Causal Inference in case of feedback

Joris Mooij
j.m.mooij@uva.nl

Institute for Computing
and Information Sciences

Radboud University Nijmegen

Informatics Institute

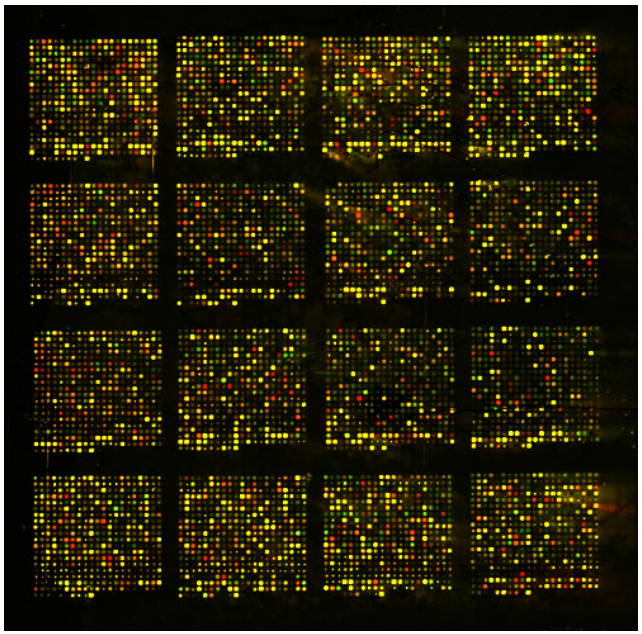
University of Amsterdam

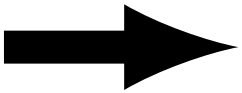
December 12th, 2013

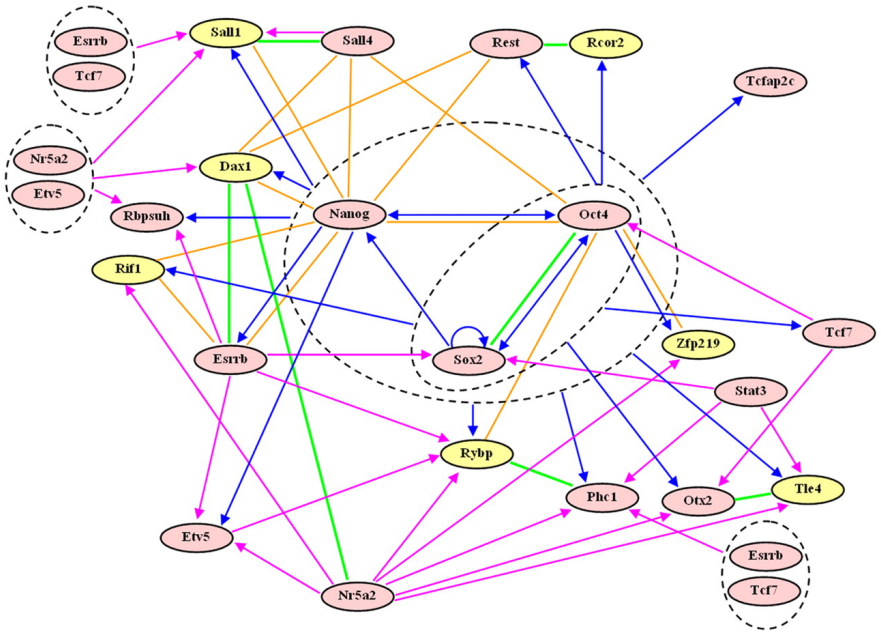
Part I

Introduction

Genetics:
how to infer gene regulatory networks from micro-array data?

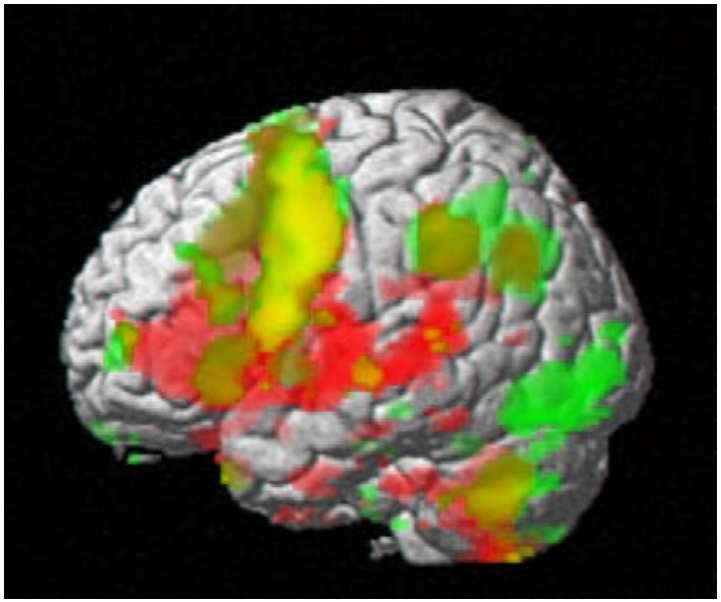


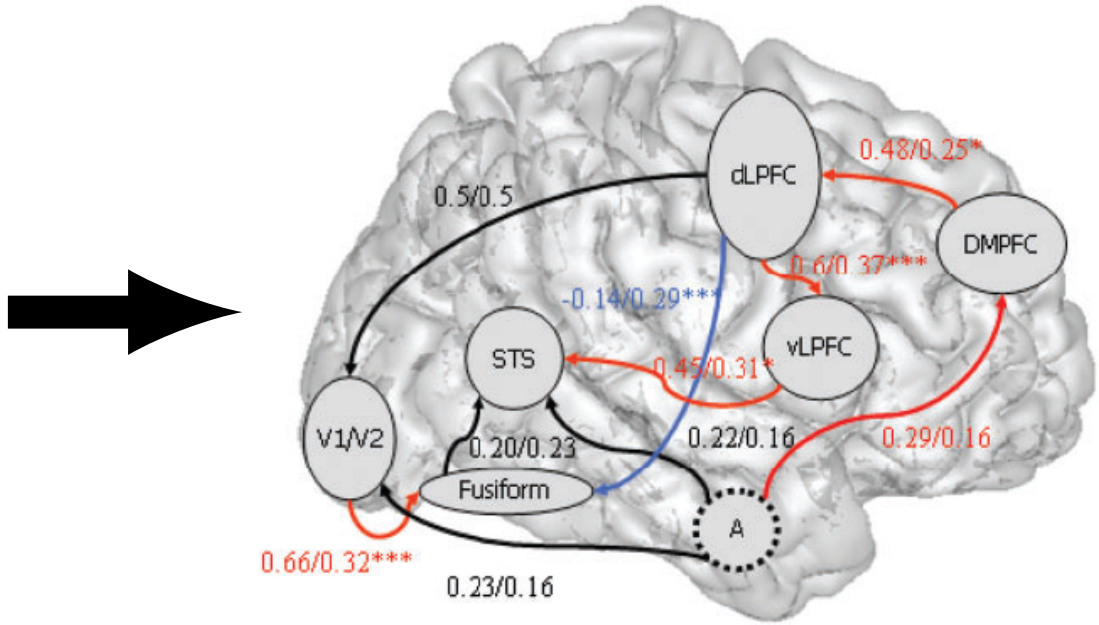




Causality: ubiquitous in the sciences

Neuroscience:
how to infer functional connectivity networks from fMRI data?





Causality: ubiquitous in the sciences

Social sciences:
does playing violent computer games cause aggressive behavior?

Causality: ubiquitous in the sciences

Economy:
does austerity reduce national debt?

Probabilistic Inference

- **Modeling**: modeling the joint distribution of a set of random variables

$$p(X, Y, Z, \dots) = f(X, Y, Z, \dots; \boldsymbol{\theta})$$

Probabilistic Inference

- **Modeling**: modeling the joint distribution of a set of random variables

$$p(X, Y, Z, \dots) = f(X, Y, Z, \dots; \boldsymbol{\theta})$$

- **Reasoning**: using the rules of probability theory to express different marginal and conditional distributions in terms of each other

$$\text{Bayes' rule: } p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}$$

Probabilistic Inference

- **Modeling**: modeling the joint distribution of a set of random variables

$$p(X, Y, Z, \dots) = f(X, Y, Z, \dots; \boldsymbol{\theta})$$

- **Reasoning**: using the rules of probability theory to express different marginal and conditional distributions in terms of each other

$$\text{Bayes' rule: } p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}$$

- **Learning**: find the best model(s) to describe the data

$$\text{ML estimation: } \arg\max_{\boldsymbol{\theta}} \prod_{i=1}^{N} f(x_i, y_i, z_i, \dots; \boldsymbol{\theta})$$

Probabilistic Inference

- **Modeling**: modeling the joint distribution of a set of random variables

  $$p(X, Y, Z, \dots) = f(X, Y, Z, \dots; \boldsymbol{\theta})$$

- **Reasoning**: using the rules of probability theory to express different marginal and conditional distributions in terms of each other

  Bayes' rule: $p(X|Y) = \dfrac{p(Y|X)p(X)}{p(Y)}$

- **Learning**: find the best model(s) to describe the data

  ML estimation: $\arg\max_{\boldsymbol{\theta}} \prod_{i=1}^{N} f(x_i, y_i, z_i, \dots; \boldsymbol{\theta})$

- **Prediction**: given a model and an observation of some random variables, what are the values of other random variables?

  $p(Y|X = x) = ?$

Causal Inference

- **Causal Modeling**: modeling the joint distribution of a set of random variables and how this changes under interventions

$$p(X, Y, Z, \dots) = \dots,$$
$$p(Y, Z \dots | \, \mathrm{do}(X = x)) = \dots,$$
$$p(X, Z, \dots | \, \mathrm{do}(Y = y)) = \dots,$$
$$p(Z, \dots | \, \mathrm{do}(X = x, Y = y)) = \dots$$

Causal Inference

- **Causal Modeling**: modeling the joint distribution of a set of random variables and how this changes under interventions

  $p(X, Y, Z, \dots) = \dots,$
  $p(Y, Z \dots | \, \mathrm{do}(X = x)) = \dots,$
  $p(X, Z, \dots | \, \mathrm{do}(Y = y)) = \dots,$
  $p(Z, \dots | \, \mathrm{do}(X = x, Y = y)) = \dots$

- **Causal Reasoning**: using rules for expressing different marginal, conditional and interventional distributions in terms of each other

  Pearl's "do-calculus", SGS's "manipulation theorem"

Causal Inference

- **Causal Modeling**: modeling the joint distribution of a set of random variables and how this changes under interventions

  $p(X, Y, Z, \dots) = \dots,$
  $p(Y, Z \dots | \, \mathrm{do}(X = x)) = \dots,$
  $p(X, Z, \dots | \, \mathrm{do}(Y = y)) = \dots,$
  $p(Z, \dots | \, \mathrm{do}(X = x, Y = y)) = \dots$

- **Causal Reasoning**: using rules for expressing different marginal, conditional and interventional distributions in terms of each other

  Pearl's "do-calculus", SGS's "manipulation theorem"

- **Causal Discovery**: find the best causal model(s) to describe the observational data and interventional data

  PC, FCI algorithms (use only observational data)

Causal Inference

- **Causal Modeling**: modeling the joint distribution of a set of random variables and how this changes under interventions

  $p(X, Y, Z, \dots) = \dots,$
  $p(Y, Z \dots | \, \mathrm{do}(X = x)) = \dots,$
  $p(X, Z, \dots | \, \mathrm{do}(Y = y)) = \dots,$
  $p(Z, \dots | \, \mathrm{do}(X = x, Y = y)) = \dots$

- **Causal Reasoning**: using rules for expressing different marginal, conditional and interventional distributions in terms of each other

  Pearl's "do-calculus", SGS's "manipulation theorem"

- **Causal Discovery**: find the best causal model(s) to describe the observational data and interventional data

  PC, FCI algorithms (use only observational data)

- **Causal Prediction**: given a causal model and given an intervention, what are the values of other random variables?

  "Covariate adjustment": $p(Y | \, \mathrm{do}(X)) = \sum_{\mathbf{W}} p(Y | X, \mathbf{W}) p(\mathbf{W})$

Probabilistic inference vs. causal inference

Traditional statistics, machine learning

- Models the distribution of the data
- Focuses on predicting results of observations
- Useful e.g. in medical diagnosis: given the symptoms, what is the most likely disease?

Causal Inference

- Models the mechanism that generates the data
- Also allows to predict results of interventions
- Useful e.g. in medical treatment: if we treat the patient with a drug, will it cure the disease?

Outline

Introduction to Causal Inference:

1. Introduction
2. Causal Modeling

Some recent developments:

3. Causal Modeling in case of feedback[1]
4. Causal Discovery in case of feedback[2]
5. Outlook

[1] Joint work with Dominik Janzing and Bernhard Schölkopf

[2] Joint work with Tom Heskes

Part II

Causal Modeling

Structural Causal Models: Definition

Definition [Pearl, 2000; Wright, 1921]

A Structural Causal Model (SCM), also known as Structural Equation Models (SEM), $\mathcal{M}$ is defined by:

Structural Causal Models: Definition

Definition [Pearl, 2000; Wright, 1921]

A Structural Causal Model (SCM), also known as Structural Equation Models (SEM), $\mathcal{M}$ is defined by:

1. N observed random variables $X_1, \ldots, X_N$ and N latent random variables $E_1, \ldots, E_N$

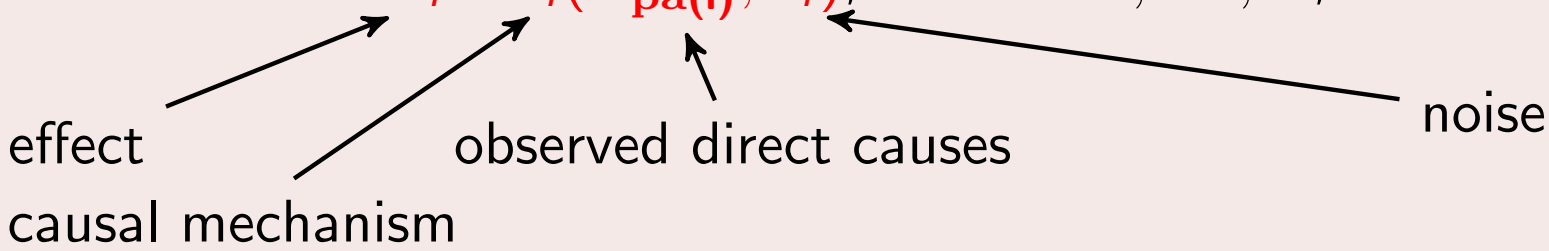
Structural Causal Models: Definition

Definition [Pearl, 2000; Wright, 1921]

A Structural Causal Model (SCM), also known as Structural Equation Models (SEM), $\mathcal{M}$ is defined by:

1. N observed random variables $X_1, \ldots, X_N$ and N latent random variables $E_1, \ldots, E_N$
2. N structural equations:

$$X_i = f_i(\mathbf{X_{pa(i)}}, E_i), \qquad i = 1, \ldots, N;$$



where the subsets $\mathrm{pa}(i) \subseteq \{1, \ldots, N\} \setminus \{i\}$ define the observed direct causes of X_i,

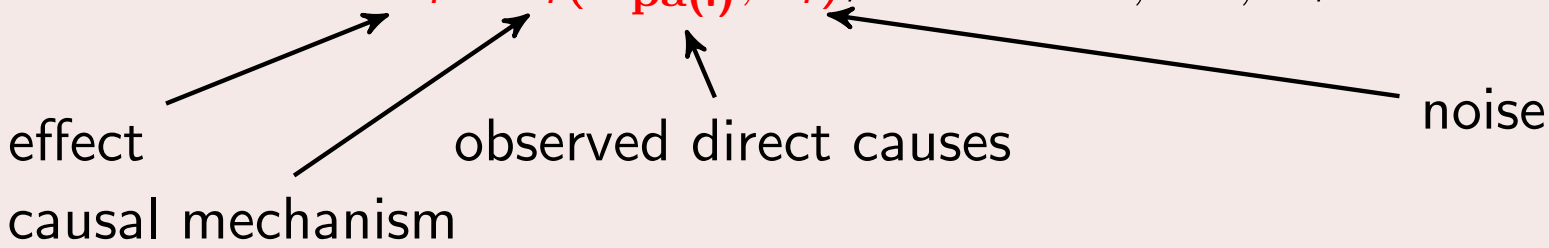
Structural Causal Models: Definition

Definition [Pearl, 2000; Wright, 1921]

A Structural Causal Model (SCM), also known as Structural Equation Models (SEM), $\mathcal{M}$ is defined by:

1. N observed random variables $X_1, \ldots, X_N$ and N latent random variables $E_1, \ldots, E_N$
2. N structural equations:

$$X_i = f_i(\mathbf{X_{pa(i)}}, E_i), \qquad i = 1, \ldots, N;$$



   where the subsets $\mathrm{pa}(i) \subseteq \{1, \ldots, N\} \setminus \{i\}$ define the observed direct causes of X_i,

3. a joint probability distribution $p(E_1, \ldots, E_N)$ on latent variables.

