Computational Hypergraph Discovery

Théo Bourdais

January 18, 2024

California Institute of Technology

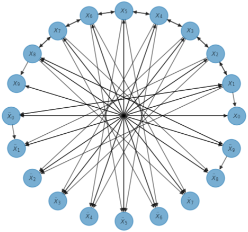
x_1	$\ddot{x}_1$	...	$\ddot{x}_{10}$
0.45	0.66	...	-0.23
⋮	⋮	⋱	⋮
-0.78	-0.12	...	0.89

$\xrightarrow{\text{CHD}}$

$$\ddot{x}_1 = \frac{c^2}{h^2}(x_2 + x_0 - 2x_1)(1 + (x_2 - x_0)^3)$$

$$\vdots$$

$$\ddot{x}_{10} = \frac{c^2}{h^2}(x_9 - 2x_{10})(1 - x_9^3)$$



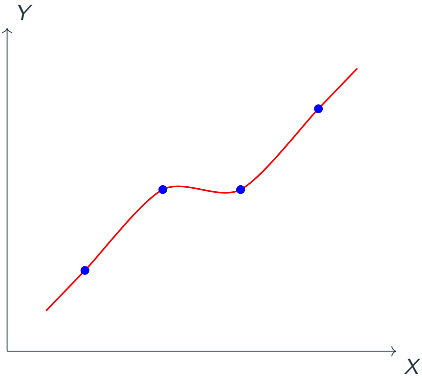
Computational Hypergraph Discovery, a Gaussian Process Framework for Connecting the Dots.

Théo Bourdais, Pau Batlle, Xianjin Yang, Ricardo Baptista, Nicolas Rouquette, and Houman Owhadi

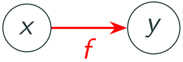
ArXiv, (2023). /abs/2311.17007

The regression problem

Suppose $y = f(x)$, given samples (X_i, Y_i) for $i = 1, .., N$, approximate f



Graph representation



Caltech

For $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$, $i = 1, .., N$, approximate f s.t. $y = f(x)$.

Linear Ridge regression

Our approximation is a linear function $\tilde{f}(x) = {\beta^*}^T x$ with

$$\beta^* = \underset{\beta \in \mathbb{R}^p}{\arg\min} \sum_{i=1}^{N} |Y_i - \beta^T X_i|^2 + \gamma \|\beta\|^2$$

We know that, for $k(x, y) = x^T y$,

$$\tilde{f} \in \mathcal{H}_k = \overline{\{\sum_i \alpha_i k(\cdot, z_i), \text{for some } z_i, \alpha_i\}}$$

Caltech

For $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$, $i = 1, .., N$, approximate f s.t. $y = f(x)$.

Quadratic Ridge regression

Our approximation is a quadratic function $\tilde{f}(x) = {\beta^*}^T \psi(x)$, $\psi(x) = (1, x, x^2)$,

$$\beta^* = \underset{\beta \in \mathbb{R}^{2p+1}}{\arg\min} \sum_{i=1}^{N} |Y_i - \beta^T \psi(X_i)|^2 + \gamma \|\beta\|^2$$

We know that, for $k(x, y) = \psi(x)^T \psi(y)$,

$$\tilde{f} \in \mathcal{H}_k = \overline{\{\sum_i \alpha_i k(\cdot, z_i), \text{for some } z_i, \alpha_i\}}$$

For $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$, $i = 1, .., N$, approximate f s.t. $y = f(x)$.

Kernel Ridge regression

Our approximation is a function in a space $\mathcal{H}_k$[1] defined by the kernel k.

$$\tilde{f} = \arg\min_{f \in \mathcal{H}_k} \sum_{i=1}^{N} |Y_i - f(X_i)|^2 + \gamma \|f\|^2$$

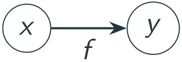
We know that,

$$\tilde{f} \in \mathcal{H}_k = \overline{\{\sum_i \alpha_i k(\cdot, z_i), \text{for some } z_i, \alpha_i\}}$$

[1] $\mathcal{H}_k$ is called the Reproducing Kernel Hilbert Space (RKHS) of k

Computational Hypergraphs

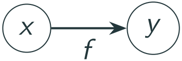
A computational hypergraph is a graphical representation of a set of equations



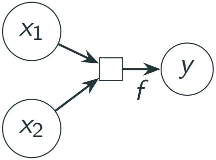
$$y = f(x)$$

Computational Hypergraphs

A computational hypergraph is a graphical representation of a set of equations



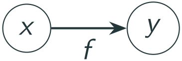
$$y = f(x)$$



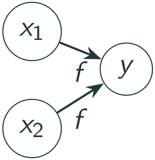
$$y = f(x_1, x_2)$$

Computational Hypergraphs

A computational hypergraph is a graphical representation of a set of equations



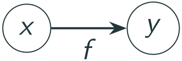
$$y = f(x)$$



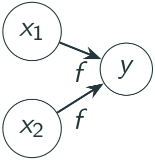
$$y = f(x_1, x_2)$$

Computational Hypergraphs

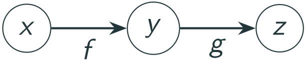
A computational hypergraph is a graphical representation of a set of equations



$$y = f(x)$$

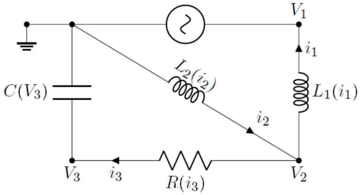


$$y = f(x_1, x_2)$$



$$y = f(x)$$
$$z = g(y)$$

The electrical circuit example[2]



