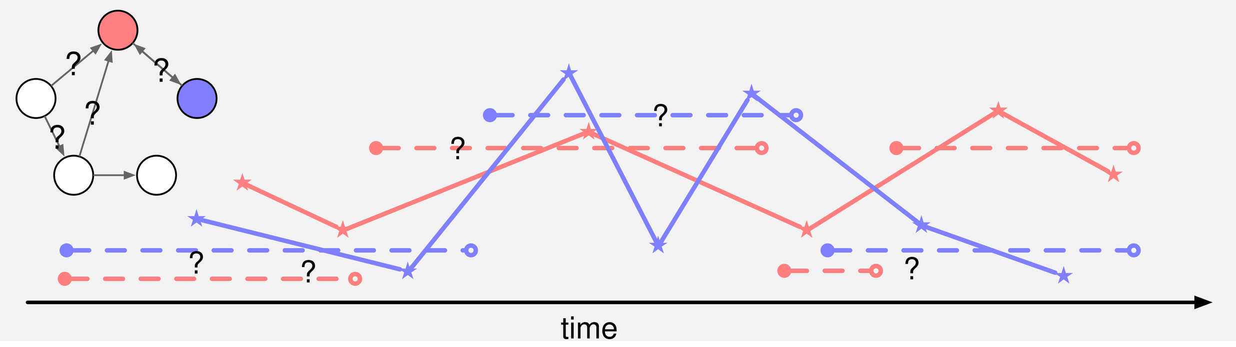
Scalable Structure Learning of Continuous-Time Bayesian Networks from Incomplete Data

Dominik Linzner, Michael Schmidt and Heinz Koeppl, Technische Universität Darmstadt, Germany

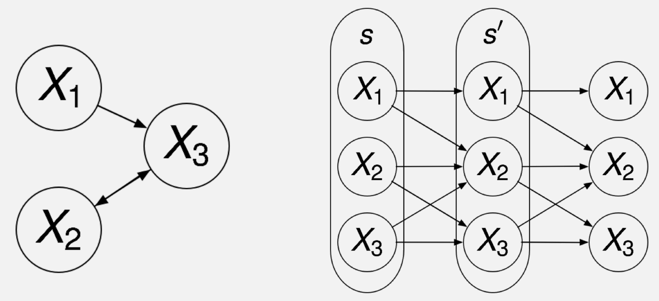
Motivation

- Scalable learning of directed networks from time series data
- If data is noisy there exist two bottlenecks:
  - I. Latent state estimation (**exponential**)
  - II. Structure search (**super-exponential**)



Background: Continuous-time Bayesian Nets (CTBN)

- Factorized Markov Chain



- Can be expressed in terms of individual processes:

$$\{X_i(t)\}_{i:\{1,\dots,N\}} \qquad U_i(t) \equiv \{X_j(t)\}_{j\in \mathrm{par}(i)}$$

- Dynamics in continuous-time via **conditional intensities**:

$$r_i(x', x \mid u) = \lim_{h\to 0} \frac{P_i(X_i(t+h) = x' \mid X_i(t) = x, U_i(t) = u)}{h}$$

- Path likelihood, $P(X_{[0,T]} \mid \{r_i\}_{i:1,\dots,N}) = \prod_{i=1}^N \prod_{x,x',u} r_i(x,x' \mid u)^{M_i(x,x'|u)} e^{-T_i(x|u) r_i(x,x|u)}$

  in terms of **sufficient statistics:** $M_i(x,x' \mid u)$ = number of transitions, $T_i(x \mid u)$ = time spend in configuration

Background: CTBN structure learning

- Realization of conditional intensities unimportant for structure learning
- Marginalzation possible analytically under Gamma-prior

$$r_i(x,x' \mid u) \sim \mathrm{Gam}(\alpha, \beta)$$

$$P(X_{[0,T]} \mid \mathcal{G}) \propto \prod_i \prod_{x,x',u} \Gamma(\bar{\alpha}_i(x,x' \mid u))\, \bar{\beta}_i(x \mid u)^{-\bar{\alpha}_i(x,x'|u)}$$

$$\bar{\alpha}_i(x,x' \mid u) = M_i(x,x' \mid u) + \alpha, \quad \bar{\beta}_i(x \mid u) = T_i(x \mid u) + \beta$$

Model: Mixture Conditional Intensities

- **Central assumption**: Conditional intensities can be written as a mixture of conditional intensities of different parent-support

$$r_i(x,x' \mid u) = \sum_{m\in\mathcal{P}(\mathrm{par}_{\mathcal{G}}(i))} \pi_i(m) r_i(x,x' \mid u_m)$$

- Probability for an edge can be represented via these mixtures

$$p(e_{ij} = 1) = \sum_{m\in\mathcal{P}(\mathrm{par}_{\mathcal{G}}(j))} \pi_j(m) 1(i \in m)$$

