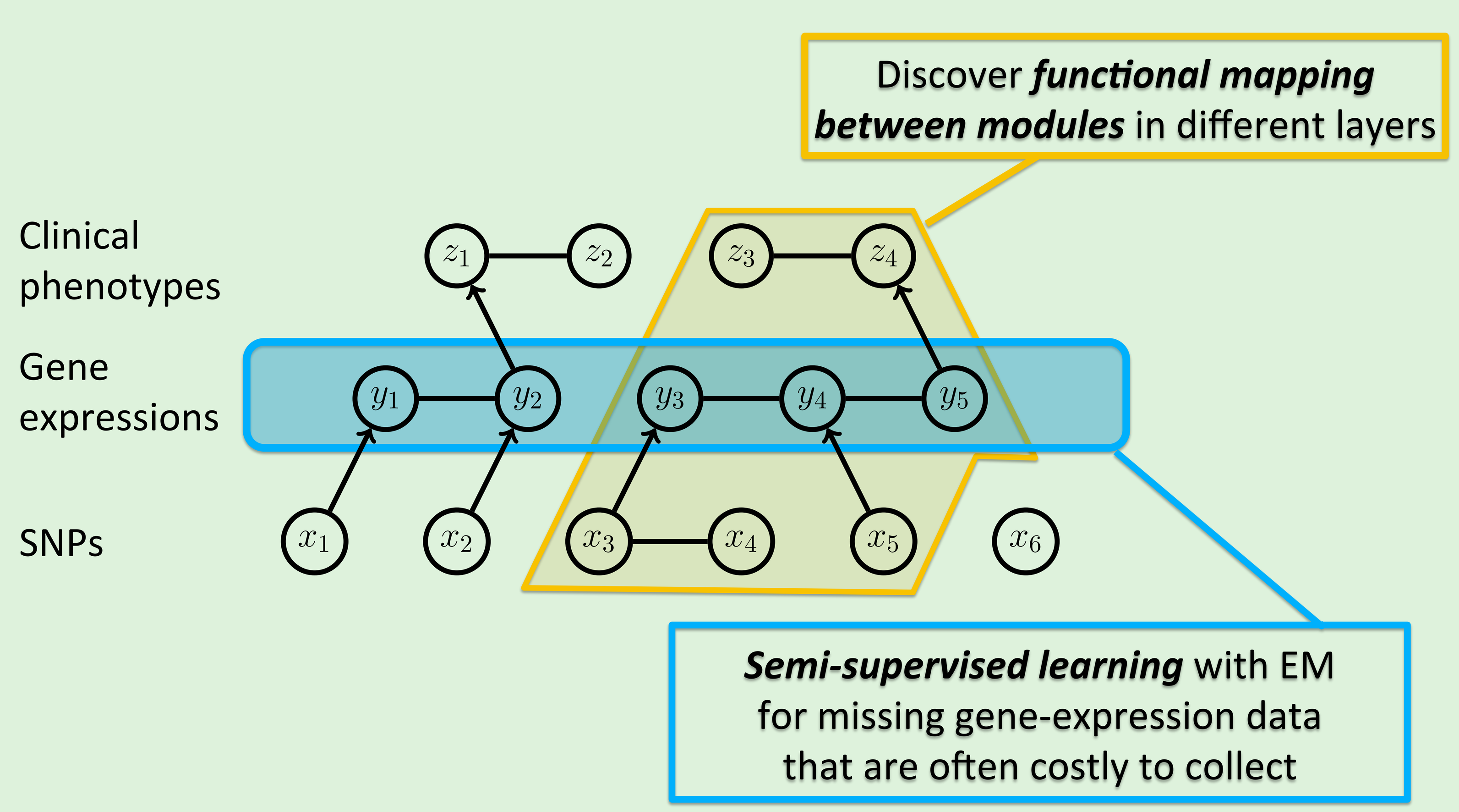
$$= N\left(\mathbf{0}, \begin{pmatrix}\Theta_{\mathbf{yy}} & \Theta_{\mathbf{xy}}^T \\ \Theta_{\mathbf{xy}} & \Theta_{\mathbf{xx}} + \Theta_{\mathbf{xy}}\Theta_{\mathbf{yy}}^{-1}\Theta_{\mathbf{xy}}^T\end{pmatrix}^{-1}\right)$$





Sparse Multi-Layer Gaussian Chain Graph Models for Integrative Genomic Data Analysis

Learn ***cascades of networks*** with multiple data types instead of a ***single network*** from gene expression data