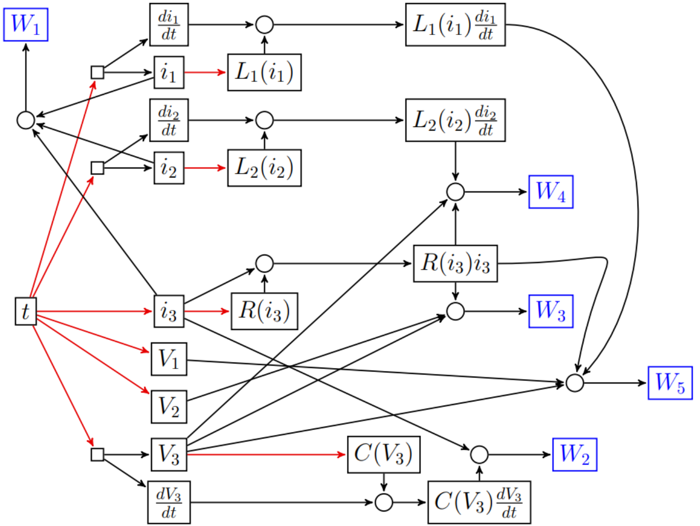
$$
\begin{aligned}
i_1 + i_3 &= i_2 \\
i_3 &= C(V_3)\frac{dV_3}{dt} \\
V_2 - V_3 &= R(i_3)i_3 \\
-V_2 &= L_2(i_2)\frac{di_2}{dt} \\
V_2 - V_1 &= L_1(i_1)\frac{di_1}{dt}
\end{aligned}
$$

[2] Owhadi, *Computational Graph Completion*.

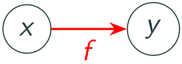
The electrical circuit example

Caltech

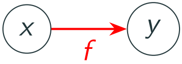
Since any set of equations can be represented as a Computational Hypergraph, we can obtain:



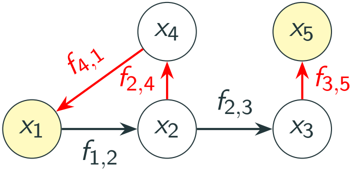
Regression Given samples $Y_i = f(X_i)$ for $i = 1, .., N$, approximate f



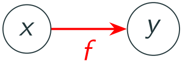
Regression Given samples $Y_i = f(X_i)$ for $i = 1, .., N$, approximate f



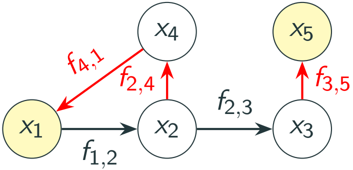
Hypergraph Completion Given the graph's structure and samples of its variables, approximate unknown edges, and missing data.



Regression Given samples $Y_i = f(X_i)$ for $i = 1, .., N$, approximate f



Hypergraph Completion Given the graph's structure and samples of its variables, approximate unknown edges, and missing data.



Hypergraph discovery Given samples of the variables, find the structure of the graph.

- Brain networks

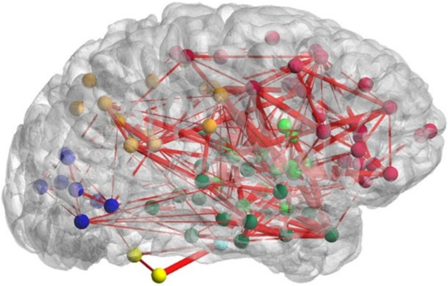


Figure 1: Image from Shu-Hsien Chu et al.

Objective: Discover functional dependencies between the activities of different brain regions.

- Brain networks
- Economic networks

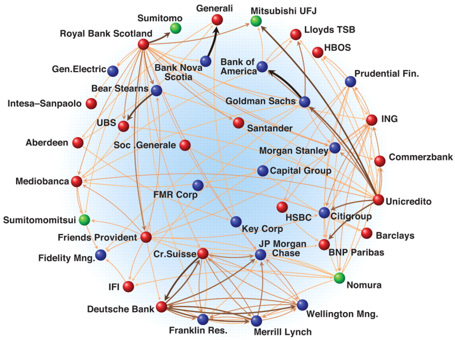


Figure 1: Image from Schweitzer et al.

Objective: Discover functional dependencies between economic markers of different banks

Caltech

- Brain networks
- Economic networks
- Weather modelling

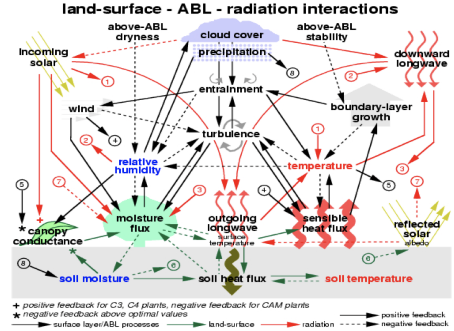


Figure 1: Image from Michael Ek.

Objective: Discover functional dependencies between the different variables

Causal inference and Probabilistic graphs

- Usually tackle a different problem (e.g., conditional independence or causality)
- Relies on strong assumptions (e.g., access to a distribution)

Causal inference and Probabilistic graphs

- Usually tackle a different problem (e.g., conditional independence or causality)
- Relies on strong assumptions (e.g., access to a distribution)

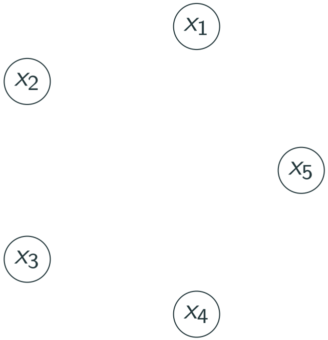
Sparse regressions

- Uses knowledge of sparse representations in a dictionary of functions
- Example: SINDY

Caltech

The CHD problem

Given N samples of our variables, recover the functional dependencies between them (i.e., the structure of the graph).



The CHD problem

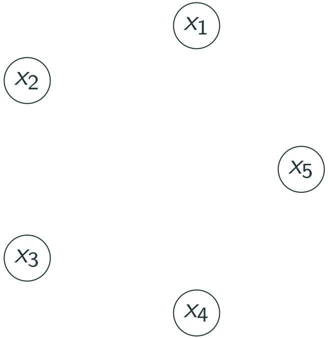
Given N samples of x_1, x_2, x_3, x_4, x_5, recover the functional dependencies (i.e. the structure of the graph).