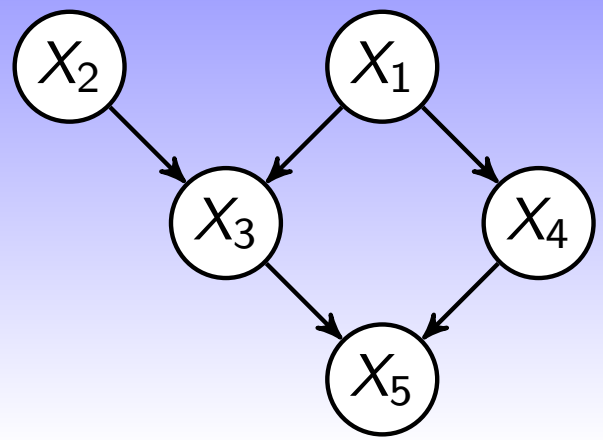
Structural Causal Models: Example

Example

Structural causal model $\mathcal{M}$:

$$
\begin{aligned}
X_1 &= f_1(E_1) & p(E_1) &= \dots \\
X_2 &= f_2(E_2) & p(E_2) &= \dots \\
X_3 &= f_3(X_1, X_2, E_3) & p(E_3) &= \dots \\
X_4 &= f_4(X_1, E_4) & p(E_4) &= \dots \\
X_5 &= f_5(X_3, X_4, E_5) & p(E_5) &= \dots \\
& & p(\mathbf{E}) &= \textstyle\prod_i p(E_i)
\end{aligned}
$$

Causal graph $\mathcal{G}_{\mathcal{M}}$:



Definition

Given a SCM $\mathcal{M}$, the causal graph $\mathcal{G}_{\mathcal{M}}$ is the directed graph with vertices $\{X_1, \dots, X_N\}$ and edges $X_j \to X_i$ iff f_i depends on X_j (i.e., if $j \in \mathrm{pa}(i)$).

Structural Causal Models: Interventions

For a *causal* model, we also need to model interventions.

Structural Causal Models: Interventions

For a *causal* model, we also need to model interventions.

Interventions in SCMs

An intervention $\text{do}(X_i = \xi_i)$ on a variable X_i, forcing it to attain the value ξ_i, changes the structural equation for X_i as follows:

Original SCM $\mathcal{M}$:

$$
\begin{aligned}
X_i &= f_i(\mathbf{X}_{\text{pa}(i)}, E_i) \\
X_j &= f_j(\mathbf{X}_{\text{pa}(j)}, E_j) \quad \forall j \neq i \\
p(\mathbf{E}) &= \dots
\end{aligned}
$$

Intervened SCM $\mathcal{M}_{\xi_i}$:

$$
\begin{aligned}
X_i &= \xi_i \\
X_j &= f_j(\mathbf{X}_{\text{pa}(j)}, E_j) \quad \forall j \neq i \\
p(\mathbf{E}) &= \dots
\end{aligned}
$$

Structural Causal Models: Interventions

For a *causal* model, we also need to model interventions.

Interventions in SCMs

An intervention $\text{do}(X_i = \xi_i)$ on a variable X_i, forcing it to attain the value ξ_i, changes the structural equation for X_i as follows:

Original SCM $\mathcal{M}$:	Intervened SCM $\mathcal{M}_{\xi_i}$:
$X_i = f_i(\mathbf{X}_{\text{pa}(i)}, E_i)$	$X_i = \xi_i$
$X_j = f_j(\mathbf{X}_{\text{pa}(j)}, E_j) \quad \forall j \neq i$	$X_j = f_j(\mathbf{X}_{\text{pa}(j)}, E_j) \quad \forall j \neq i$
$p(\mathbf{E}) = \dots$	$p(\mathbf{E}) = \dots$

Interpretation: overriding default causal mechanisms that normally would determine the values of the intervened variables.

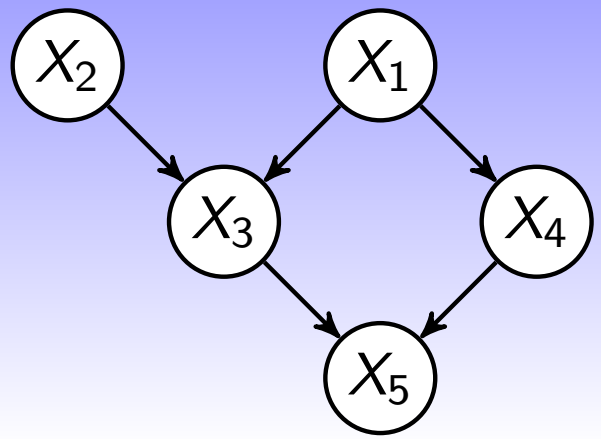
Example

Observational (no intervention):

Structural causal model $\mathcal{M}$:

$$
\begin{array}{ll}
X_1 = f_1(E_1) & p(E_1) = \dots \\
X_2 = f_2(E_2) & p(E_2) = \dots \\
X_3 = f_3(X_1, X_2, E_3) & p(E_3) = \dots \\
X_4 = f_4(X_1, E_4) & p(E_4) = \dots \\
X_5 = f_5(X_3, X_4, E_5) & p(E_5) = \dots \\
 & p(\mathbf{E}) = \prod_i p(E_i)
\end{array}
$$

Causal graph $\mathcal{G}_{\mathcal{M}}$:



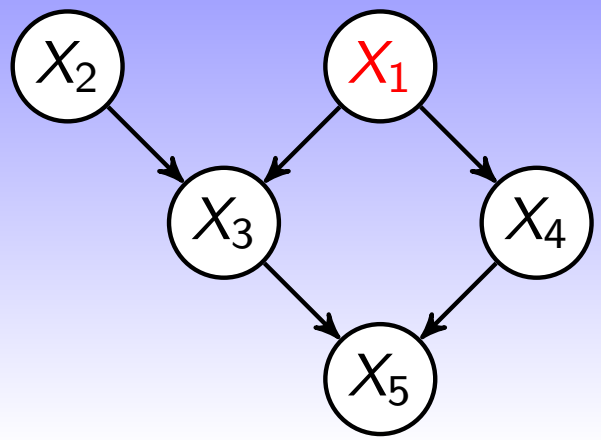
Example

Intervention $\text{do}(X_1 = \xi_1)$:

Structural causal model $\mathcal{M}_{\xi_1}$:

$$
\begin{array}{ll}
X_1 = \xi_1 & p(E_1) = \dots \\
X_2 = f_2(E_2) & p(E_2) = \dots \\
X_3 = f_3(X_1, X_2, E_3) & p(E_3) = \dots \\
X_4 = f_4(X_1, E_4) & p(E_4) = \dots \\
X_5 = f_5(X_3, X_4, E_5) & p(E_5) = \dots \\
 & p(\mathbf{E}) = \prod_i p(E_i)
\end{array}
$$

Causal graph $\mathcal{G}_{\mathcal{M}_{\xi_1}}$:



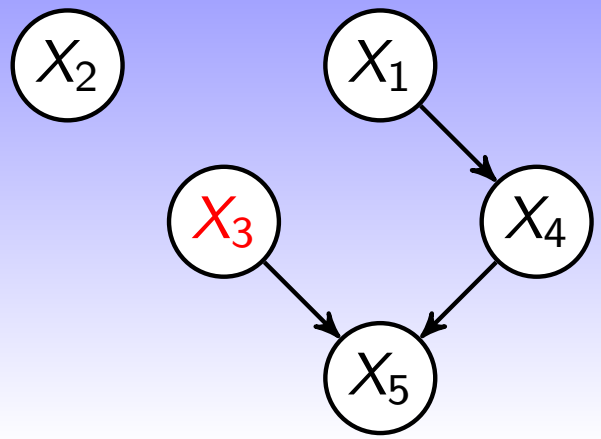
Example

Intervention $\text{do}(X_3 = \xi_3)$:

Structural causal model $\mathcal{M}_{\xi_3}$:

$$
\begin{array}{ll}
X_1 = f_1(E_1) & p(E_1) = \dots \\
X_2 = f_2(E_2) & p(E_2) = \dots \\
X_3 = \xi_3 & p(E_3) = \dots \\
X_4 = f_4(X_1, E_4) & p(E_4) = \dots \\
X_5 = f_5(X_3, X_4, E_5) & p(E_5) = \dots \\
 & p(\mathbf{E}) = \prod_i p(E_i)
\end{array}
$$

Causal graph $\mathcal{G}_{\mathcal{M}_{\xi_3}}$:



Confounders and causal sufficiency

Definition: Confounder

A confounder is a latent common cause of two or more observed variables.

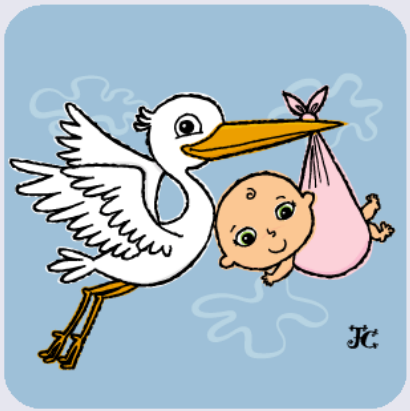
Confounders and causal sufficiency

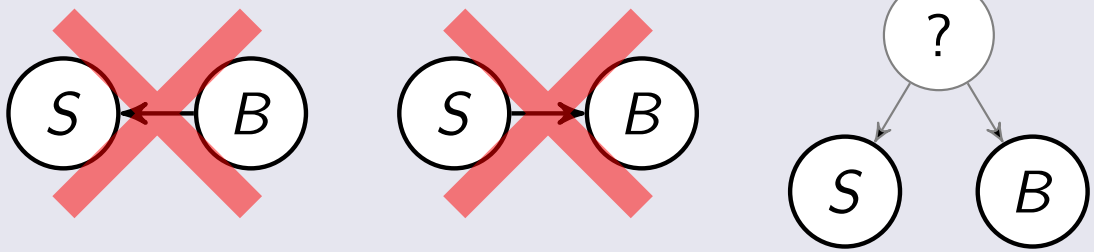
Definition: Confounder

A confounder is a latent common cause of two or more observed variables.

Example

- Significant correlation ($p = 0.008$) between human birth rate and number of stork populations in European countries [Matthews, 2000]
- Most people nowadays do not believe that storks deliver babies (nor that babies deliver storks)
- There must be some confounder explaining the correlation





Confounders and causal sufficiency

Definition: Confounder

A confounder is a latent common cause of two or more observed variables.

Absence of confounders implies causal sufficiency.

Definition: Causal Sufficiency

If all latent variables $E_1, \ldots, E_N$ in an SCM are jointly independent, i.e., if

$$p(\mathbf{E}) = \prod_{i=1}^{N} p(E_i)$$

then we say that the observed variables $\mathbf{X}$ are causally sufficient.

Causal feedback and (A)cyclicity

Definition: causal feedback

A SCM incorporates causal feedback if its graph contains a directed cycle

$$X_{i_0} \to X_{i_1} \to \cdots \to X_{i_n}, \qquad X_{i_0} = X_{i_n}$$

If it does not contain such a directed cycle, the model is called acyclic. If it is also causally sufficient, its graph is a Directed Acyclic Graph (DAG).

Causal feedback and (A)cyclicity

Definition: causal feedback

A SCM incorporates causal feedback if its graph contains a directed cycle

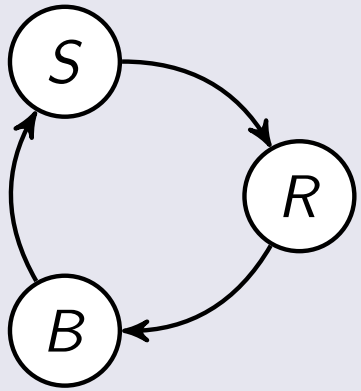
$$X_{i_0} \to X_{i_1} \to \cdots \to X_{i_n}, \qquad X_{i_0} = X_{i_n}$$

If it does not contain such a directed cycle, the model is called acyclic. If it is also causally sufficient, its graph is a Directed Acyclic Graph (DAG).

Example

In economy, causal feedback is often present:

R: risks taken by bank;
B: imminent bankruptcy;
S: saved by the government.



Factorization: Bayesian Networks

Theorem

Any probability distribution induced by an *acyclic*, *causally sufficient* SCM $\mathcal{M}$ can be factorized as:

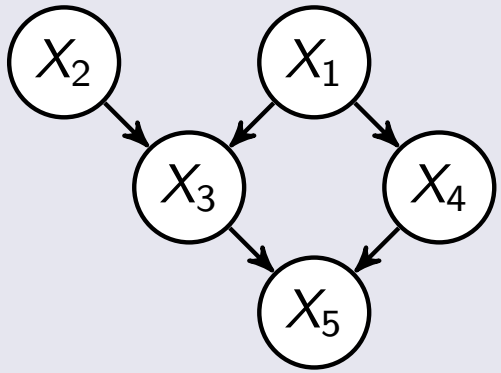
$$p_{\mathcal{M}}(X_1, \ldots, X_N) = \prod_{i=1}^{N} p_{\mathcal{M}}(X_i \mid \mathbf{X}_{\mathrm{pa}(i)})$$

Example

Structural causal model $\mathcal{M}$:

$$\begin{aligned} X_1 &= f_1(E_1) & p(E_1) &= \ldots \\ X_2 &= f_2(E_2) & p(E_2) &= \ldots \\ X_3 &= f_3(X_1, X_2, E_3) & p(E_3) &= \ldots \\ X_4 &= f_4(X_1, E_4) & p(E_4) &= \ldots \\ X_5 &= f_5(X_3, X_4, E_5) & p(E_5) &= \ldots \end{aligned}$$

Causal graph $\mathcal{G}_{\mathcal{M}}$:



$$p(X_1, \ldots, X_5) = p(X_1)\, p(X_2)\, p(X_3 \mid X_1, X_2)\, p(X_4 \mid X_1)\, p(X_5 \mid X_3, X_4)$$

Causal Reasoning: Truncated factorization

The following theorem expresses the joint distribution of a Bayesian network after an intervention. It is an example of causal reasoning.

Theorem: Truncated factorization

Any probability distribution induced by an *acyclic*, *causally sufficient* SCM $\mathcal{M}$ can be factorized as:

$$p_{\mathcal{M}}(X_1, \ldots, X_N) = \prod_{i=1}^{N} p_{\mathcal{M}}(X_i \mid \mathbf{X}_{\mathrm{pa}(i)})$$

After an intervention $\mathrm{do}(\mathbf{X}_I = \boldsymbol{\xi}_I)$, the probability distribution becomes:

$$p_{\mathcal{M}_{\boldsymbol{\xi}_I}}\left(X_1, \ldots, X_N \mid \mathrm{do}(\mathbf{X}_I = \boldsymbol{\xi}_I)\right) = \prod_{\substack{i=1 \\ i \notin I}}^{N} p_{\mathcal{M}}(X_i \mid \mathbf{X}_{\mathrm{pa}(i)}) \prod_{i \in I} \mathbf{1}_{[X_i = \xi_i]}$$

Part III

Causal Modeling in case of feedback

Feedback

Cyclic causal dependencies are also called feedback loops. Examples:

- Holding a microphone too close to a loudspeaker.
- Predator-prey relationships in biology.
- Computer programs running on a single core are *acyclic*; parallel programs running on multiple cores can be *cyclic*.