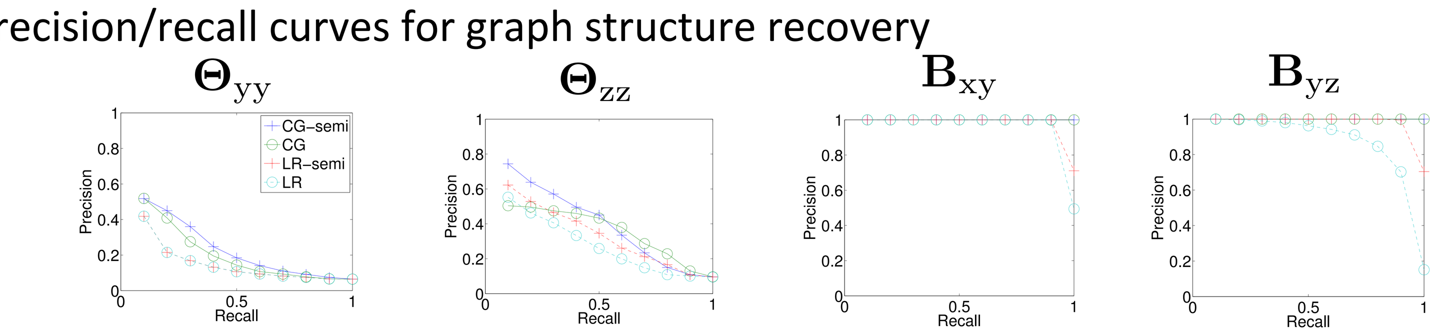
Simulation Results

Better graph structure recovery and prediction accuracy, regardless of true component model!

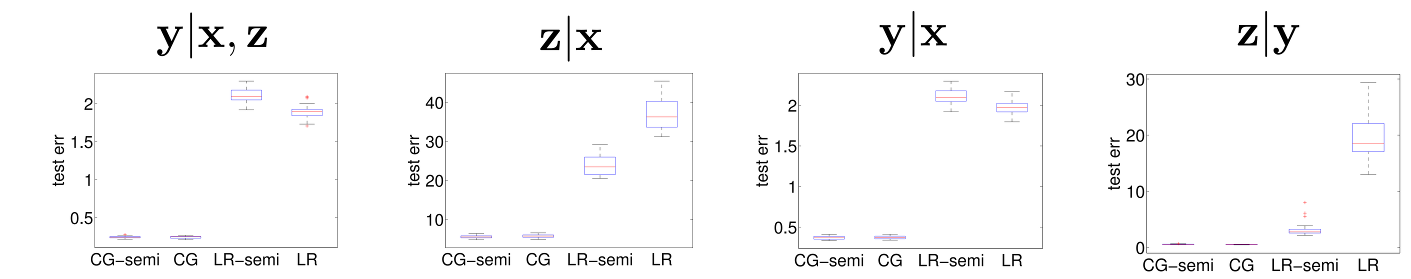
- Problem size: 500 x's, 100 y's, 50 z's
- 400 training samples with 200 samples missing y's

Linear Regression-based True Component Model

Precision/recall curves for graph structure recovery

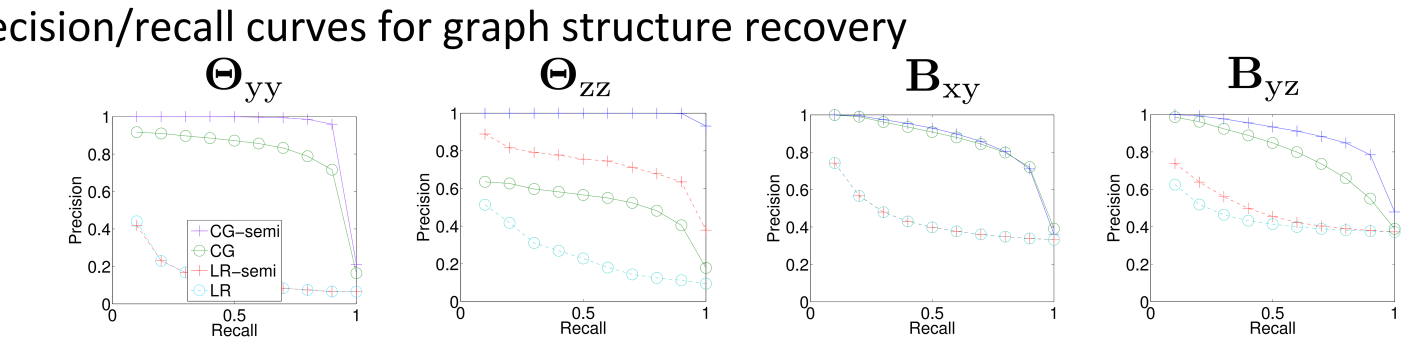


Prediction errors (MSE on test set)