Ancestors: If $x_5 = f(x_1, x_4)$ for some f, x_1, x_4 are ancestors of x_5.

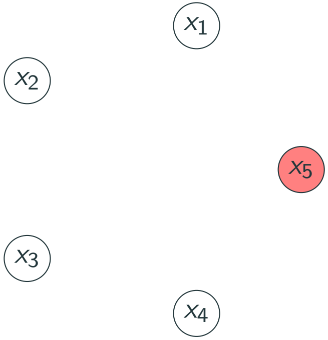
Our solution

Ancestors: If $x_5 = f(x_1, x_4)$ for some f, x_1, x_4 are ancestors of x_5.

For each node, identify its ancestors



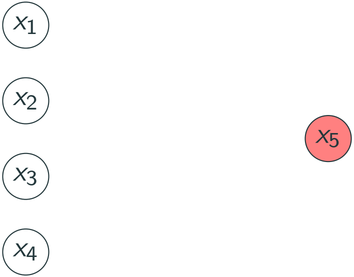
Ancestors: If $x_5 = f(x_1, x_4)$ for some f, x_1, x_4 are ancestors of x_5.
For each node, identify its ancestors



Our solution

Ancestors: If $x_5 = f(x_1, x_4)$ for some f, x_1, x_4 are ancestors of x_5.

For each node, identify its ancestors

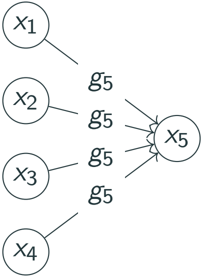


Our solution

x_5 is not a function of

$x_1, \ldots, x_4$

or



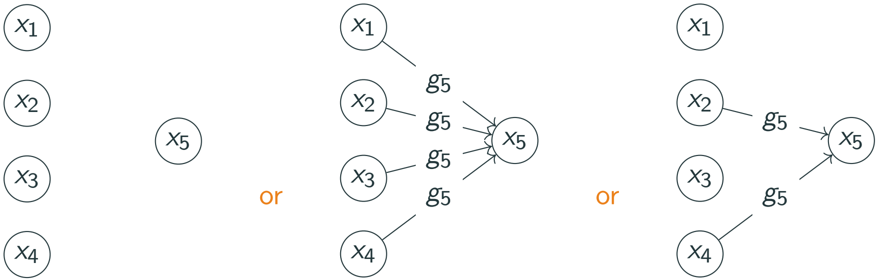
There is a function g_5 s.t.

$$x_5 = g_5(x_1, \ldots, x_4)$$

There are three questions:

- Does x_5 have any ancestors?

Our solution



x_5 is not a function of

$x_1, \ldots, x_4$

There is a function g_5 s.t.

$x_5 = g_5(x_1, \ldots, x_4)$

There is a function g_5 s.t.

$x_5 = g_5(x_2, x_4)$

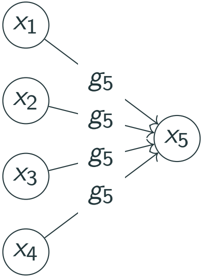
There are three questions:

- Does x_5 have any ancestors?
- If so, what is the minimum set of ancestors?

x_5 is not a function of

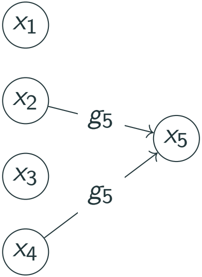
$x_1, \ldots, x_4$

or



There is a function g_5 s.t.

$x_5 = g_5(x_1, \ldots, x_4)$



There is a function g_5 s.t.

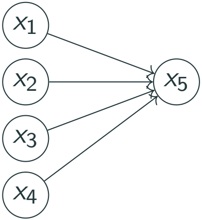
$x_5 = g_5(x_2, x_4)$

There are three questions:

- Does x_5 have any ancestors?
- If so, what is the minimum set of ancestors?
- What kind of function is g_5?

Does x_5 have any ancestors?

Ancestors,
samples gathered in X

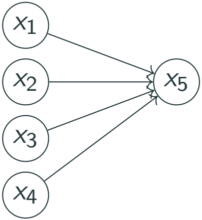


Target
samples gathered in Y

Does x_5 have any ancestors?

Caltech

Ancestors,
samples gathered in X



Target
samples gathered in Y

Let's see if there is g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a Gaussian Process (kernel k and noise variance γ):

$$g_5 = \arg\min_f \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2 \tag{1}$$

Does x_5 have any ancestors?

Caltech

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \; \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

Does x_5 have any ancestors?

Caltech

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \; \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

$$s \;=\; \|g_5\|_k^2$$

(variance of data explained by model)

Does x_5 have any ancestors?

Caltech

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \ \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

$$s = \|g_5\|_k^2$$

(variance of data explained by model)

$$n = \frac{1}{\gamma}|g_5(X) - Y|^2$$

(variance of data explained by noise)

Does x_5 have any ancestors?

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \; \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

$$s \;=\; \|g_5\|_k^2$$

(variance of data explained by model)

$$n \;=\; \frac{1}{\gamma}|g_5(X) - Y|^2$$

(variance of data explained by noise)

The noise n comes from two sources:

- True noise from the data

Does x_5 have any ancestors?

Caltech

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \; \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

$$s = \|g_5\|_k^2$$

(variance of data explained by model)

$$n = \frac{1}{\gamma}|g_5(X) - Y|^2$$

(variance of data explained by noise)

The noise n comes from two sources:

- True noise from the data
- Unexplained data variance

Does x_5 have any ancestors?

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \; \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

$$s = \|g_5\|_k^2$$

(variance of data explained by model)

$$n = \frac{1}{\gamma}|g_5(X) - Y|^2$$

(variance of data explained by noise)

The noise n comes from two sources:

- True noise from the data
- Unexplained data variance
    - Quantifies model misspecification

Does x_5 have any ancestors?