- A **marginal likelihood lower-bound** can be computed

$$P(X_{[0,T]} \mid \pi) \geq \prod_i \prod_{m,u_m,x,x'} \frac{\Gamma(\bar{\alpha}_i(x,x'|u_m))}{\bar{\beta}_i(x|u_m)^{\bar{\alpha}_i(x,x'|u_m)}}$$

$$\bar{\alpha}_i(x,x' \mid u_m) = \pi_i(m) M_i(x,x' \mid u_m) + \alpha,$$

$$\bar{\beta}_i(x \mid u_m) \equiv \pi_i(m) T_i(x \mid u_m) + \beta$$

Method

- Based on marginal likelihood we define our **objective function** for structure learning (log-posterior lower-bound)

$$\mathcal{F}_i[\pi] \equiv \sum_{m,u_m,x,x'} \{\ln\Gamma(\bar{\alpha}_i(x,x' \mid u_m)) - \bar{\alpha}_i(x,x' \mid u_m) \ln \bar{\beta}_i(x \mid u_m)\} + \ln \mathrm{Dir}(\pi_i \mid c_i)$$

- Optimal mixture (structure) can be computed via: $\pi_i^* = \arg\max_{\pi_i\in\Delta_i} \{\mathcal{F}_i[\pi]\}$

Incomplete data

- For incomplete data, sufficient statistics are latent and need to be estimated → intractable for large systems
- Solution: **Variational inference [1]**

$$\mathcal{F}_i[\pi] \to \mathcal{F}_i[q,\pi], \quad M_i(x,x' \mid u_m) \to \mathrm{E}_q[M_i(x,x' \mid u_m)], \quad T_i(x \mid u_m) \to \mathrm{E}_q[T_i(x \mid u_m)]$$

- Yields an EM-Algorithm: $\pi_i^* = \arg\max_{\pi_i\in\Delta_i} \{\max_q \{\mathcal{F}_i[q,\pi]\}\}$
- Restricting $\pi_i(m)$ yields different strategies:

Exhaustive:	**Restricted exhaustive:**	**Greedy:**
$\in \mathcal{P}(\mathrm{par}_{\mathcal{G}}(i))$	$\in \mathcal{P}(\mathrm{par}_{g}(i)), \quad g \subseteq \mathcal{G}$	$\in \{m : \mathcal{P}(\mathrm{par}_{\mathcal{G}}(i)) \mid \lvert m\rvert \leq K\}$

Variational Perturbation theory

- **Central assumption**: KL-divergence can be expanded [1]

$$KL(q\|p) = f^{(0)}[q] + \varepsilon f^{(1)}[q] + \mathcal{O}(\varepsilon^2)$$

- KL-divergence decomposes: $KL(q\|p) = -\sum_{i=1}^N \mathcal{F}_i[q_i,\pi_i] + \mathcal{O}(\varepsilon^2)$
- Expectation step is performed by solving set of ODEs (forward-backward):

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_i(t) = \tilde{\Omega}_i^{\pi}(t)\rho_i(t), \quad \frac{\mathrm{d}}{\mathrm{d}t} q_i(t) = q_i(t)\Omega_i^{\pi}(t),$$