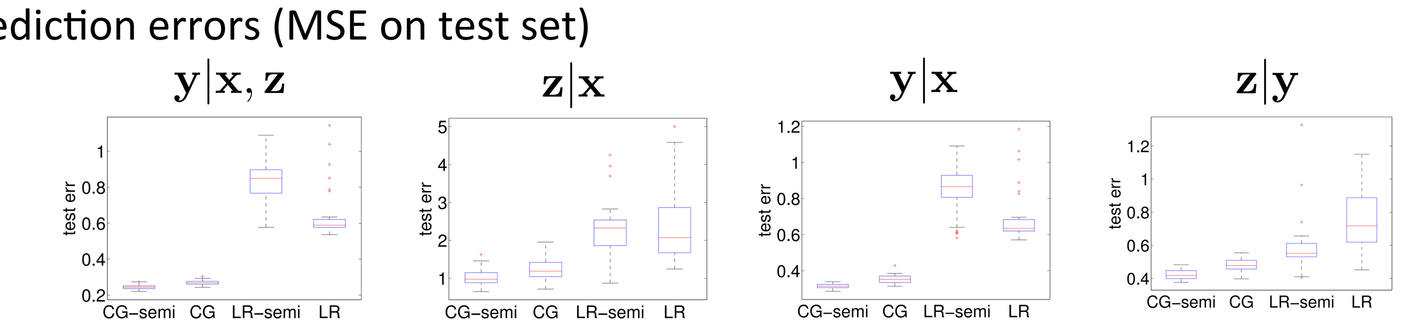


CGGM-based True Component Model

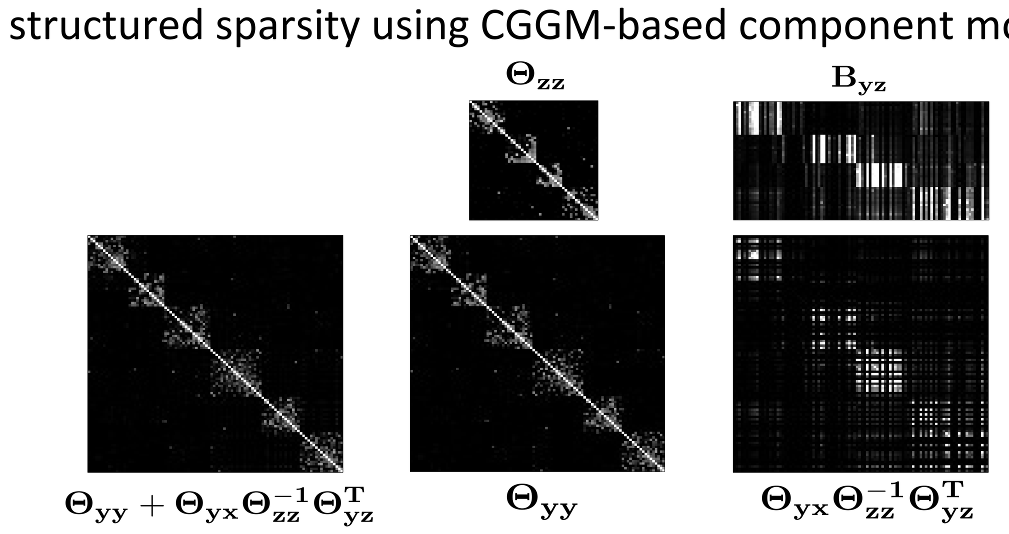
Precision/recall curves for graph structure recovery



Prediction errors (MSE on test set)



Recovery of structured sparsity using CGGM-based component model



Integrative Genomic Data Analysis

- 3 layer chain graph model.
- 1000 SNPs, 200 gene expressions, and 100 phenotypes
- from pancreatic islets study for diabetic mice.
- 306 training samples, 100 validation samples, 100 test samples
- Gene expression data missing for 150 mice.

Task	CG-semi	CG	LR-semi	LR
$\mathbf{y} \mid \mathbf{x},\mathbf{z}$	0.9070	0.9996	1.0958	0.9671
$\mathbf{z} \mid \mathbf{x}$	1.0661	1.0585	1.0505	1.0614
$\mathbf{y} \mid \mathbf{x}$	0.8989	0.9382	0.9332	0.9103
$\mathbf{z} \mid \mathbf{y}$	1.0712	1.0861	1.1095	1.0765

References

Sohn and Kim. Joint estimation of structured sparsity and output structure in multiple-output regression via inverse-covariance regularization. *AISTATS 2012.*

Abegaz and Wit. Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics 2013.*

Wytock and Kolter. Sparse Gaussian conditional random fields: algorithms, theory, and application to energy forecasting. *ICML 2013.*