Caltech

Noise-to-signal ratio

$\frac{n}{n+s} \in [0,1]$, quantifies how much the data agrees with x_5 having $x_1, .., x_4$ as ancestors

- if $\frac{n}{n+s} \approx 0$: The model is well specified.
- if $\frac{n}{n+s} \approx 1$: The model is misspecified.

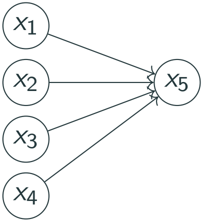
Does x_5 have any ancestors?

Noise-to-signal ratio

$\frac{n}{n+s} \in [0, 1]$, quantifies how much the data agrees with x_5 having $x_1, .., x_4$ as ancestors

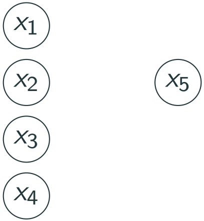
- if $\frac{n}{n+s} \approx 0$: The model is well specified.
- if $\frac{n}{n+s} \approx 1$: The model is misspecified.

If $\frac{n}{n+s} < 0.5$, x_5 has ancestors



or

If $\frac{n}{n+s} > 0.5$, x_5 has no ancestors

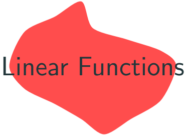


Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions we are searching g_5 in.

Choosing the kernel

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions we are searching g_5 in.

g_5 linear

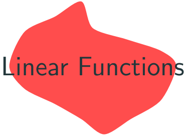


Current kernel: Linear

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i$$

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions we are searching g_5 in.

g_5 linear $\xrightarrow{\frac{n}{n+s} < 0.5}$ x_5 has ancestors and g_5 is linear

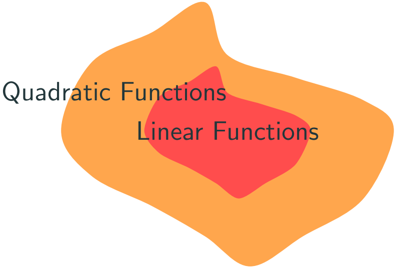


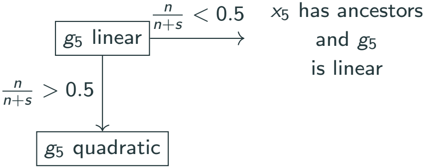
Current kernel: Linear

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i$$

Choosing the kernel

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions we are searching g_5 in.



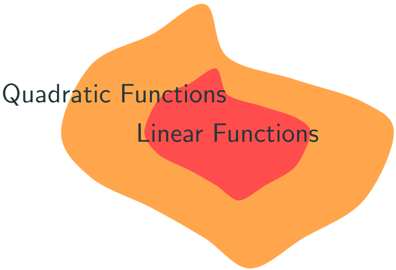


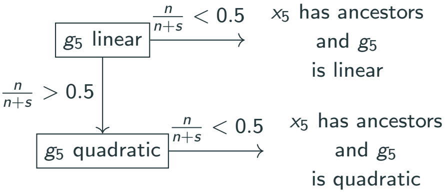
Current kernel: Quadratic

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

Choosing the kernel

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions we are searching g_5 in.



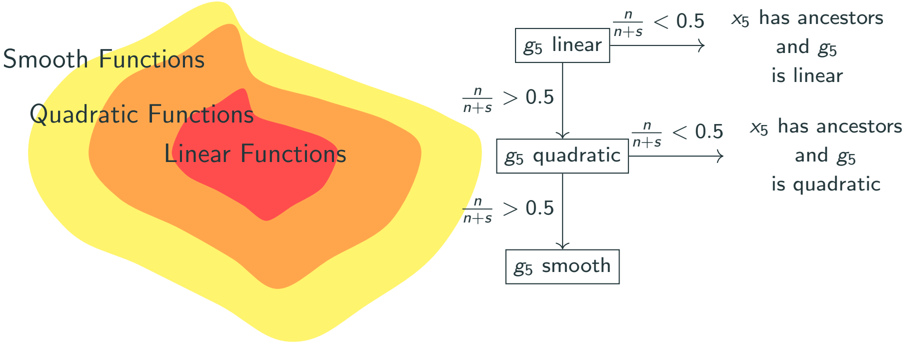


Current kernel: Quadratic

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

Choosing the kernel

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions we are searching g_5 in.

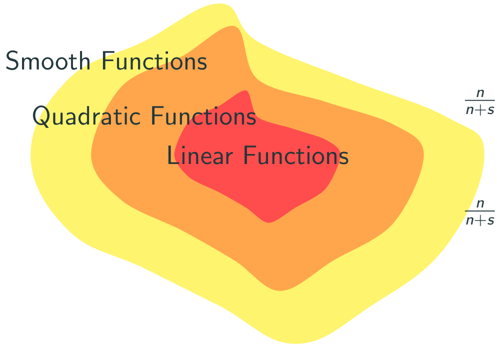


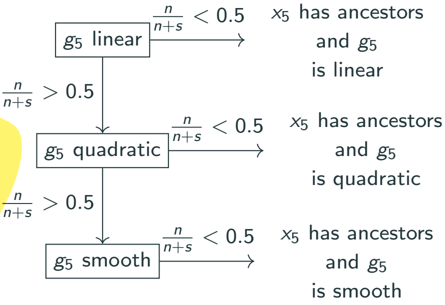
Current kernel: Nonlinear

$$k(x,y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j + \prod_{i=1}^{n} (1 + e^{-(x_i - y_i)^2})$$

Choosing the kernel

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions we are searching g_5 in.