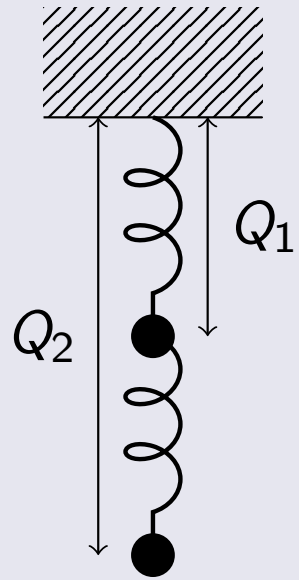
Feedback

Cyclic causal dependencies are also called feedback loops. Examples:

- Holding a microphone too close to a loudspeaker.
- Predator-prey relationships in biology.
- Computer programs running on a single core are *acyclic*; parallel programs running on multiple cores can be *cyclic*.

Example

- Two masses, connected by a spring, suspended from the ceiling by another spring.
- Vertical *equilibrium* positions Q_1 and Q_2.
- Q_1 causes Q_2.
- Q_2 causes Q_1.
- Example of a two-cycle: cannot be modeled with (causal) Bayesian network.



Causal modeling of feedback systems

Question: What are good mathematical representations of cyclic causal models?

Causal modeling of feedback systems

Question: What are good mathematical representations of cyclic causal models?

No consensus in the field...

- (Causal) Bayesian networks are *acyclic* by definition, and extending the definition to cyclic graphs [Schmidt & Murphy, 2009; Itani *et al.*, 2010] seems problematic.
- Extending the global Markov condition to cyclic models for *linear* models works [Spirtes, 1993], but nonlinear and discrete models yield problems [Spirtes, 1995; Pearl & Dechter, 1996; Neal 2000].
- Structural Causal Models have a "natural" extension to the cyclic case. But how to interpret these models in terms of a data generating process? Is this "the right" mathematical framework?

Causal modeling of feedback systems

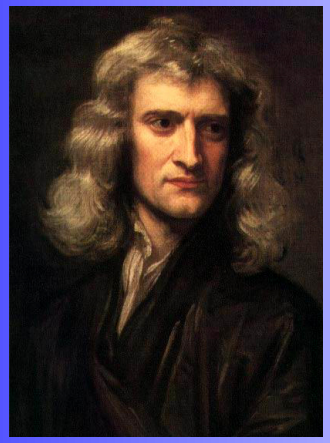
Question: What are good mathematical representations of cyclic causal models?

No consensus in the field...

- (Causal) Bayesian networks are *acyclic* by definition, and extending the definition to cyclic graphs [Schmidt & Murphy, 2009; Itani *et al.*, 2010] seems problematic.
- Extending the global Markov condition to cyclic models for *linear* models works [Spirtes, 1993], but nonlinear and discrete models yield problems [Spirtes, 1995; Pearl & Dechter, 1996; Neal 2000].
- Structural Causal Models have a "natural" extension to the cyclic case. But how to interpret these models in terms of a data generating process? Is this "the right" mathematical framework?

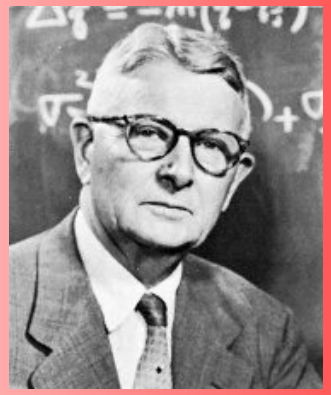
How do scientists usually model systems with feedback?

Two different worlds?



Ordinary Differential Equations

?



Structural Causal Models

From dynamical systems to causal models (in a nutshell)

1. Ordinary Differential Equations

$$\begin{cases} \dot{X} = -0.5X + Y, & X(0) = 1 \\ \dot{Y} = -X + 0.2Y, & Y(0) = 2 \end{cases}$$

From dynamical systems to causal models (in a nutshell)

1. Ordinary Differential Equations

$$\begin{cases} \dot{X} = -0.5X + Y, & X(0) = 1 \\ \dot{Y} = -X + 0.2Y, & Y(0) = 2 \end{cases}$$

2. Labeled Equilibrium Equations

$$\begin{cases} X: & 0 = -0.5X + Y \\ Y: & 0 = -X + 0.2Y \end{cases}$$

From dynamical systems to causal models (in a nutshell)

1. Ordinary Differential Equations

$$\begin{cases} \dot{X} = -0.5X + Y, & X(0) = 1 \\ \dot{Y} = -X + 0.2Y, & Y(0) = 2 \end{cases}$$

2. Labeled Equilibrium Equations

$$\begin{cases} X: & 0 = -0.5X + Y \\ Y: & 0 = -X + 0.2Y \end{cases}$$

3. Structural Causal Model

$$\begin{cases} X = 2Y \\ Y = 5X \end{cases}$$

From dynamical systems to causal models (in a nutshell)

1. Ordinary Differential Equations

$$\begin{cases} \dot{X} = -0.5X + Y, & X(0) = 1 \\ \dot{Y} = -X + 0.2Y, & Y(0) = 2 \end{cases}$$

2. Labeled Equilibrium Equations

$$\begin{cases} X: & 0 = -0.5X + Y \\ Y: & 0 = -X + 0.2Y \end{cases}$$

3. Structural Causal Model

$$\begin{cases} X = 2Y \\ Y = 5X \end{cases}$$

4. Dealing with Uncertainty

$$\begin{cases} X = 2Y + E_X \\ Y = 5X + E_Y \\ p(E_X, E_Y) = \dots \end{cases}$$

ODEs: Definition

Definition (ODE)

- An Ordinary Differential Equation model (ODE) is a dynamical system $\mathcal{D}$ described by D coupled first-order ordinary differential equations and initial condition $\mathbf{X}_0$:

$$\begin{cases} \dot{X}_i(t) := \frac{dX_i}{dt}(t) & = f_i(\mathbf{X}_{\mathrm{pa}_{\mathcal{D}}(i)}) \\ X_i(0) & = (\mathbf{X}_0)_i \end{cases} \qquad \forall i = 1, \ldots, D$$

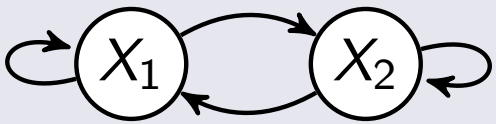
- $\mathrm{pa}_{\mathcal{D}}(i) \subseteq \{1, \ldots, D\}$ is the set of parents of variable X_i.
- Each $f_i : \mathcal{R}_{\mathrm{pa}_{\mathcal{D}}(i)} \to \mathcal{R}_i$ is a (sufficiently smooth) function.
- The *structure* can be represented as a directed graph $\mathcal{G}_{\mathcal{D}}$, with nodes $\{X_i\}_{i \in \mathcal{I}}$ and a directed edge $X_i \to X_j$ iff $\dot{X}_j$ depends on X_i.

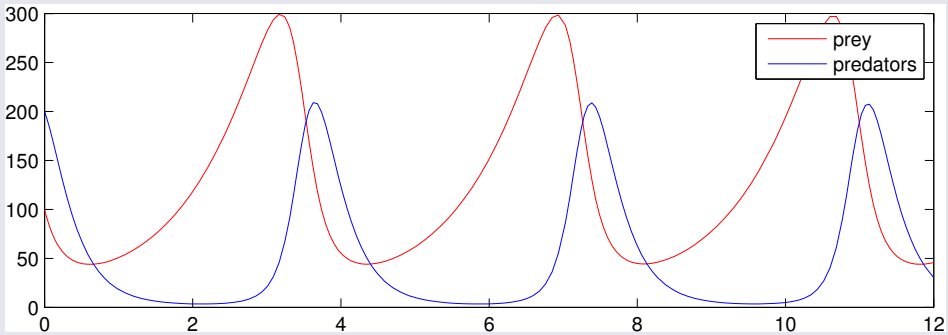
Example (Lotka-Volterra model)

- *Lotka-Volterra model*: well-known model from population biology
- Abundance of prey $X_1 \in [0, \infty)$ (e.g., rabbits)
- Abundance of predators $X_2 \in [0, \infty)$ (e.g., wolves)
- ODE $\mathcal{D}$:

$$\begin{cases} \dot{X}_1 &= X_1(\theta_{11} - \theta_{12}X_2) \\ \dot{X}_2 &= -X_2(\theta_{22} - \theta_{21}X_1) \end{cases} \qquad \begin{cases} X_1(0) = a \\ X_2(0) = b \end{cases}$$

- Graph $\mathcal{G}_{\mathcal{D}}$:





Perfect Interventions

- The dynamical system $\mathcal{D}$ is assumed to describe the "natural" or observational state of the system.
- Causal models aim to predict also the effects of interventions in which the system is actively perturbed from its natural state.
- Interventions can be modeled in different ways. Here we look at perfect interventions.

Definition (Perfect Interventions)

- The perfect intervention $\mathrm{do}(\mathbf{X}_I = \boldsymbol{\xi}_I)$ means that $\mathbf{X}_I$ is enforced to attain the value $\boldsymbol{\xi}_I$ for all times $t \in [0, \infty)$.
- This changes the ODE $\mathcal{D}$ into the intervened system $\mathcal{D}_{\mathrm{do}(\mathbf{X}_I = \boldsymbol{\xi}_I)}$:

$$\dot{X}_i(t) = \begin{cases} 0 & i \in I \\ f_i(\mathbf{X}_{\mathrm{pa}_{\mathcal{D}}(i)}) & i \in \mathcal{I} \setminus I, \end{cases} \qquad X_i(0) = \begin{cases} \xi_i & i \in I \\ (\mathbf{X}_0)_i & i \in \mathcal{I} \setminus I \end{cases}$$

Perfect Interventions in ODEs: Example

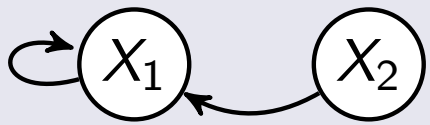
Example (Lotka-Volterra model)

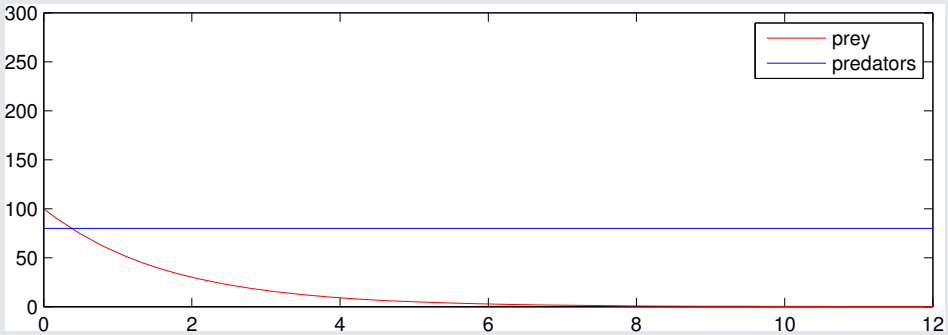
- $\mathcal{D}$: $\begin{cases} \dot{X}_1 &= X_1(\theta_{11} - \theta_{12}X_2) \\ \dot{X}_2 &= -X_2(\theta_{22} - \theta_{21}X_1) \end{cases}$ $\begin{cases} X_1(0) = a \\ X_2(0) = b \end{cases}$

- Perfect intervention $\mathrm{do}(X_2 = \xi_2)$: Monitor the abundance of wolves and make sure that the number equals the target value ξ_2 at all time.

- $\mathcal{D}_{\mathrm{do}(X_2=\xi_2)}$: $\begin{cases} \dot{X}_1 &= X_1(\theta_{11} - \theta_{12}X_2) \\ \dot{X}_2 &= \cancel{-X_2(\theta_{22} - \theta_{21}X_1)}\ 0 \end{cases}$ $\begin{cases} X_1(0) = a \\ X_2(0) = \cancel{b}\ \xi_2 \end{cases}$

- $\mathcal{G}_{\mathcal{D}_{\mathrm{do}(X_2=\xi_2)}}$:





ODEs: Stability

When studying the system in the limit $t \to \infty$, an important concept is stability:

Definition (Stability)

The ODE $\mathcal{D}$ is called stable if there exists a unique equilibrium state $\mathbf{X}^* \in \mathcal{R}_\mathcal{I}$ such that for any initial state $\mathbf{X}_0 \in \mathcal{R}_\mathcal{I}$, the system converges to this equilibrium state as $t \to \infty$:

$$\exists!_{\mathbf{X}^* \in \mathcal{R}_\mathcal{I}} \ \forall_{\mathbf{X}_0 \in \mathcal{R}_\mathcal{I}} : \lim_{t \to \infty} \mathbf{X}(t) = \mathbf{X}^*.$$

ODEs: Stability

When studying the system in the limit $t \to \infty$, an important concept is stability:

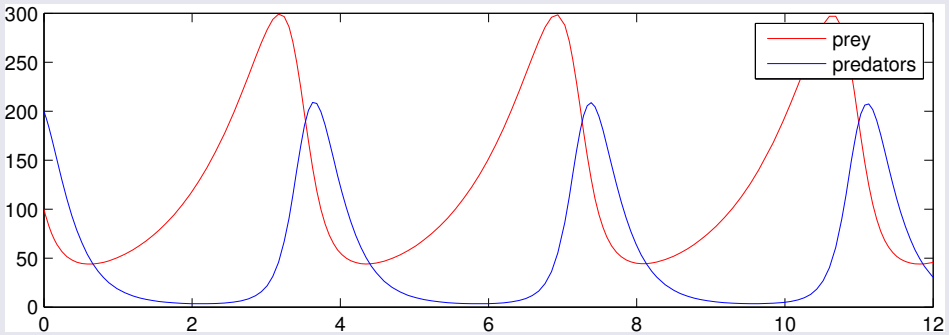
Definition (Stability)

The ODE $\mathcal{D}$ is called stable if there exists a unique equilibrium state $\mathbf{X}^* \in \mathcal{R}_\mathcal{I}$ such that for any initial state $\mathbf{X}_0 \in \mathcal{R}_\mathcal{I}$, the system converges to this equilibrium state as $t \to \infty$:

$$\exists!_{\mathbf{X}^* \in \mathcal{R}_\mathcal{I}} \; \forall_{\mathbf{X}_0 \in \mathcal{R}_\mathcal{I}} : \lim_{t \to \infty} \mathbf{X}(t) = \mathbf{X}^*.$$

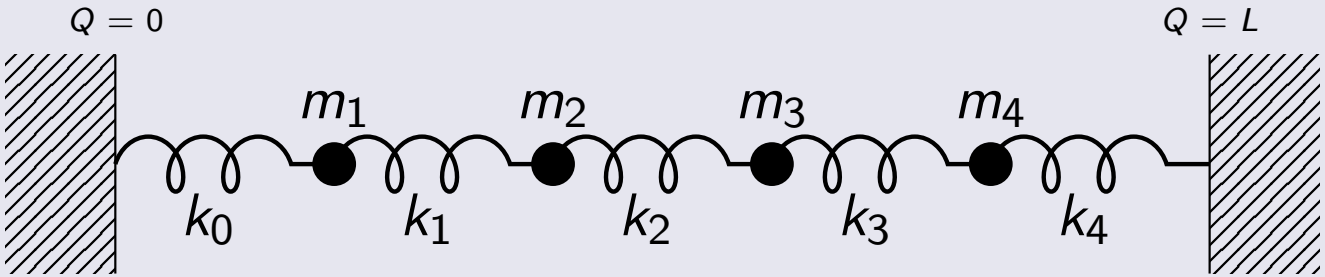
Example (Counter-example: Lotka-Volterra model)

The Lotka-Volterra model is *not* stable (it keeps oscillating).