Results

Synthetic data

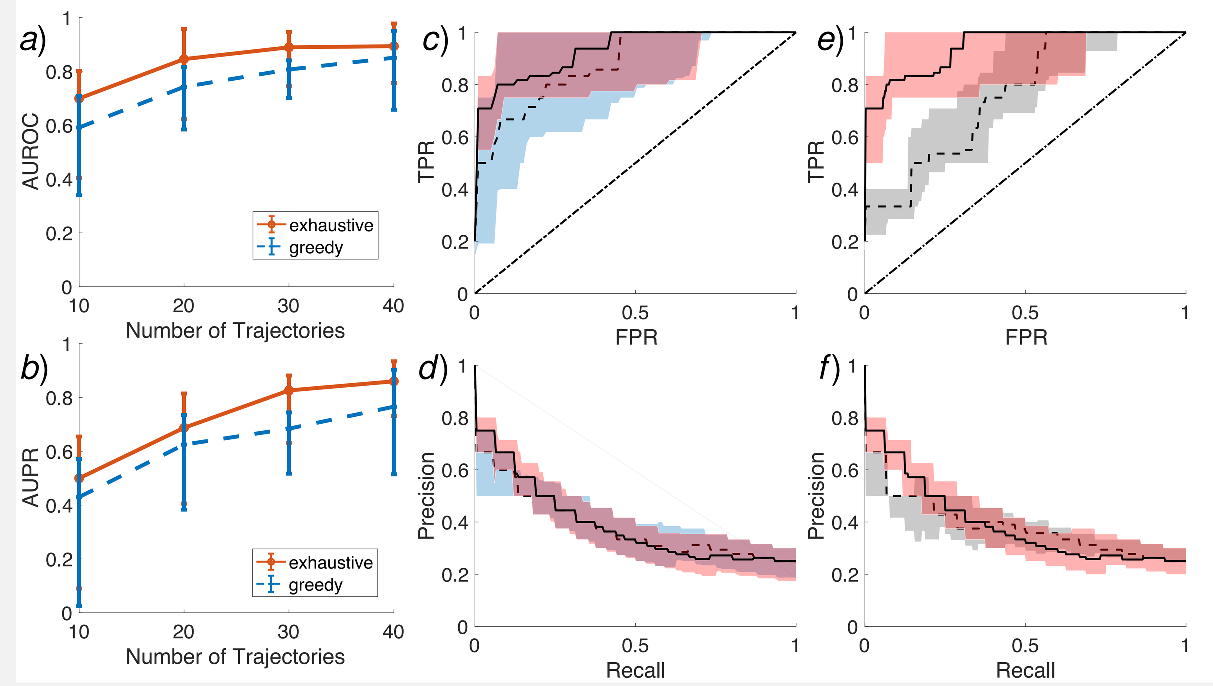


Figure 1: a) AUROCs and b) AUPRs for varying number of trajectories. c) ROC and d) PR curve for 40 trajectories. In all plots (red) denotes the exhaustive, (blue/dashed) the greedy-algorithm. e) ROC-curve f) PR-curve for different initials, where (red) denotes heuristic and (grey/dashed) random. Confidence intervals are given by 75% and 25% percentiles of the results from 30 random graphs, generated as explained in the main text.

Real-world data

- **IRMA gene-regulatory network** dataset [2]
- Has been implemented on cultures of yeast for. benchmarking

method		switch on	switch off
steady state	knockout	0.68 (0.42)	0.81 (0.50)
DBN	G1DBN	0.78 (0.64)	0.61 (0.34)
	VBSSM	0.79 (0.70)	0.76 (0.60)
ODE	TNSI	0.68 (0.51)	0.68 (0.42)
NDS	GP4GRN	0.73 (0.61)	0.76 (0.57)
	CSI^d	0.63 (0.46)	0.86 (0.72)
	CSI^c	0.64 (0.39)	0.73 (0.59)
GC	GCCA	0.71 (0.55)	0.74 (0.65)
CTBN	exhaustive	0.81 (**0.86**)	**0.93** (**0.92**)
	greedy K=2	**0.88** (0.85)	0.91 (0.89)
random		0.65 (0.45)	0.65 (0.45)

Table 1: AUROC (AUPR) of different methods on IRMA-data (top performers in bold).

- **British Household** dataset [3]
- Time-course of questionaires of british citicents about various facts about their lives

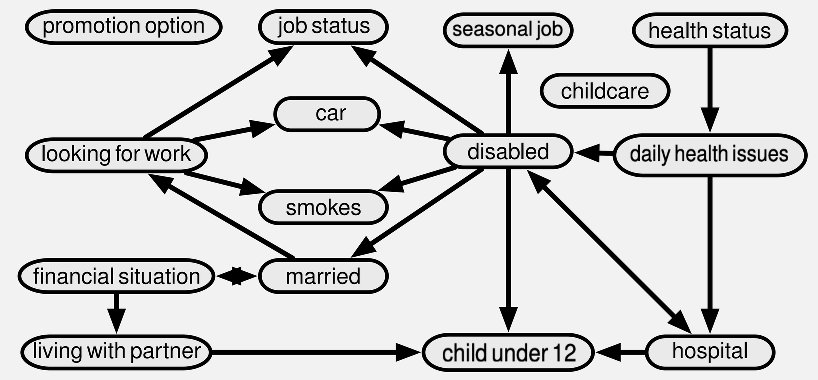


Figure 2: Learned structure using gradient-based greedy structure learning with maximal K=2 parents from 600 trajectories.

Acknowledgements

Dominik Linzner and Michael Schmidt are funded by the European Union's Horizon 2020 research and innovation programme (iPC--Pediatric Cure, No. 826121). Heinz Koeppl acknowledges support by the European Research Council (ERC) within the CONSYN project, No. 773196, and by the Hessian research priority programme LOEWE within the project CompuGene.

References

[1] Linzner, D., & Koeppl, H. (2018). *Advances in Neural Information Processing Systems 31*, 7880–7890.
[2] Cantone, I., Marucci, L., Iorio, F., Ricci, M. A., Belcastro, V., Bansal, M., ... Cosma, M. P. (2009).. *Cell*, *137*(1), 172–181.
[3] British Household Panel Survey (ESRC Research Centre on Micro-social Change, 2003)