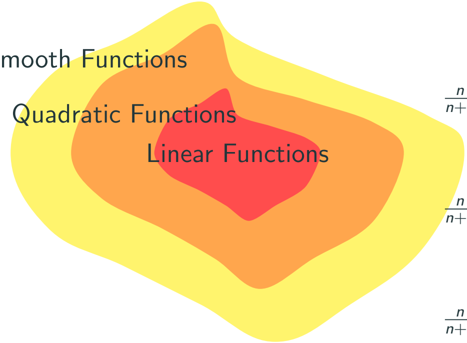


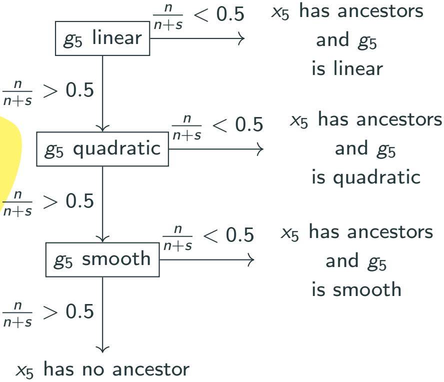
Current kernel: **Nonlinear**

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j + \prod_{i=1}^{n} (1 + e^{-(x_i - y_i)^2})$$

Choosing the kernel

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions we are searching g_5 in.





Current kernel: **Nonlinear**

$$k(x,y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j + \prod_{i=1}^{n} (1 + e^{-(x_i - y_i)^2})$$

To find the minimum set of ancestors, we will:

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors

what is the minimum set of ancestors?

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:
  - Identify the least important potential ancestor

what is the minimum set of ancestors?

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:
    - Identify the least important potential ancestor
    - Remove it from potential ancestors

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:
  - Identify the least important potential ancestor
  - Remove it from potential ancestors
  - Check that the data agrees

Caltech

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:
  - Identify the least important potential ancestor
  - Remove it from potential ancestors
  - Check that the data agrees $\Leftarrow$ we use $\frac{n}{n+s}$

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:
    - Identify the least important potential ancestor
    - Remove it from potential ancestors
    - Check that the data agrees $\Leftarrow$ we use $\frac{n}{n+s}$

Identify the least important ancestor

if we found g_5 s.t.$x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

Identify the least important ancestor

Caltech

if we found g_5 s.t.$x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

Suppose we are using the quadratic kernel:

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

if we found g_5 s.t.$x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

Suppose we are using the quadratic kernel:

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

Observe $k = k_2 + k_{-2}$

if we found g_5 s.t.$x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

Suppose we are using the quadratic kernel:

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

Observe $k = k_2 + k_{-2}$

- k_2 depends on x_2: $k_2(x, y) = x_2 y_2 + x_2^2 y_2^2 + 2 \sum_{j \neq 2} x_2 x_j y_2 y_j$

if we found g_5 s.t.$x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

Suppose we are using the quadratic kernel:

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

Observe $k = k_2 + k_{-2}$

- k_2 depends on x_2: $k_2(x, y) = x_2 y_2 + x_2^2 y_2^2 + 2 \sum_{j \neq 2} x_2 x_j y_2 y_j$
- k_{-2} does not depend on x_2: $k_{-2} = k - k_2$

Identify the least important ancestor

if we found g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

$$k = k_2 + k_{-2}$$

Then we there is

Identify the least important ancestor

if we found g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

$$k = k_2 + k_{-2}$$

Then we there is

- $f_2 \in \mathcal{H}_{k_2}$, depends on x_2

Identify the least important ancestor

if we found g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

$$k = k_2 + k_{-2}$$

Then we there is

- $f_2 \in \mathcal{H}_{k_2}$, depends on x_2
- $f_{-2} \in \mathcal{H}_{k_{-2}}$, does not depend on x_2

Identify the least important ancestor

Caltech

if we found g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

$$k = k_2 + k_{-2}$$

Then we there is

- $f_2 \in \mathcal{H}_{k_2}$, depends on x_2
- $f_{-2} \in \mathcal{H}_{k_{-2}}$, does not depend on x_2

such that

$$g_5 = f_2 + f_{-2}$$
$$s = \|g_5\|_k^2 = \|f_2\|_{k_2}^2 + \|f_{-2}\|_{k_{-2}}^2$$

Identify the least important ancestor

Caltech

if we found g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

$$k = k_2 + k_{-2}$$

Then we there is

- $f_2 \in \mathcal{H}_{k_2}$, depends on x_2
- $f_{-2} \in \mathcal{H}_{k_{-2}}$, does not depend on x_2

such that

$$g_5 = f_2 + f_{-2}$$
$$s = \|g_5\|_k^2 = \|f_2\|_{k_2}^2 + \|f_{-2}\|_{k_{-2}}^2$$

We can define the activation a_2, which quantifies the contribution of x_2 to the signal data variance:

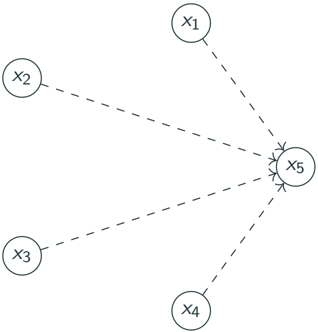
$$a_2 = \frac{\|f_2\|_{k_2}^2}{\|g_5\|_k^2}$$

The Algorithm

Caltech

Alg. for discovering the ancestors of x_5

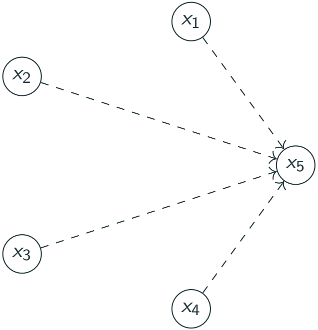
Assign all other nodes as ancestors
Compute $\frac{n}{n+s}$ for each kernel
if No kernel has low noise **then**
    x_5 has no ancestors
else
    Pick first kernel with low noise
end if
while there are some ancestors left **do**
    compute the contribution of each node
    remove the node that contributes the least
    recompute $\frac{n}{n+s}$
end while
using the evolution of $\frac{n}{n+s}$, choose the number of ancestors