Stability: Example

Example (Damped coupled harmonic oscillators)

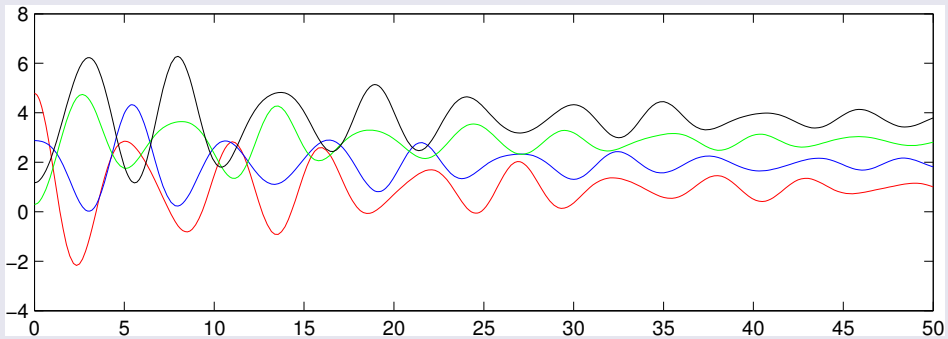


- Equations of motion (with $Q_0 := 0, Q_{D+1} := L$):

$$\dot{P}_i = k_i(Q_{i+1} - Q_i - l_i) - k_{i-1}(Q_i - Q_{i-1} - l_{i-1}) - \frac{b_i}{m_i} P_i$$

$$\dot{Q}_i = P_i / m_i$$

- Because of the friction, this system is stable (oscillations die out):



Equilibrium of Observational system

- Given an ODE $\mathcal{D}$:

$$\begin{cases} \dot{X}_i(t) & = f_i(\mathbf{X}_{\mathrm{pa}_{\mathcal{D}}(i)}) \\ X_i(0) & = (\mathbf{X}_0)_i \end{cases} \qquad \forall i \in \mathcal{I}$$

- At equilibrium, the rate of change of any variable is zero.
- This yields the following equilibrium equations:

$$0 = f_i(\mathbf{X}_{\mathrm{pa}_{\mathcal{D}}(i)}) \qquad \forall i \in \mathcal{I}$$

- This is a set of D coupled equations with unknowns $X_1, \ldots, X_D$.
- The stability assumption implies that there exists a unique solution $\mathbf{X}^*$ of the equilibrium equations.

Labeling Equilibrium Equations

- Note that the dynamical system contains "labels" for the equations: in case of an intervention on X_i, simply change the dynamical equation for $\dot{X}_i$.
- This information is lost when considering the equilibrium equations.
- In order to model perfect interventions, we introduce labels for the equilibrium equations.

Definition

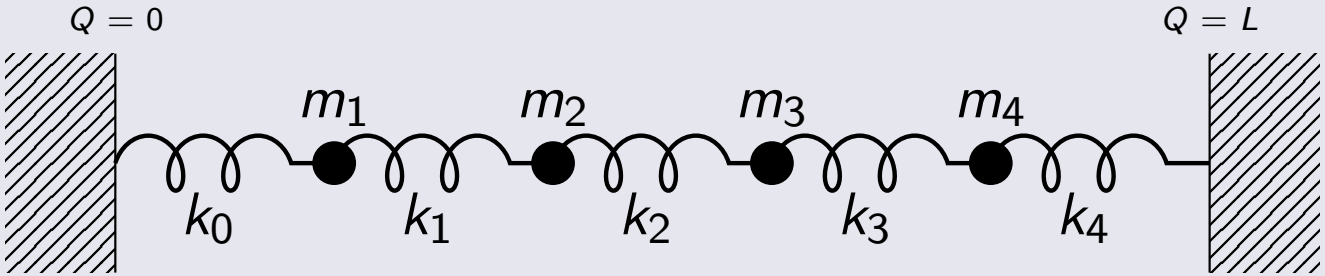
Given an ODE $\mathcal{D}$:

$$\begin{cases} \dot{X}_i(t) & = f_i(\mathbf{X}_{\mathrm{pa}_{\mathcal{D}}(i)}) \\ X_i(0) & = (\mathbf{X}_0)_i \end{cases} \qquad \forall i \in \mathcal{I}$$

its system $\mathcal{E}_{\mathcal{D}}$ of Labeled Equilibrium Equations (LEE) is given by:

$$i: \quad 0 = f_i(\mathbf{X}_{\mathrm{pa}_{\mathcal{D}}(i)}) \qquad \forall i \in \mathcal{I}$$

Example (Damped coupled harmonic oscillators)



- Equations of motion:

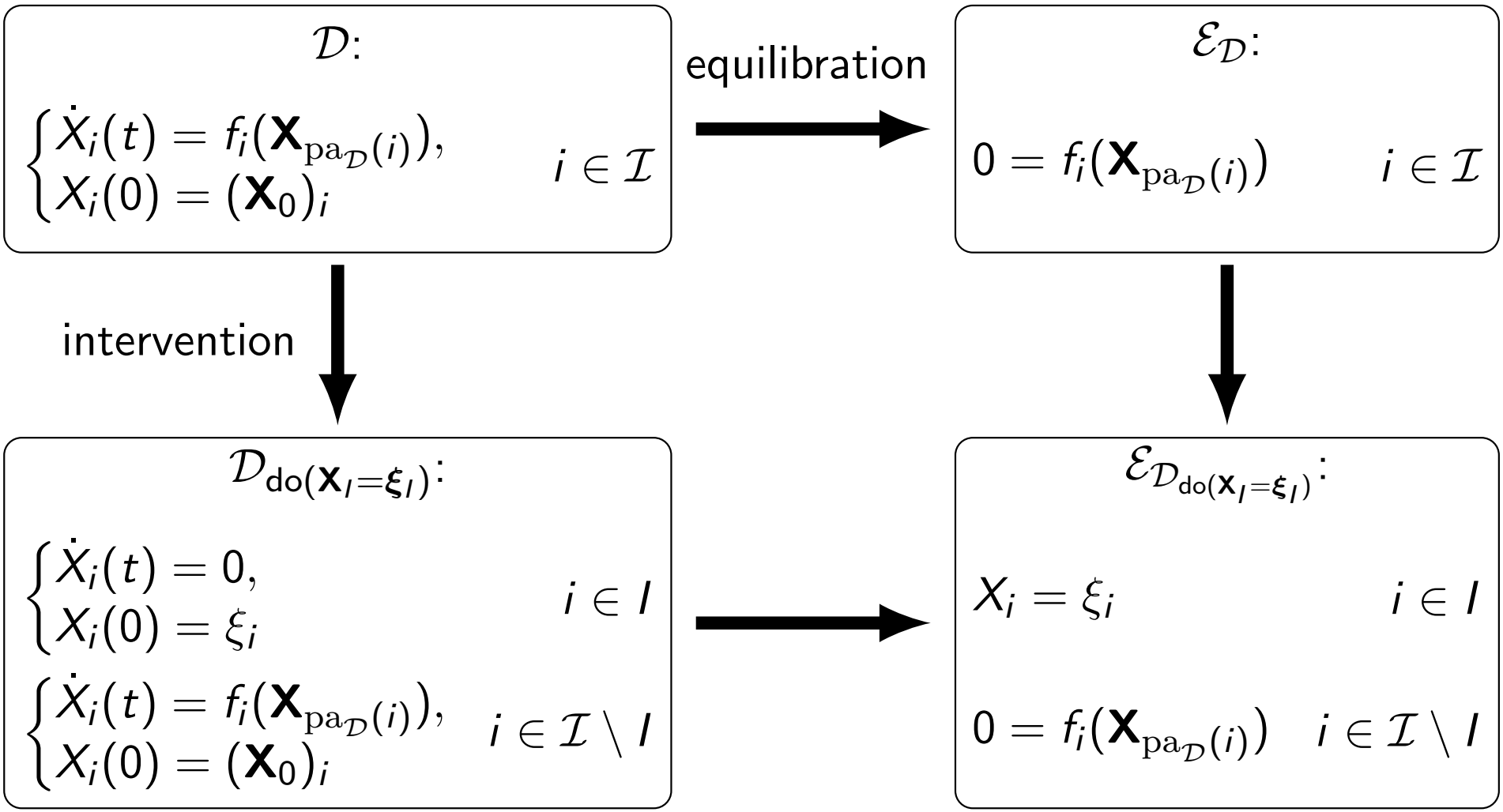
$$\dot{P}_i = k_i(Q_{i+1} - Q_i - l_i) - k_{i-1}(Q_i - Q_{i-1} - l_{i-1}) - \frac{b_i}{m_i}P_i$$

$$\dot{Q}_i = P_i/m_i$$

- The induced Labeled Equilibrium Equations are given by:

$$\mathcal{E}_i : \quad \begin{cases} 0 &= k_i(Q_{i+1} - Q_i - l_i) - k_{i-1}(Q_i - Q_{i-1} - l_{i-1}) - \frac{b_i}{m_i}P_i \\ 0 &= P_i \end{cases}$$

Equilibrium of Intervened systems



Definition

Given a system of Labeled Equilibrium Equations (LEE) $\mathcal{E}$:

$$i: \quad 0 = f_i(\mathbf{X}_{\mathrm{pa}_{\mathcal{E}}(i)}) \qquad \forall i \in \mathcal{I}$$

the induced SCM is obtained by solving each equation $\mathcal{E}_i$ for X_i in terms of the other variables:

$$X_i = g_i(\mathbf{X}_{\mathrm{pa}_{\mathcal{E}}(i)\setminus\{i\}}) \qquad \forall i \in \mathcal{I}$$

Note: This definition only makes sense if each labeled equilibrium equation $\mathcal{E}_i$ has a unique solution for X_i.

Induced SCM: Example

Example (Damped coupled harmonic oscillators)

- ODE $\mathcal{D}$:

$$\dot{P}_i = k_i(Q_{i+1} - Q_i - l_i) - k_{i-1}(Q_i - Q_{i-1} - l_{i-1}) - \frac{b_i}{m_i} P_i$$

$$\dot{Q}_i = P_i / m_i$$

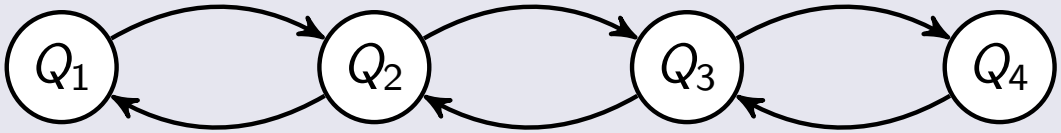
- Induced LEE $\mathcal{E}_{\mathcal{D}}$:

$$\mathcal{E}_i : \quad \begin{cases} 0 & = k_i(Q_{i+1} - Q_i - l_i) - k_{i-1}(Q_i - Q_{i-1} - l_{i-1}) - \frac{b_i}{m_i} P_i \\ 0 & = P_i \end{cases}$$

- Induced SCM $\mathcal{M}_{\mathcal{E}_{\mathcal{D}}}$:

$$Q_i = \frac{k_i(Q_{i+1} - l_i) + k_{i-1}(Q_{i-1} + l_{i-1})}{k_i + k_{i+1}}, \qquad P_i = 0.$$

- Graph of induced SCM $\mathcal{G}_{\mathcal{M}_{\mathcal{E}_{\mathcal{D}}}}$:

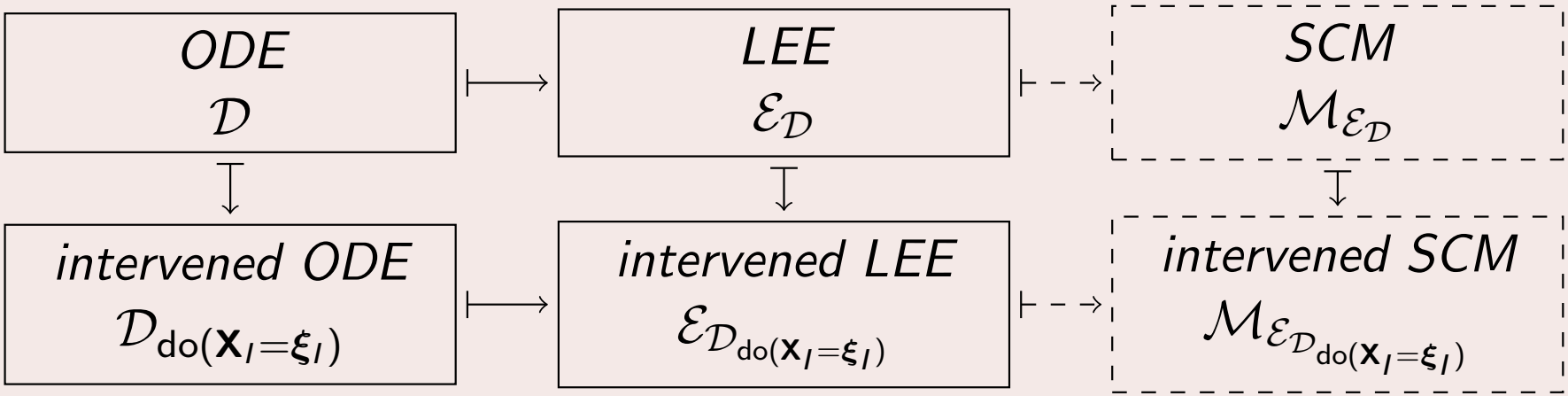


From ODEs to SCMs

Theorem (Mooij, Janzing, Schölkopf, UAI 2013)

Under certain stability conditions on the ODE $\mathcal{D}$ and the intervened ODE $\mathcal{D}_{\mathrm{do}(\mathbf{X}_I=\boldsymbol{\xi}_I)}$:

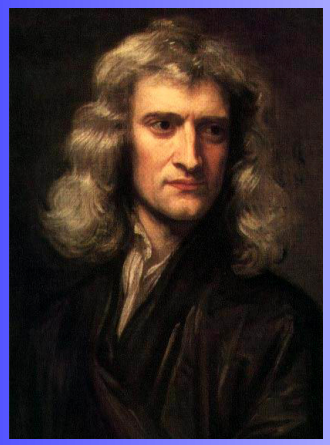
1. *The following diagram commutes:*

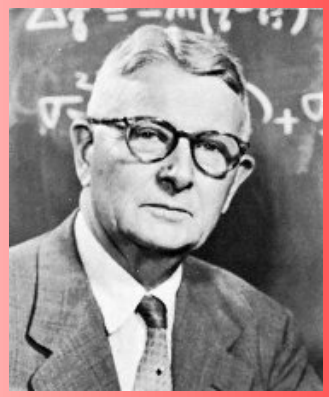


2. *If the intervened ODE $\mathcal{D}_{\mathrm{do}(\mathbf{X}_I=\boldsymbol{\xi}_I)}$ is stable, the induced intervened SCM $\mathcal{M}_{\mathcal{E}_{\mathcal{D}_{\mathrm{do}(\mathbf{X}_I=\boldsymbol{\xi}_I)}}}$ has a unique solution that coincides with the stable equilibrium of the intervened ODE $\mathcal{D}_{\mathrm{do}(\mathbf{X}_I=\boldsymbol{\xi}_I)}$.*

(Similar result was derived by [Dash, 2003] for the *acyclic* case.)

Conclusion: There is a bridge between the two worlds!





Ordinary Differential Equations → Structural Causal Models

Discussion

- We have shown one particular way in which structural causal models can be "derived".
- This shows that cyclic SCMs (and cyclic LEEs) are a very natural way to model causal systems with feedback.
- This work dealt with the deterministic case. Uncertainty can arise in several ways:
  1. uncertainty about (constant) parameters of the differential equations;
  2. uncertainty about the initial condition (in the case of constants of motion);
  3. latent variables (in the case of confounding).

  Dealing with uncertainty is work in progress (similar ideas, but more involved).