The Algorithm

Alg. for discovering the ancestors of x_5

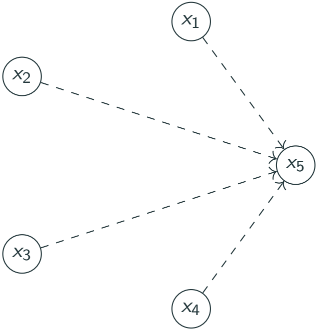
```
Assign all other nodes as ancestors
Compute n/(n+s) for each kernel
if No kernel has low noise then
    x_5 has no ancestors
else
    Pick first kernel with low noise
end if
while there are some ancestors left do
    compute the contribution of each node
    remove the node that contributes the least
    recompute n/(n+s)
end while
using the evolution of n/(n+s), choose the number of ancestors
```



- **Linear kernel**: $\frac{n}{n+s} = 0.81$
- **Quadratic kernel**: $\frac{n}{n+s} = 0.12$
- **Nonlinear kernel**: $\frac{n}{n+s} = 0.44$

Alg. for discovering the ancestors of x_5

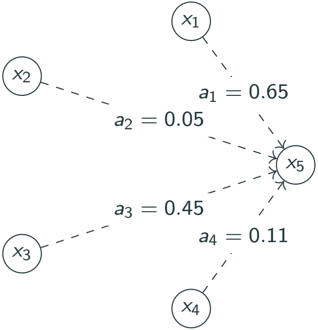
```
Assign all other nodes as ancestors
Compute n/(n+s) for each kernel
if No kernel has low noise then
    x_5 has no ancestors
else
    Pick first kernel with low noise
end if
while there are some ancestors left do
    compute the contribution of each node
    remove the node that contributes the least
    recompute n/(n+s)
end while
using the evolution of n/(n+s), choose the number of ancestors
```

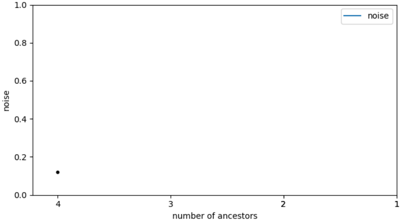


- **Linear kernel**: $\frac{n}{n+s} = 0.81$
- **Quadratic kernel**: $\frac{n}{n+s} = 0.12$
- **Nonlinear kernel**: $\frac{n}{n+s} = 0.44$

Alg. for discovering the ancestors of x_5

```
Assign all other nodes as ancestors
Compute n/(n+s) for each kernel
if No kernel has low noise then
    x5 has no ancestors
else
    Pick first kernel with low noise
end if
while there are some ancestors left do
    compute the contribution of each node
    remove the node that contributes the least
    recompute n/(n+s)
end while
using the evolution of n/(n+s), choose the number of ancestors
```

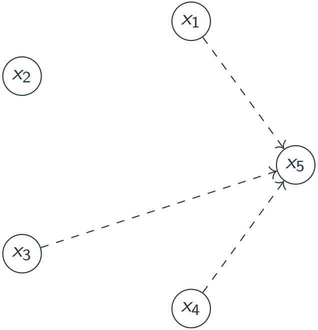


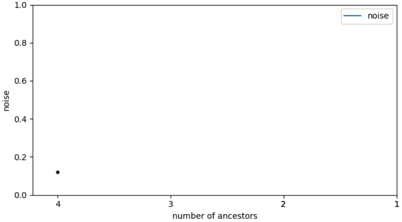


Caltech

Alg. for discovering the ancestors of x_5

1: Assign all other nodes as ancestors
2: Compute $\frac{n}{n+s}$ for each kernel
3: **if** No kernel has low noise **then**
4: x_5 has no ancestors
5: **else**
6: Pick first kernel with low noise
7: **end if**
8: **while** there are some ancestors left **do**
9: compute the contribution of each node
10: remove the node that contributes the least
11: recompute $\frac{n}{n+s}$
12: **end while**
13: using the evolution of $\frac{n}{n+s}$, choose the number of ancestors

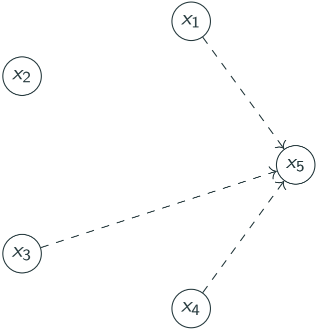


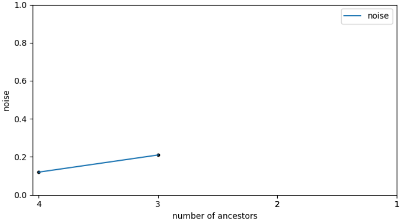


Caltech

Alg. for discovering the ancestors of x_5

1: Assign all other nodes as ancestors
2: Compute $\frac{n}{n+s}$ for each kernel
3: **if** No kernel has low noise **then**
4: x_5 has no ancestors
5: **else**
6: Pick first kernel with low noise
7: **end if**
8: **while** there are some ancestors left **do**
9: compute the contribution of each node
10: remove the node that contributes the least
11: recompute $\frac{n}{n+s}$
12: **end while**
13: using the evolution of $\frac{n}{n+s}$, choose the number of ancestors

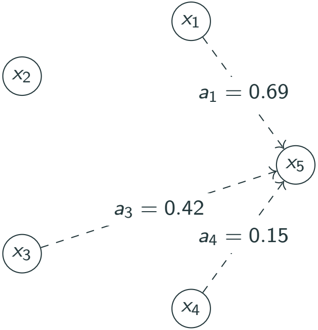


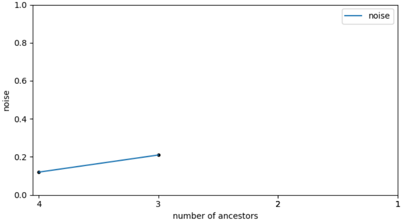


The Algorithm

Alg. for discovering the ancestors of x_5

Assign all other nodes as ancestors
Compute $\frac{n}{n+s}$ for each kernel
if No kernel has low noise **then**
    x_5 has no ancestors
else
    Pick first kernel with low noise
end if
while there are some ancestors left **do**
    compute the contribution of each node
    remove the node that contributes the least
    recompute $\frac{n}{n+s}$
end while
using the evolution of $\frac{n}{n+s}$, choose the number of ancestors