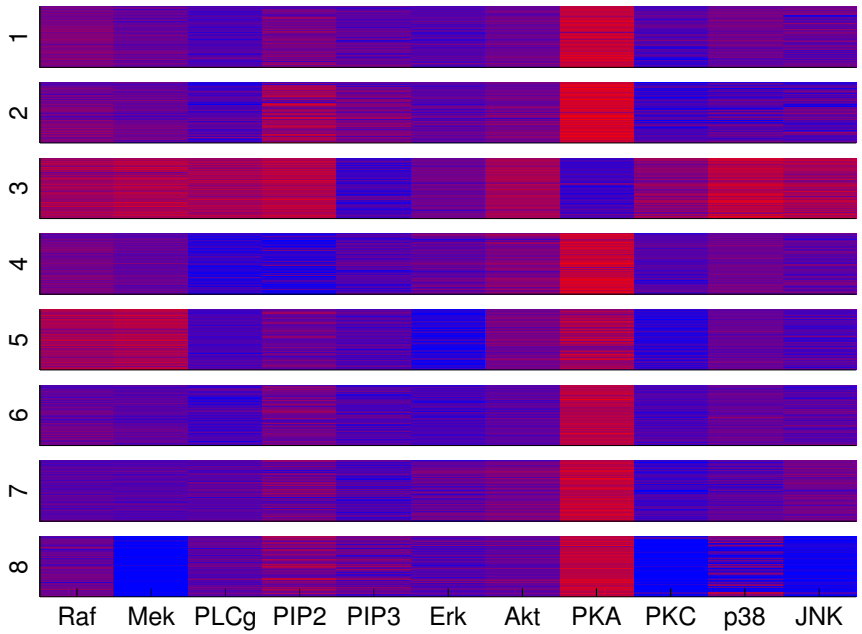
Part IV

Causal Discovery in case of feedback

Case study: Reconstructing a signalling network

Protein Abundance Data:

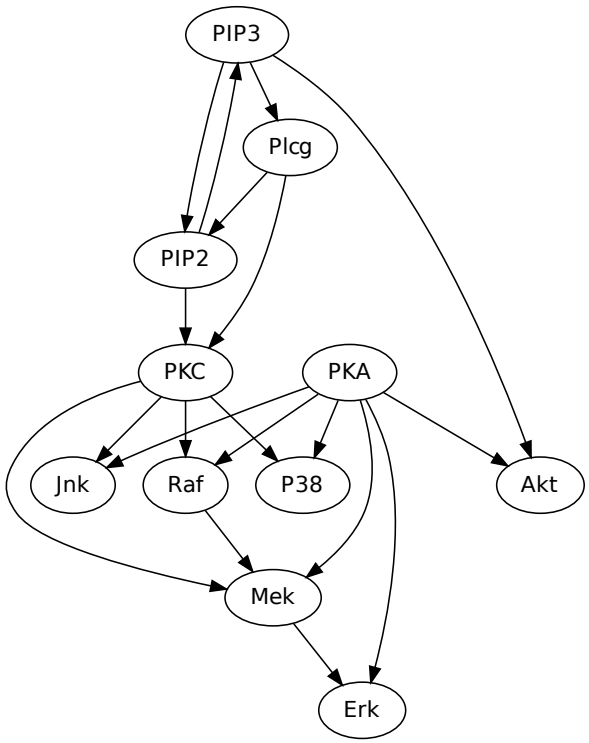
[Sachs et al., 2005]



Condition	Reagent	Intervention
1	-	observational
2	Akt-inhibitor	inhibits AKT activity
3	G0076	inhibits PKC activity
4	Psitectorigenin	inhibits PIP2 abundance
5	U0126	inhibits MEK activity
6	LY294002	inhibits PIP2/PIP3 activity
7	PMA	activates PKC activity
8	β2CAMP	activates PKA activity

Causal Mechanism:

("Signalling network")



(depicted here: "consensus" network)

Motivation

Good test case for causal discovery methods, because:

- High-quality data:
  - Single-cell measurements
  - Many data points (about 10^4)
  - Small measurement noise
- Much knowledge about "ground truth"
- Possibly important applications in cancer medicine

Motivation

Good test case for causal discovery methods, because:

- High-quality data:
  - Single-cell measurements
  - Many data points (about 10^4)
  - Small measurement noise
- Much knowledge about "ground truth"
- Possibly important applications in cancer medicine

Good results obtained by [Sachs et al., 2005] assuming acyclicity and causal sufficiency using Bayesian network learning with discretized data. But...

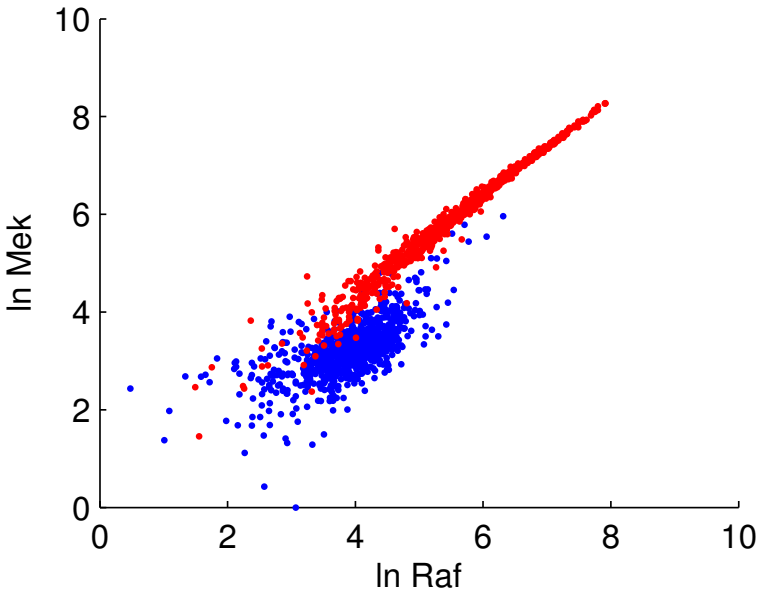
- Data shows evidence of feedback loops (cycles).
- No suitable cyclic causal discovery methods available (but: [Itani et al., 2010, Schmidt and Murphy, 2009] for discretized data).

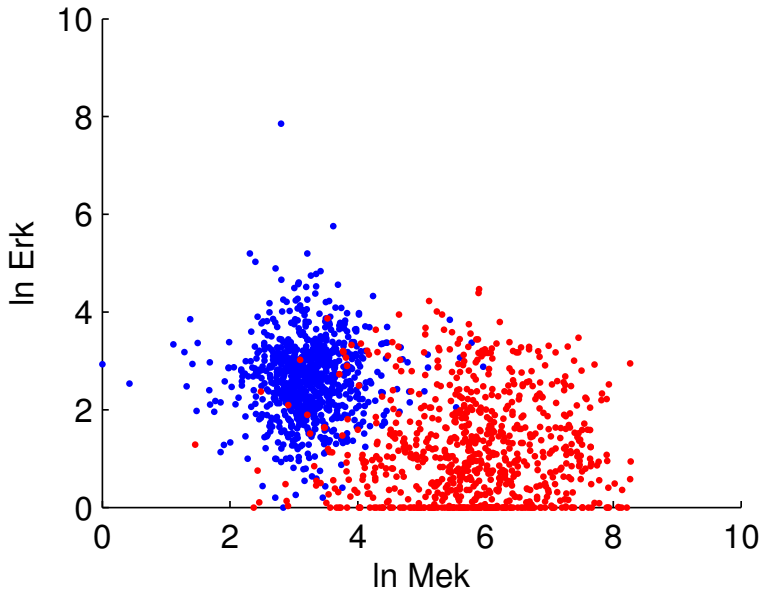
The importance of modeling feedback

- Feedback plays an important role in many biological systems.
- Ignoring feedback may lead to unwanted surprises, e.g., [Hall-Jackson et al., 1999]:

  "Here, we describe a compound (ZM 336372) that is a potent inhibitor of the protein kinase c-Raf in vitro. Paradoxically, however, incubation of mammalian cells with this compound induces an enormous activation of c-Raf and the B-Raf isoform (measured in the absence of the drug), suggesting that a feedback control loop exists by which Raf isoforms suppress their own activation. This unexpected finding may explain why ZM 336372 does not reverse the phenotype of Ras-transformed cell lines, and suggests that inhibition of the kinase activity of Raf might not be a good approach for the development of an anti-cancer drug."

The data (scatter plots)





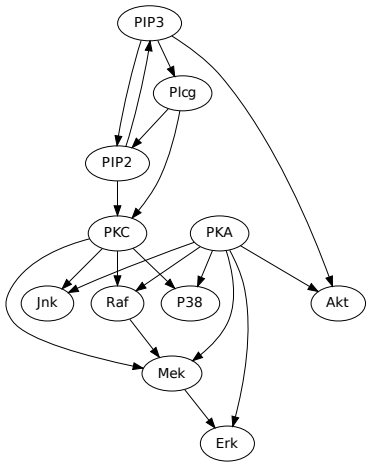
condition 1 (observational), condition 5 (MEK inhibitor)

Note:

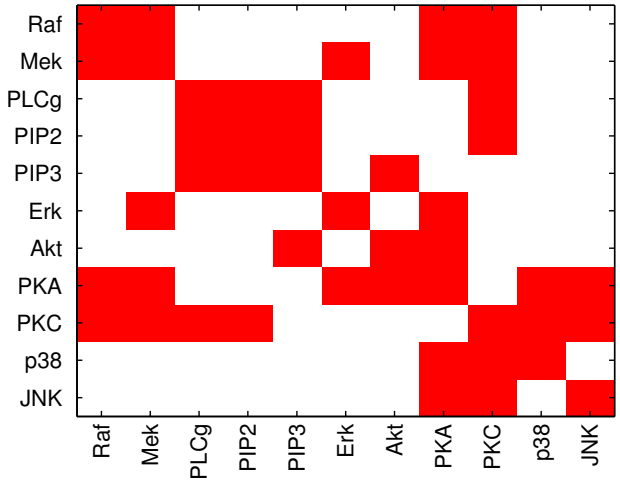
- Noise can be very small (so observation noise is small)
- Strong correlation between Raf and Mek (consensus: Raf $\rightarrow$ Mek)
- Evidence for feedback (intervening on Mek changes Raf)
- No dependence between Mek and Erk (consensus: Mek $\rightarrow$ Erk)

Challenge: faithfulness violations

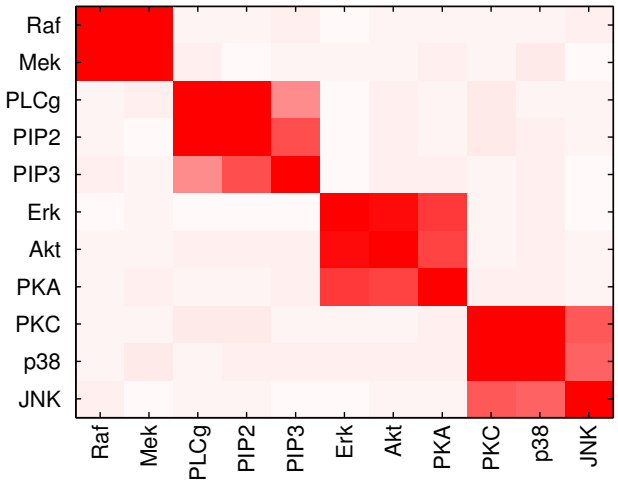
Consensus causal graph



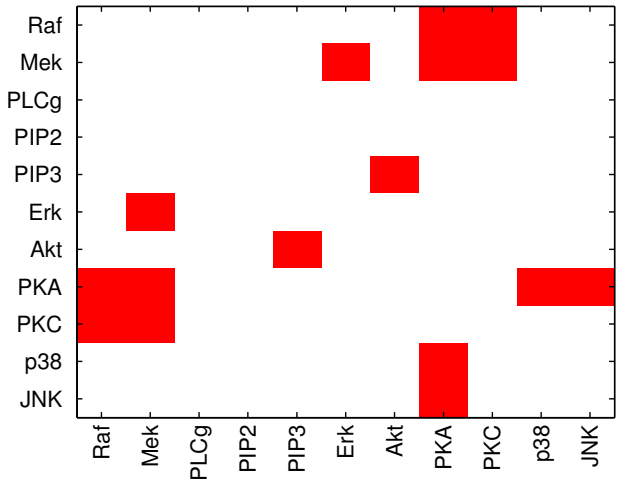
Expected correlations



Measured correlations



Faithfulness violations



This means that we need to combine observational and interventional data.

Goal

The goal of this work:

Perform more sophisticated causal analysis of the data by...

Goal

The goal of this work:

Perform more sophisticated causal analysis of the data by...

- Modeling feedback loops;
- Modeling the interventions in a realistic way;
- Using continuous data instead of a coarsely discretized version, allowing for nonlinear causal mechanisms;

Goal

The goal of this work:

Perform more sophisticated causal analysis of the data by...

- Modeling feedback loops;
- Modeling the interventions in a realistic way;
- Using continuous data instead of a coarsely discretized version, allowing for nonlinear causal mechanisms;

...and by doing so, arrive at a more realistic reconstruction of the signalling network than [Sachs et al., 2005] originally obtained by using (acyclic) discrete-valued Bayesian networks.

Basic Modeling Assumptions

We make the following assumptions for modeling the data:

Causal modeling assumptions

- No time-series data: the cells have reached equilibrium when the measurements are performed;

Basic Modeling Assumptions

We make the following assumptions for modeling the data:

Causal modeling assumptions

- No time-series data: the cells have reached equilibrium when the measurements are performed;
- The equilibrium abundances $\mathbf{X}$ are governed by a (possibly cyclic) Structural Causal Model

$$X_i = f_i(\mathbf{X}_{\mathrm{pa}(i)}, E_i), \qquad i = 1, \ldots, D;$$

Basic Modeling Assumptions

We make the following assumptions for modeling the data:

Causal modeling assumptions

- No time-series data: the cells have reached equilibrium when the measurements are performed;
- The equilibrium abundances $\mathbf{X}$ are governed by a (possibly cyclic) Structural Causal Model

$$X_i = f_i(\mathbf{X}_{\mathrm{pa}(i)}, E_i), \qquad i = 1, \ldots, D;$$

- $\mathbf{E}$ is constant in time but varies over cells;

Basic Modeling Assumptions

We make the following assumptions for modeling the data:

Causal modeling assumptions

- No time-series data: the cells have reached equilibrium when the measurements are performed;
- The equilibrium abundances $\mathbf{X}$ are governed by a (possibly cyclic) Structural Causal Model

$$X_i = f_i(\mathbf{X}_{\mathrm{pa}(i)}, E_i), \qquad i = 1, \ldots, D;$$

- $\mathbf{E}$ is constant in time but varies over cells;
- The reagents may change the structural equations locally;

Basic Modeling Assumptions

We make the following assumptions for modeling the data:

Causal modeling assumptions

- No time-series data: the cells have reached equilibrium when the measurements are performed;
- The equilibrium abundances $\mathbf{X}$ are governed by a (possibly cyclic) Structural Causal Model

$$X_i = f_i(\mathbf{X}_{\mathrm{pa}(i)}, E_i), \qquad i = 1, \ldots, D;$$

- $\mathbf{E}$ is constant in time but varies over cells;
- The reagents may change the structural equations locally;
- Causal sufficiency (all E_i are jointly independent).

Induced distributions of cyclic SCMs

Lemma (Induced distribution of cyclic SCMs)

- *If for each value of the noise* $\mathbf{E}$*, there exists a unique solution* $\mathbf{X}(\mathbf{E})$ *of the structural equations* $\{X_i = f_i(\mathbf{X}_{\mathrm{pa}(i)}, E_i)\}$*, a SCM induces a unique observational distribution* $p(\mathbf{X})$*.*
- *In the acyclic case, that assumption is automatically satisfied.*
- *If the mapping* $\mathbf{E} \mapsto \mathbf{X}(\mathbf{E})$ *is invertable, the induced density satisfies:*

$$p(\mathbf{X}) = p_{\mathbf{E}}\big(\mathbf{E}(\mathbf{X})\big) \left| \det \frac{\partial \mathbf{E}}{\partial \mathbf{X}} \right| .$$

This means that under these assumptions, we can write down the likelihood of the data as a function of the model parameters.

Modeling Interventions with a SCM

Following [Sachs et al., 2005], we distinguish two types of interventions:

- abundance interventions that alter the abundance of some compound;
- activity interventions that alter the activity of some compound.

Modeling Interventions with a SCM

Following [Sachs et al., 2005], we distinguish two types of interventions:

- abundance interventions that alter the abundance of some compound;
- activity interventions that alter the activity of some compound.

Here, we propose to model these interventions as follows:

- An abundance intervention on X_i replaces the structural equation for X_i with $X_i = \xi_i$ (standard "perfect" interventions);
- An activity intervention on X_i replaces the causal mechanisms for its children X_j, $i \in \mathrm{pa}(j)$ by other causal mechanisms $X_j = \tilde{f}_j(\mathbf{X}_{\mathrm{pa}(j)}, E_j)$.

Modeling Interventions with a SCM

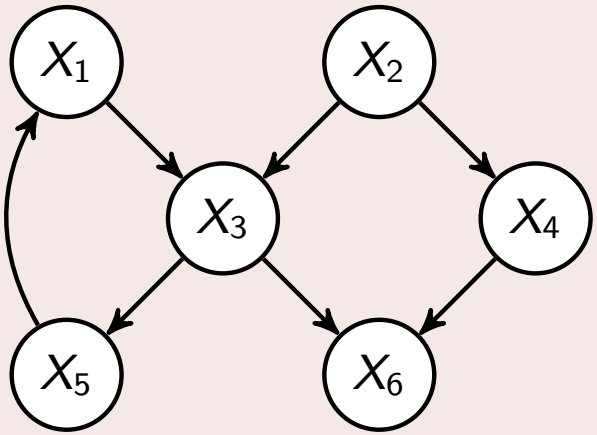
Following [Sachs et al., 2005], we distinguish two types of interventions:

- abundance interventions that alter the abundance of some compound;
- activity interventions that alter the activity of some compound.