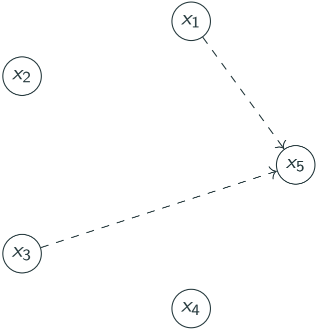


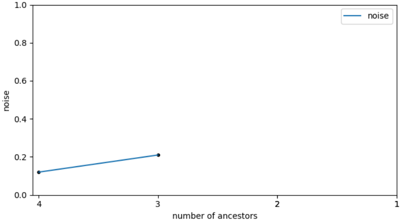


Caltech

Alg. for discovering the ancestors of x_5

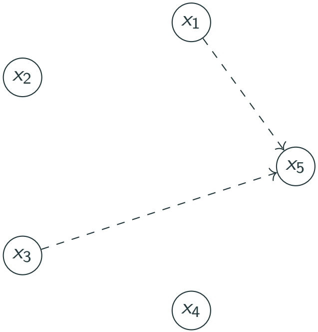
Assign all other nodes as ancestors
Compute $\frac{n}{n+s}$ for each kernel
if No kernel has low noise **then**
    x_5 has no ancestors
else
    Pick first kernel with low noise
end if
while there are some ancestors left **do**
    compute the contribution of each node
    remove the node that contributes the least
    recompute $\frac{n}{n+s}$
end while
using the evolution of $\frac{n}{n+s}$, choose the number of ancestors

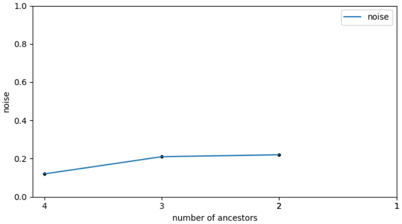




Alg. for discovering the ancestors of x_5

```
Assign all other nodes as ancestors
Compute n/(n+s) for each kernel
if No kernel has low noise then
    x5 has no ancestors
else
    Pick first kernel with low noise
end if
while there are some ancestors left do
    compute the contribution of each node
    remove the node that contributes the least
    recompute n/(n+s)
end while
using the evolution of n/(n+s), choose the number of ancestors
```

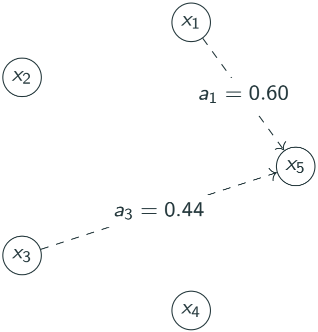


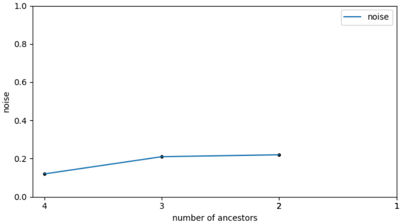


The Algorithm

Alg. for discovering the ancestors of x_5

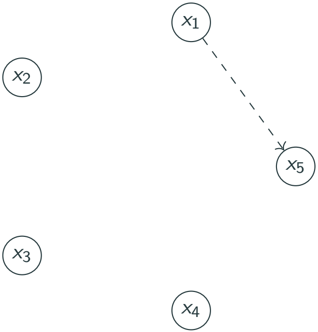
Assign all other nodes as ancestors
Compute $\frac{n}{n+s}$ for each kernel
if No kernel has low noise **then**
    x_5 has no ancestors
else
    Pick first kernel with low noise
end if
while there are some ancestors left **do**
    compute the contribution of each node
    remove the node that contributes the least
    recompute $\frac{n}{n+s}$
end while
using the evolution of $\frac{n}{n+s}$, choose the number of ancestors

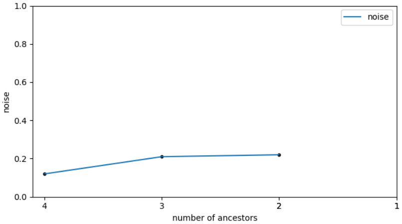




Alg. for discovering the ancestors of x_5

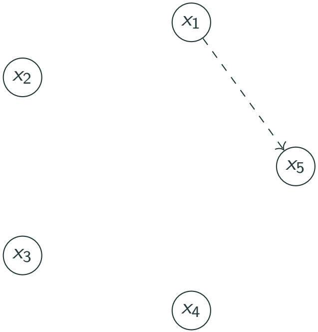
```
Assign all other nodes as ancestors
Compute n/(n+s) for each kernel
if No kernel has low noise then
    x5 has no ancestors
else
    Pick first kernel with low noise
end if
while there are some ancestors left do
    compute the contribution of each node
    remove the node that contributes the least
    recompute n/(n+s)
end while
using the evolution of n/(n+s), choose the number of ancestors
```

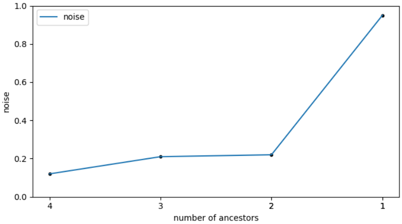




Alg. for discovering the ancestors of x_5

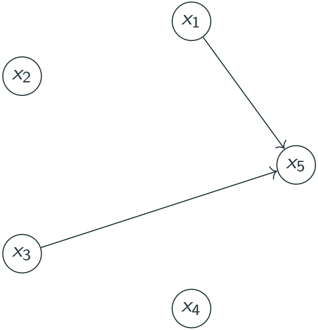
Assign all other nodes as ancestors
Compute $\frac{n}{n+s}$ for each kernel
if No kernel has low noise **then**
    x_5 has no ancestors
else
    Pick first kernel with low noise
end if
while there are some ancestors left **do**
    compute the contribution of each node
    remove the node that contributes the least
    recompute $\frac{n}{n+s}$
end while
using the evolution of $\frac{n}{n+s}$, choose the number of ancestors