Here, we propose to model these interventions as follows:

- An abundance intervention on X_i replaces the structural equation for X_i with $X_i = \xi_i$ (standard "perfect" interventions);
- An activity intervention on X_i replaces the causal mechanisms for its children X_j, $i \in \mathrm{pa}(j)$ by other causal mechanisms $X_j = \tilde{f}_j(\mathbf{X}_{\mathrm{pa}(j)}, E_j)$.

Example: No intervention



$$
\begin{aligned}
X_1 &= f_1(X_5, E_1) \\
X_2 &= f_2(E_2) \\
X_3 &= f_3(X_1, X_2, E_3) \\
X_4 &= f_4(X_2, E_4) \\
X_5 &= f_5(X_3, E_5) \\
X_6 &= f_6(X_3, X_4, E_6)
\end{aligned}
$$

Modeling Interventions with a SCM

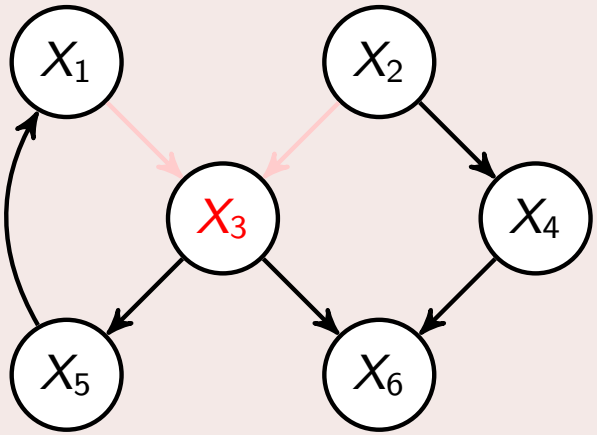
Following [Sachs et al., 2005], we distinguish two types of interventions:

- abundance interventions that alter the abundance of some compound;
- activity interventions that alter the activity of some compound.

Here, we propose to model these interventions as follows:

- An abundance intervention on X_i replaces the structural equation for X_i with $X_i = \xi_i$ (standard "perfect" interventions);
- An activity intervention on X_i replaces the causal mechanisms for its children X_j, $i \in \mathrm{pa}(j)$ by other causal mechanisms $X_j = \tilde{f}_j(\mathbf{X}_{\mathrm{pa}(j)}, E_j)$.

Example: Abundance intervention on X_3



$$
\begin{aligned}
X_1 &= f_1(X_5, E_1) \\
X_2 &= f_2(E_2) \\
X_3 &= \cancel{f_3(X_1, X_2, E_3)} \quad \xi_3 \\
X_4 &= f_4(X_2, E_4) \\
X_5 &= f_5(X_3, E_5) \\
X_6 &= f_6(X_3, X_4, E_6)
\end{aligned}
$$

Modeling Interventions with a SCM

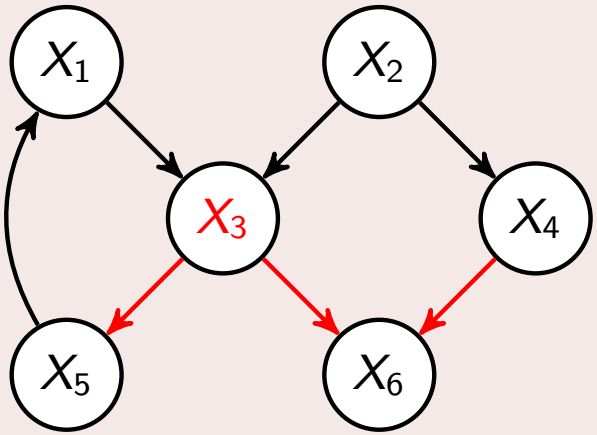
Following [Sachs et al., 2005], we distinguish two types of interventions:

- abundance interventions that alter the abundance of some compound;
- activity interventions that alter the activity of some compound.

Here, we propose to model these interventions as follows:

- An abundance intervention on X_i replaces the structural equation for X_i with $X_i = \xi_i$ (standard "perfect" interventions);
- An activity intervention on X_i replaces the causal mechanisms for its children X_j, $i \in \mathrm{pa}(j)$ by other causal mechanisms $X_j = \tilde{f}_j(\mathbf{X}_{\mathrm{pa}(j)}, E_j)$.

Example: Activity intervention on X_3



$$X_1 = f_1(X_5, E_1)$$
$$X_2 = f_2(E_2)$$
$$X_3 = f_3(X_1, X_2, E_3)$$
$$X_4 = f_4(X_2, E_4)$$
$$X_5 = \cancel{f_5(X_3, E_5)} \quad \tilde{f}_5(X_3, E_5)$$
$$X_6 = \cancel{f_6(X_3, X_4, E_6)} \quad \tilde{f}_6(X_3, X_4, E_6)$$

Algorithm: Score-based approach

- Use approximate Bayesian model selection in a multi-task learning setting to estimate the posterior probability of a putative causal graph $\mathcal{G}$, given the data (and prior assumptions).

Algorithm: Score-based approach

- Use approximate Bayesian model selection in a multi-task learning setting to estimate the posterior probability of a putative causal graph $\mathcal{G}$, given the data (and prior assumptions).
- Given a hypothetical causal graph $\mathcal{G}$, numerically optimize the posterior with respect to the parameters.

Algorithm: Score-based approach

- Use approximate Bayesian model selection in a multi-task learning setting to estimate the posterior probability of a putative causal graph $\mathcal{G}$, given the data (and prior assumptions).
- Given a hypothetical causal graph $\mathcal{G}$, numerically optimize the posterior with respect to the parameters.
- Employ Laplace approximation to approximate the evidence (marginal likelihood) for that causal graph $\mathcal{G}$.

Algorithm: Score-based approach

- Use approximate Bayesian model selection in a multi-task learning setting to estimate the posterior probability of a putative causal graph $\mathcal{G}$, given the data (and prior assumptions).
- Given a hypothetical causal graph $\mathcal{G}$, numerically optimize the posterior with respect to the parameters.
- Employ Laplace approximation to approximate the evidence (marginal likelihood) for that causal graph $\mathcal{G}$.
- Number of possible causal graphs $\mathcal{G}$ for 11 variables:

  $$\begin{aligned} &316034593964189176074 25 && \text{(acyclic)} \\ &1298074214633706907132624082305024 && \text{(cyclic).} \end{aligned}$$

  Use local search to explore posterior distribution over causal graphs.

Algorithm: Score-based approach

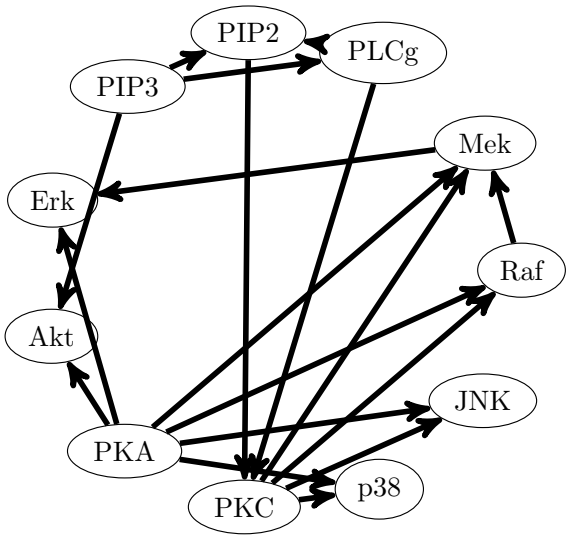
- Use approximate Bayesian model selection in a multi-task learning setting to estimate the posterior probability of a putative causal graph $\mathcal{G}$, given the data (and prior assumptions).
- Given a hypothetical causal graph $\mathcal{G}$, numerically optimize the posterior with respect to the parameters.
- Employ Laplace approximation to approximate the evidence (marginal likelihood) for that causal graph $\mathcal{G}$.
- Number of possible causal graphs $\mathcal{G}$ for 11 variables:

  $$\begin{aligned} &316034593964189176074 25 && \text{(acyclic)} \\ &1298074214633706907132624082305024 && \text{(cyclic).} \end{aligned}$$

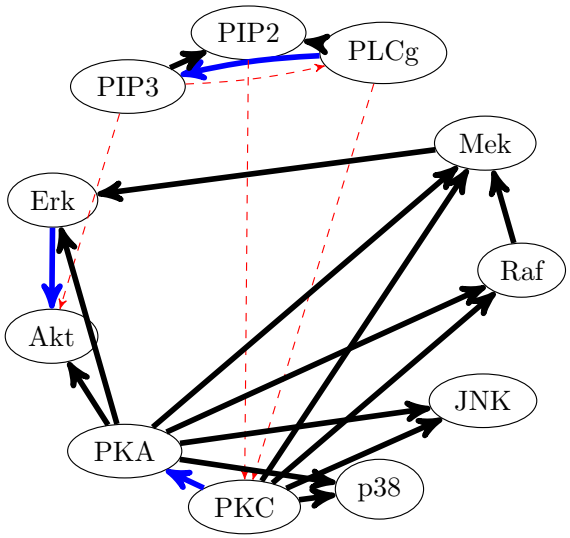
  Use local search to explore posterior distribution over causal graphs.
- Stability selection [Meinshausen *et al.*, 2010] to identify stable causal relations.

Comparison with ground truth (max. 17 edges, acyclic)

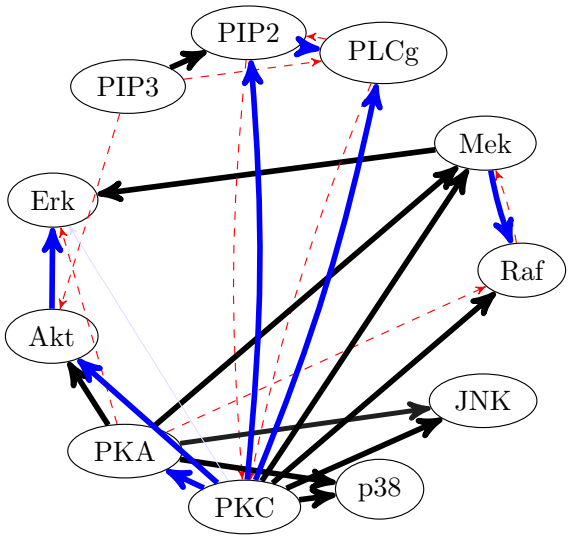
For comparison with the consensus model and the reconstructed model by Sachs *et al.*, we constrain the number of edges:



Consensus



Sachs *et al.*

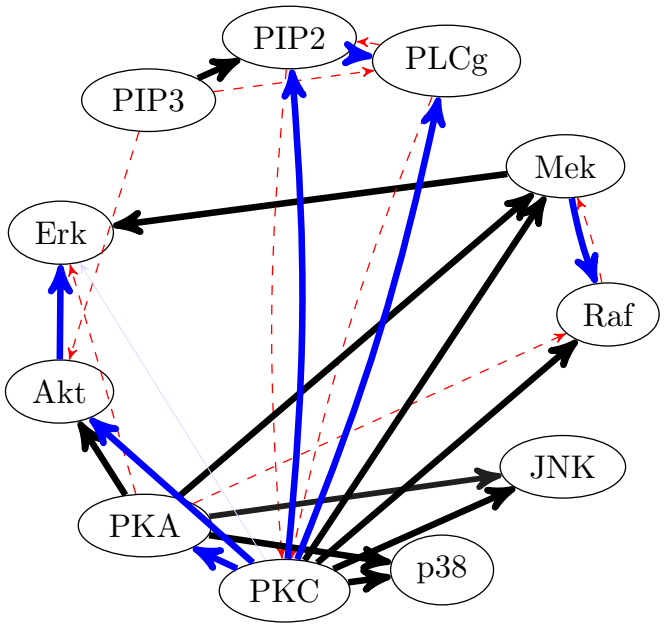


This work

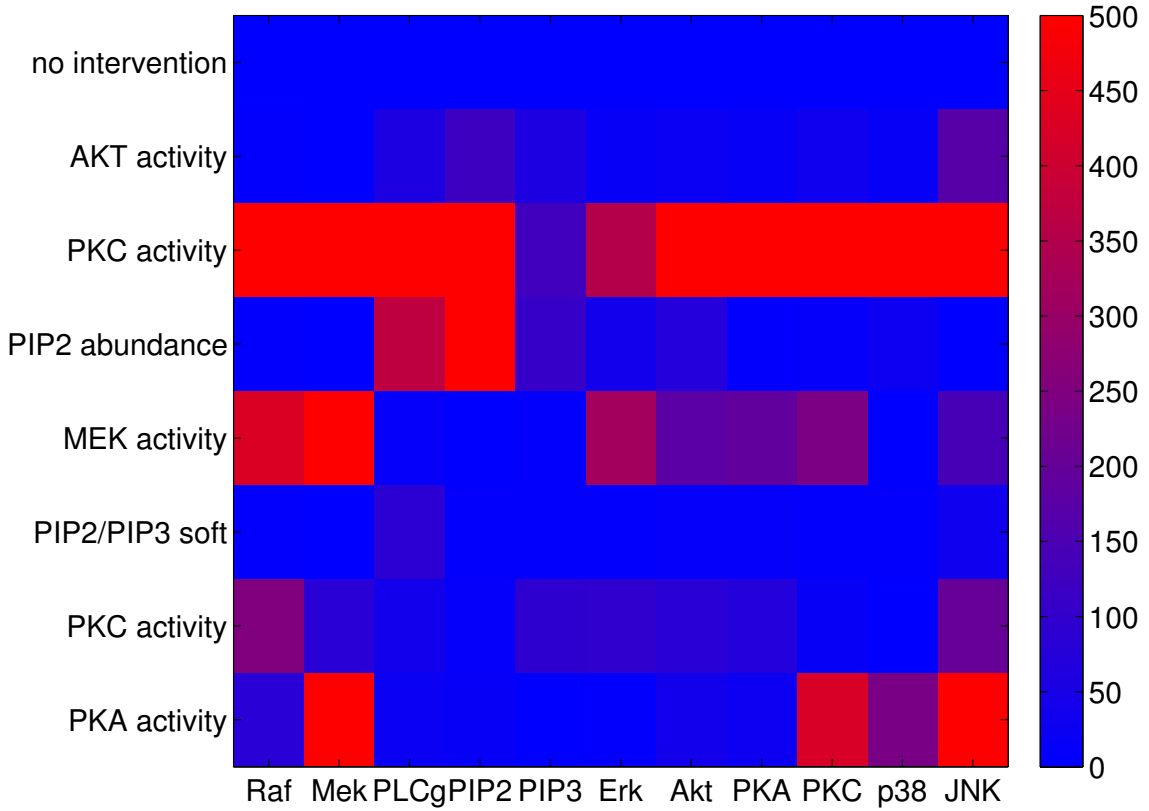
Black: expected, Blue: novel findings, Red dashed: missing.

Our acyclic, strongly regularised, result deviates more from the "consensus" network. Actually seems to be good news!

Comparison with ground truth (max. 17 edges, acyclic)



This work



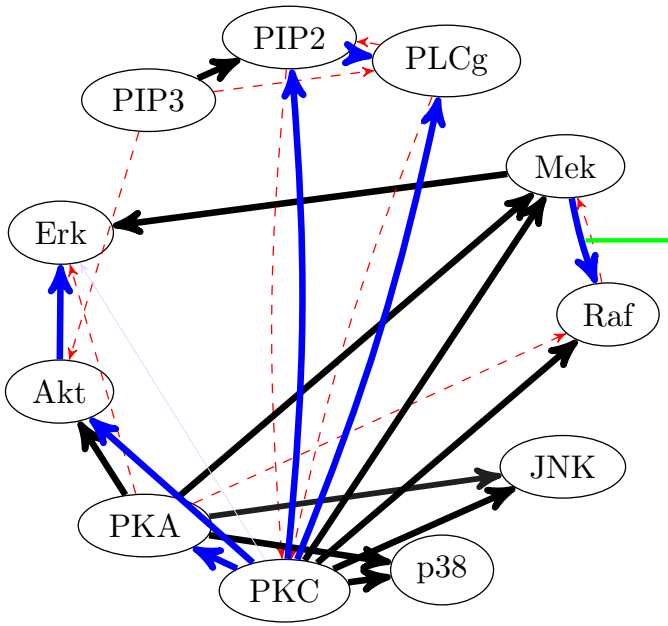
KS test w.r.t. observational data

Black: expected,

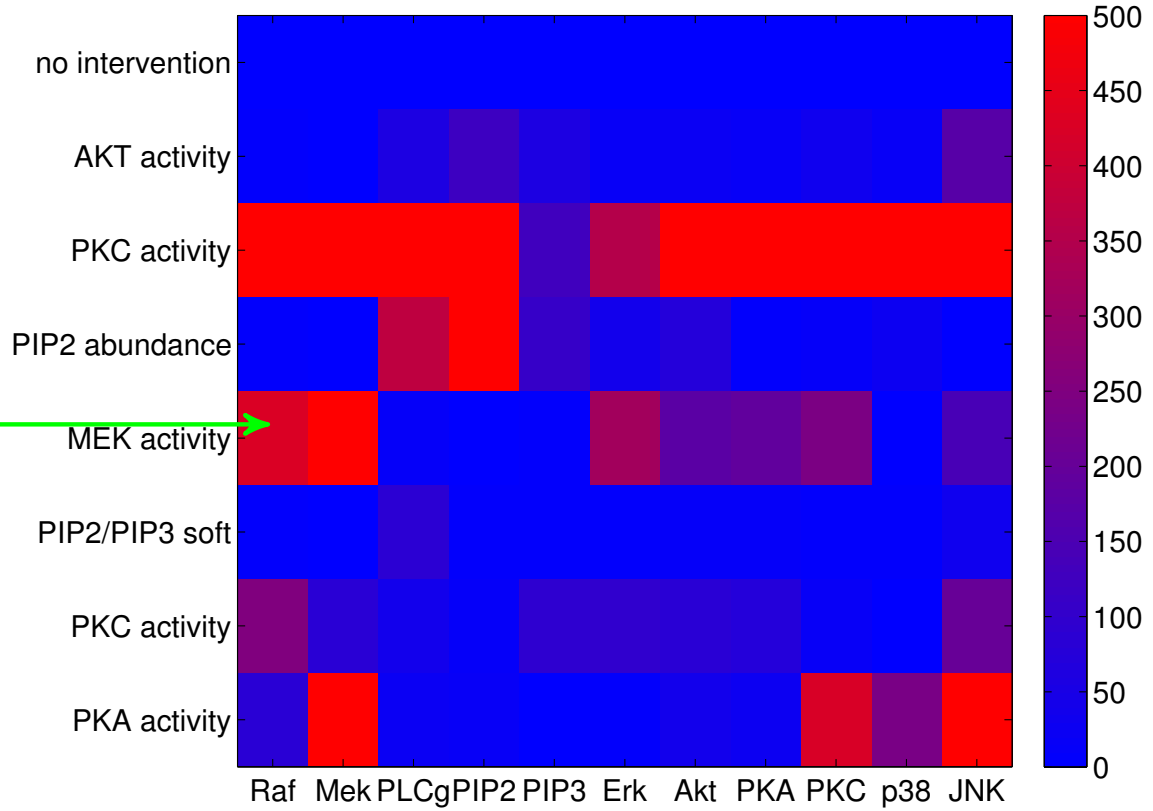
Blue: novel findings,

Red dashed: missing.

Comparison with ground truth (max. 17 edges, acyclic)



This work



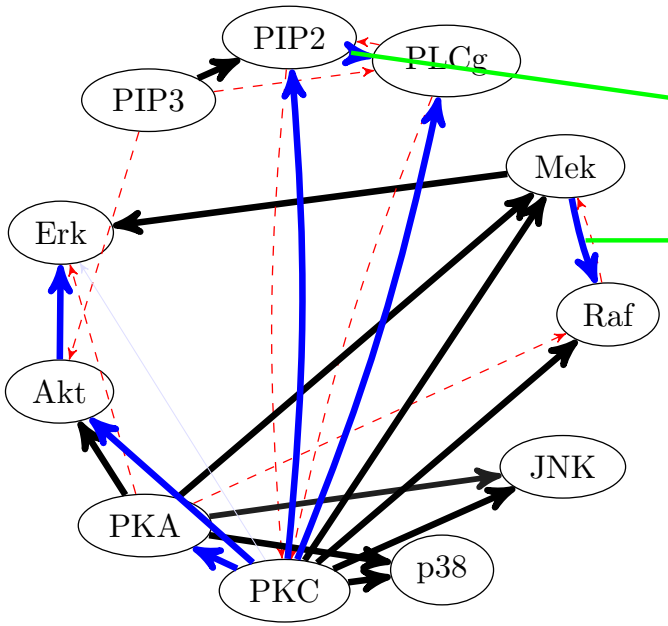
KS test w.r.t. observational data

Black: expected,

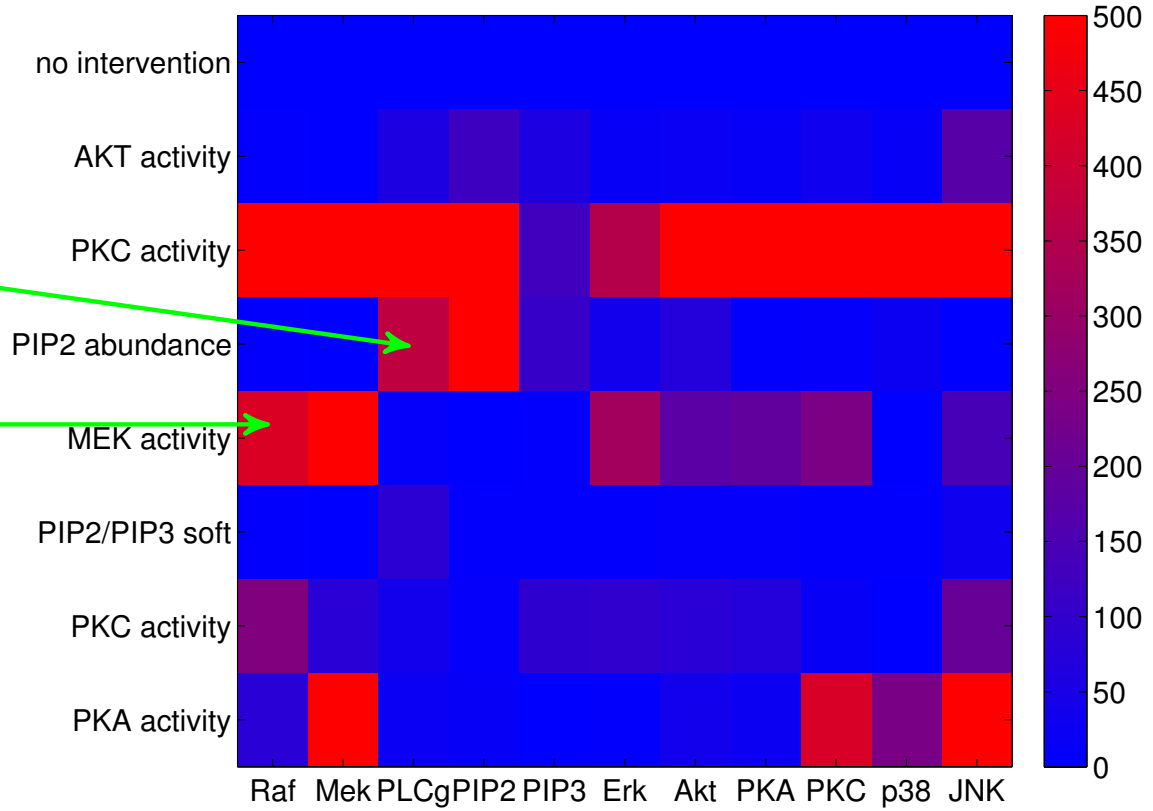
Blue: novel findings,

Red dashed: missing.

Comparison with ground truth (max. 17 edges, acyclic)



This work



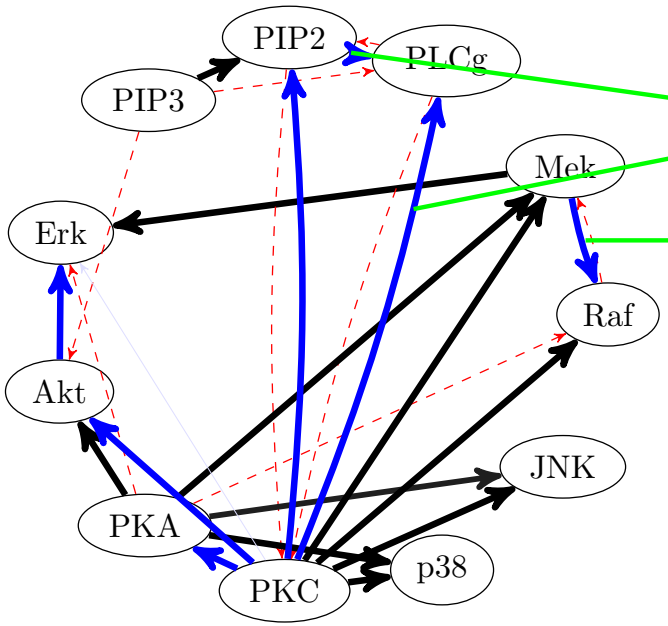
KS test w.r.t. observational data

Black: expected,

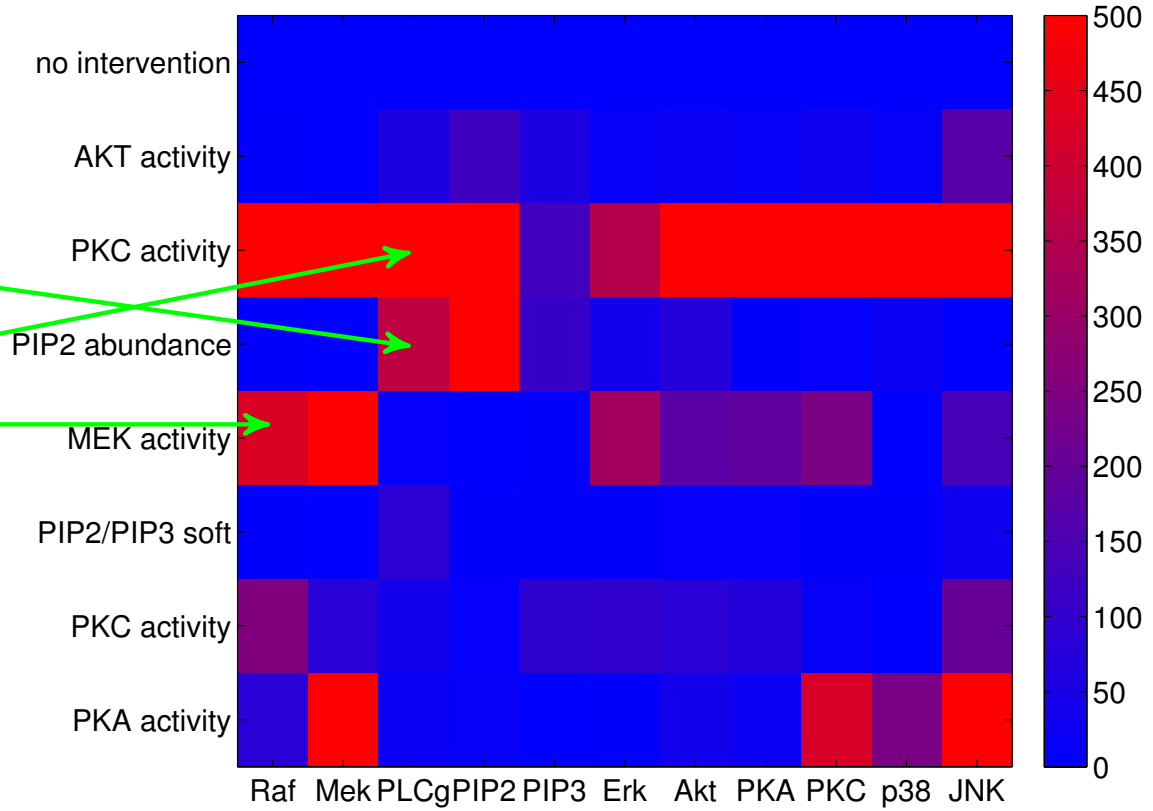
Blue: novel findings,

Red dashed: missing.

Comparison with ground truth (max. 17 edges, acyclic)



This work



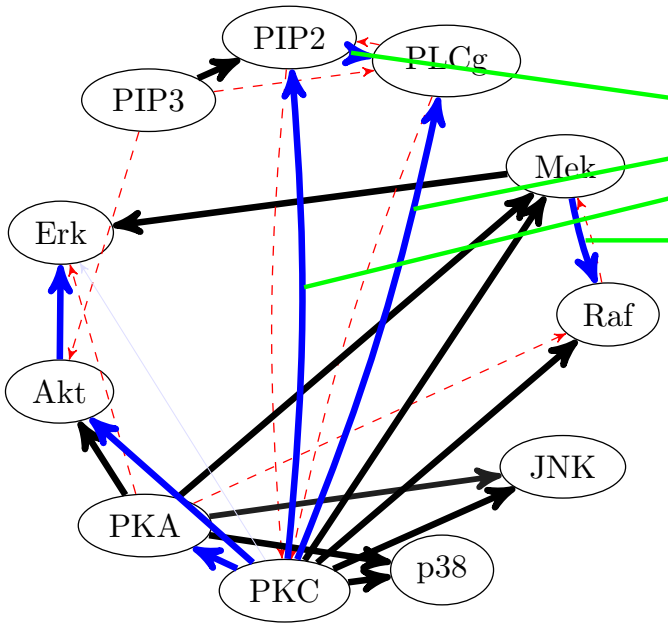
KS test w.r.t. observational data

Black: expected,

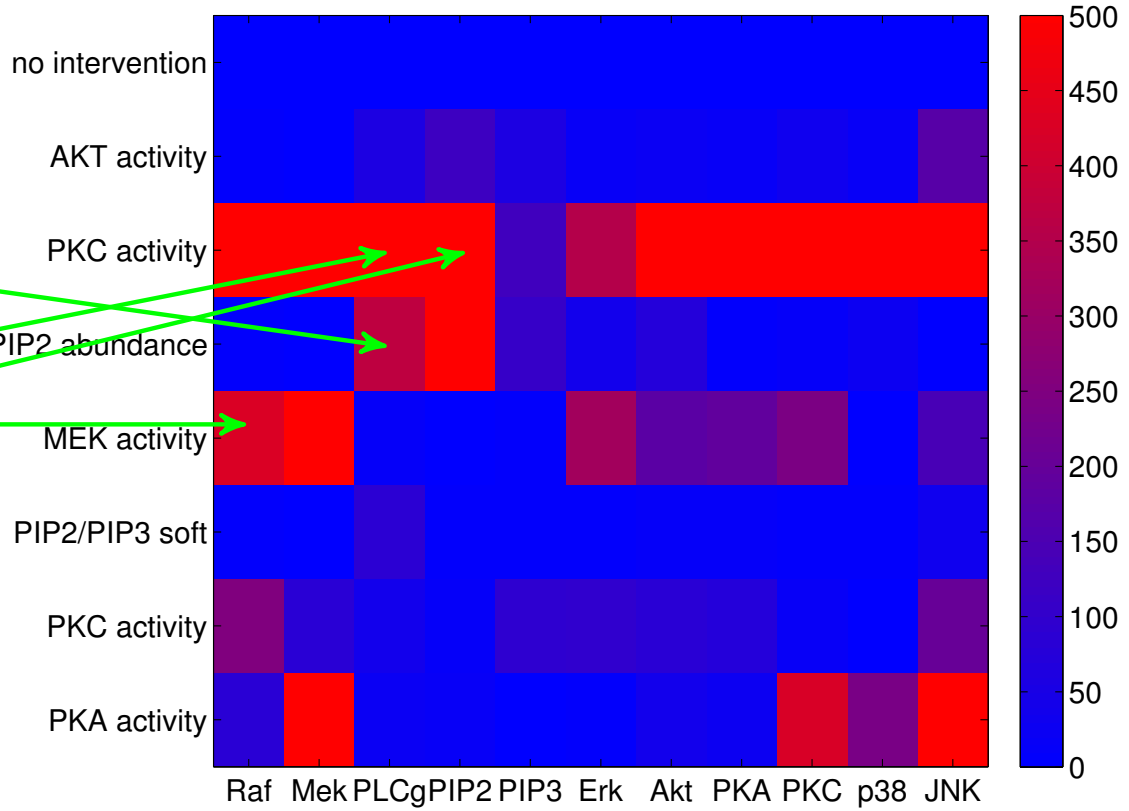
Blue: novel findings,

Red dashed: missing.

Comparison with ground truth (max. 17 edges, acyclic)



This work



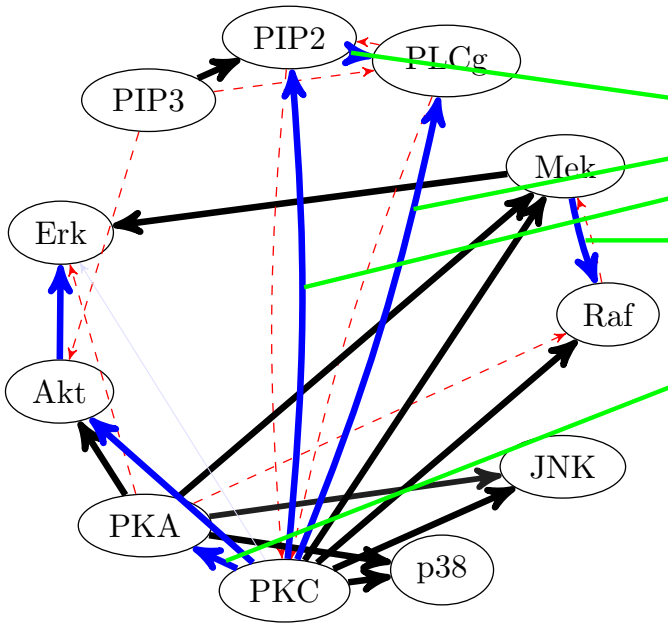
KS test w.r.t. observational data

Black: expected,

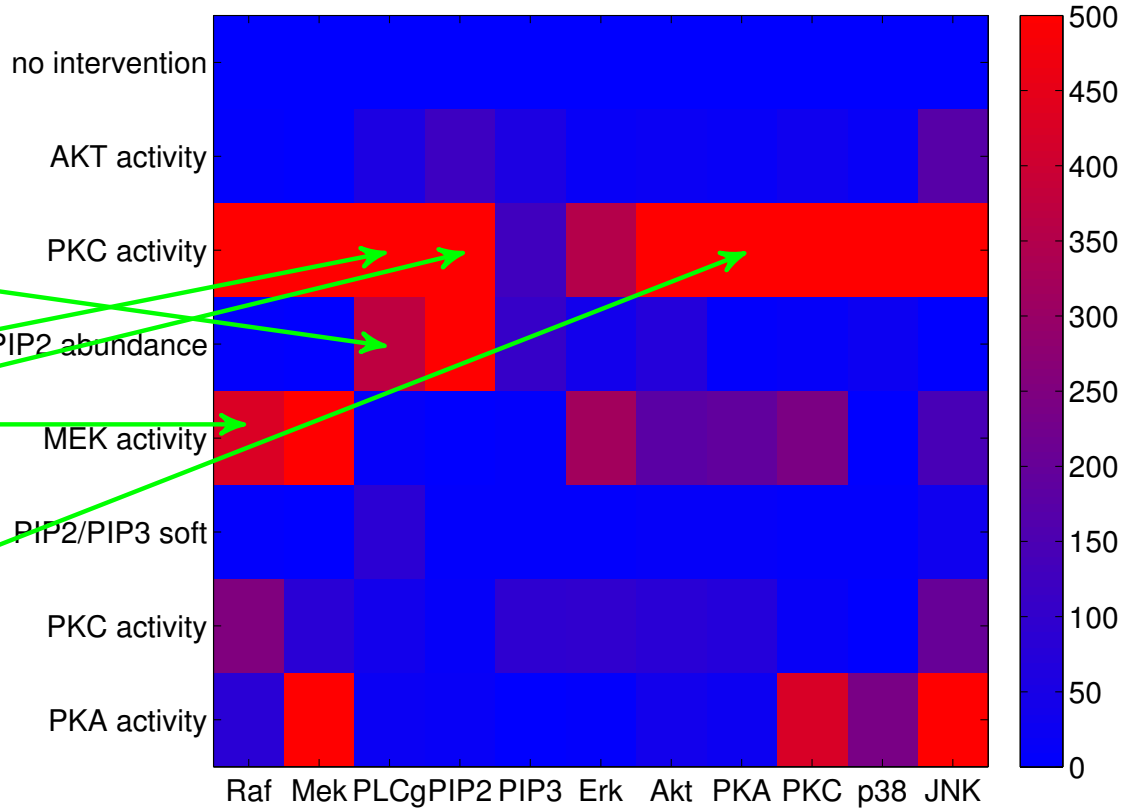
Blue: novel findings,

Red dashed: missing.

Comparison with ground truth (max. 17 edges, acyclic)



This work



KS test w.r.t. observational data

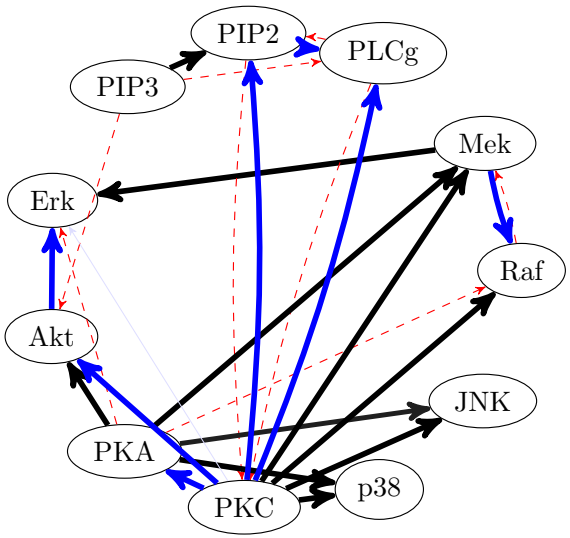
Black: expected,

Blue: novel findings,

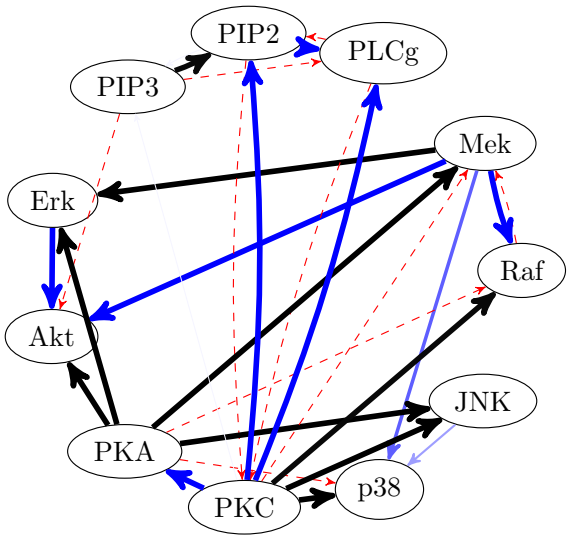
Red dashed: missing.

Results (max. 17 edges, acyclic)

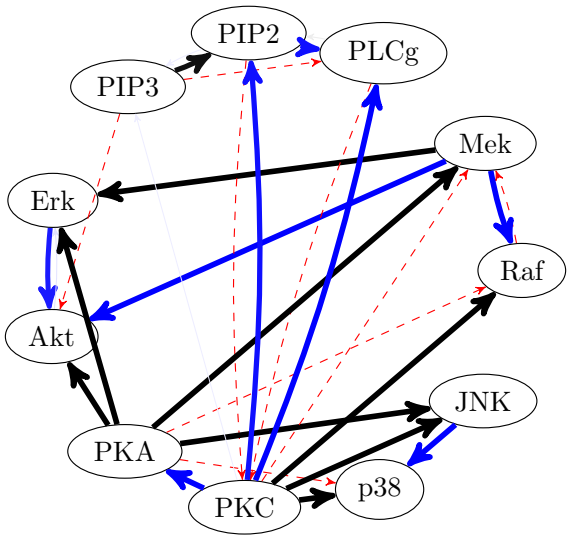
Acyclic, strongly regularized results for different priors:



linear
Gaussian



nonlinear
Gaussian



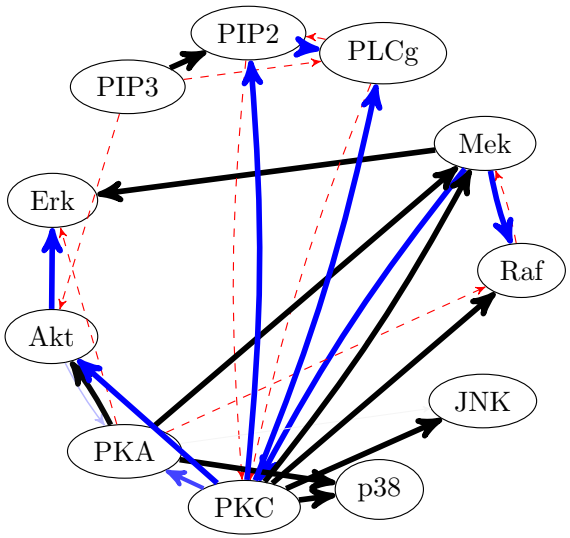
nonlinear
non-Gaussian

Black: expected, Blue: novel findings, Red dashed: missing.

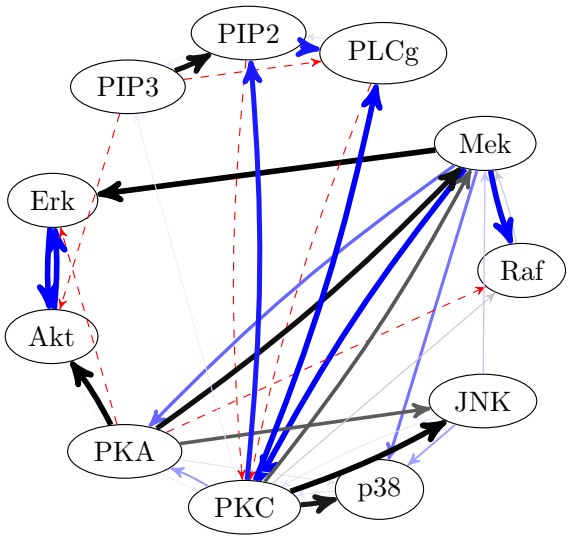
Note: no strong dependence on prior.

Results (max. 17 edges, cyclic)

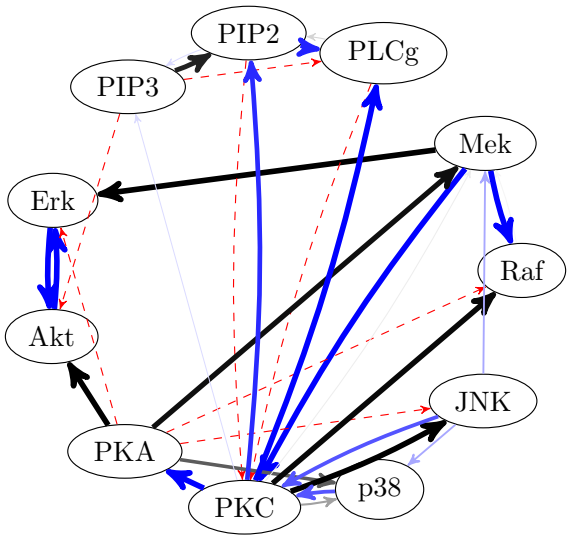
Cyclic, strongly regularized results for different priors:



linear
Gaussian



nonlinear
Gaussian



nonlinear
non-Gaussian

Black: expected, Blue: novel findings, Red dashed: missing.

Good news: Our method reveals some likely feedback cycles.
Bad news: stronger dependence on prior (more data needed?).

Discussion

Performing a proper causal analysis of this data is a **challenging task**:

- time-series data are absent, so need to assume homeostatis;
- confounders could be present;
- feedback loops are expected to be present;
- most interventions change the activity instead of the abundance;
- assumptions about the specificity of interventions may be unrealistic;
- faithfulness violations are present.

Discussion

Performing a proper causal analysis of this data is a **challenging task**:

- time-series data are absent, so need to assume homeostatis;
- confounders could be present;
- feedback loops are expected to be present;
- most interventions change the activity instead of the abundance;
- assumptions about the specificity of interventions may be unrealistic;
- faithfulness violations are present.

Main contributions:

- More principled approach to learn structure of (a)cyclic causal models from combination of observational and interventional equilibrium data.
- Natural way to model activity interventions.

Conclusions and future work

Conclusions:

- Results support the hypothesis that the underlying system contains feedback loops.
- The proposed method identifies a few likely feedback loops, but more data is probably necessary.

Conclusions and future work

Conclusions:

- Results support the hypothesis that the underlying system contains feedback loops.
- The proposed method identifies a few likely feedback loops, but more data is probably necessary.

Future work:

- Analysis of causal predictive performance: do our models give more accurate predictions, also for (new) interventions?
- Experimental evaluation of predictions.

Part V

Causal Inference: Outlook

Three interesting and important future directions

1. The field has focussed mainly on *acyclic* causal systems. Feedback occurs in many different systems in biology, economy, and other fields. A lot of interesting work remains to be done for the cyclic case.

Three interesting and important future directions

1. The field has focussed mainly on *acyclic* causal systems. Feedback occurs in many different systems in biology, economy, and other fields. A lot of interesting work remains to be done for the cyclic case.
2. The Causal Discovery literature has focussed mainly on the special case of *purely observational data*. In practice, interventional data is often available as well, and this data typically conveys important information about the underlying causal structure. Designing good methods and algorithms that can use this data may have a big impact in many empirical sciences.

Three interesting and important future directions

1. The field has focussed mainly on *acyclic* causal systems. Feedback occurs in many different systems in biology, economy, and other fields. A lot of interesting work remains to be done for the cyclic case.
2. The Causal Discovery literature has focussed mainly on the special case of *purely observational data*. In practice, interventional data is often available as well, and this data typically conveys important information about the underlying causal structure. Designing good methods and algorithms that can use this data may have a big impact in many empirical sciences.
3. Related to AI: Can we build "intelligent" systems that are able to learn a causal model of the world? An important ingredient (in addition to being able to learn from given data) is *active learning*, or *experimental design*.

Three interesting and important future directions

1. The field has focussed mainly on *acyclic* causal systems. Feedback occurs in many different systems in biology, economy, and other fields. A lot of interesting work remains to be done for the cyclic case.
2. The Causal Discovery literature has focussed mainly on the special case of *purely observational data*. In practice, interventional data is often available as well, and this data typically conveys important information about the underlying causal structure. Designing good methods and algorithms that can use this data may have a big impact in many empirical sciences.
3. Related to AI: Can we build "intelligent" systems that are able to learn a causal model of the world? An important ingredient (in addition to being able to learn from given data) is *active learning*, or *experimental design*.

Thanks for your attention!

Acknowledgments and References

I thank Bram Thijssen and Tjeerd Dijkstra for stimulating discussions. JM was supported by NWO, the Netherlands Organization for Scientific Research (VENI grant 639.031.036).


Dash, D. (2003).
Caveats for Causal Reasoning.
PhD thesis, University of Pittsburgh, Pittsburgh, PA.


Hall-Jackson, C. A., Eyers, P. A., Cohen, P., Goedert, M., Boyle, F. T., Hewitt, N., Plant, H., and Hedge, P. (1999).
Paradoxical activation of Raf by a novel Raf inhibitor.
Chemistry & Biology, 6:559–568.


Itani, S., Ohannessian, M., Sachs, K., Nolan, G. P., and Dahleh, M. A. (2010).
Structure learning in causal cyclic networks.
In *JMLR Workshop and Conference Proceedings*, volume 6, page 165176.


Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D., and Nolan, G. (2005).
Causal protein-signaling networks derived from multiparameter single-cell data.
Science, 308:523–529.


Schmidt, M. and Murphy, K. (2009).
Modeling discrete interventional data using directed cyclic graphical models.
In *Proceedings of the 25th Annual Conference on Uncertainty in Artificial Intelligence (UAI-09)*.