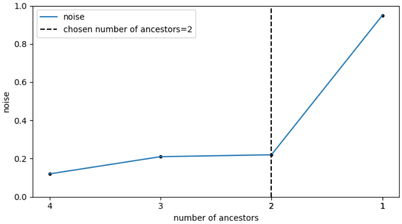




Alg. for discovering the ancestors of x_5

```
Assign all other nodes as ancestors
Compute n/(n+s) for each kernel
if No kernel has low noise then
    x5 has no ancestors
else
    Pick first kernel with low noise
end if
while there are some ancestors left do
    compute the contribution of each node
    remove the node that contributes the least
    recompute n/(n+s)
end while
using the evolution of n/(n+s), choose the number of ancestors
```





Caltech

Let $N = 10$ masses, for $i = 0, .., N-1$, their displacement from equilibrium x_i. We have:

$$\ddot{x}_i = \frac{c^2}{h^2}(x_{i+1} + x_{i-1} - 2x_i)(1 + (x_{i+1} - x_{i-1})^2) \qquad (2)$$

Boundary condition: $x_{-1} = x_N = 0$

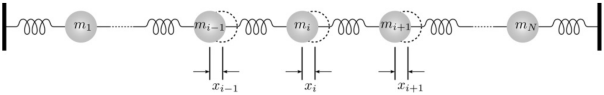


Figure 2: Nelson et al., 2018

The Fermi-Pasta-Ulam-Tsingou problem

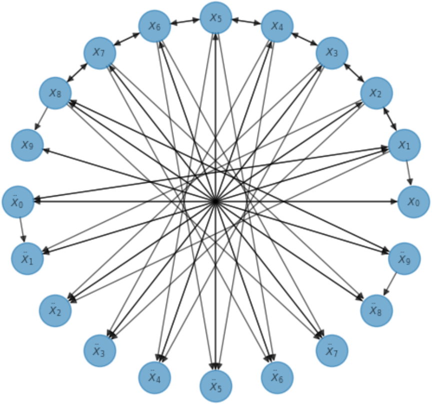
Caltech

$$\ddot{x}_i = \frac{c^2}{h^2}(x_{i+1} + x_{i-1} - 2x_i)(1 + (x_{i+1} - x_{i-1})^2) \tag{3}$$

We observe $n = 1000$ snapshots of $x_i, \dot{x}_i, \ddot{x}_i, i = 0, .., 9$.

$$\ddot{x}_i = \frac{c^2}{h^2}(x_{i+1} + x_{i-1} - 2x_i)(1 + (x_{i+1} - x_{i-1})^2) \qquad (3)$$

We observe $n = 1000$ snapshots of $x_i, \dot{x}_i, \ddot{x}_i, i = 0, .., 9$. We recover the graph perfectly, even with uninformative prior:



A typical evolution of the noise (for $\ddot{x}_7$):

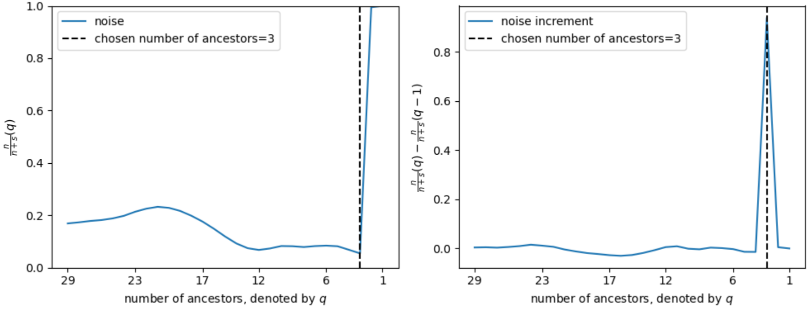


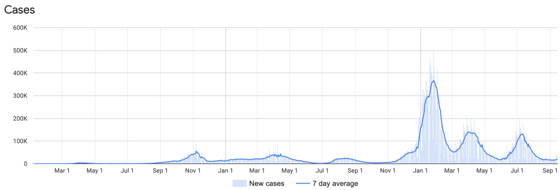
Figure 3: Left: evolution of noise-to-signal ratio . **Right**: Increment in noise ($\frac{n}{n+s}(q) - \frac{n}{n+s}(q-1)$ for q the number of ancestors)

Caltech

The dataset: Google's COVID data on France

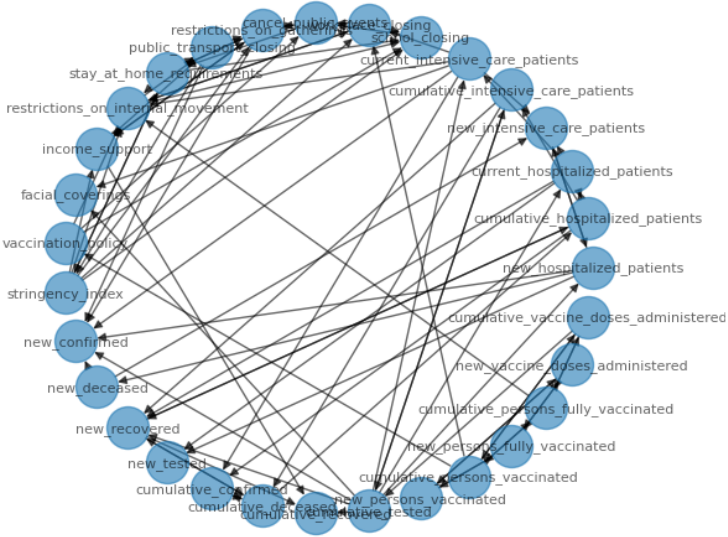
Daily values of 31 variables during 500 days:

- Epidemiology dataset (new infections, cumulative deaths,...)
- Hospital dataset (number of admitted patients, patients in intensive care, etc.)
- Vaccine dataset (number of vaccinated individuals,...)
- Policy dataset (indicators related to government responses: school closures, lockdown measures, etc.)



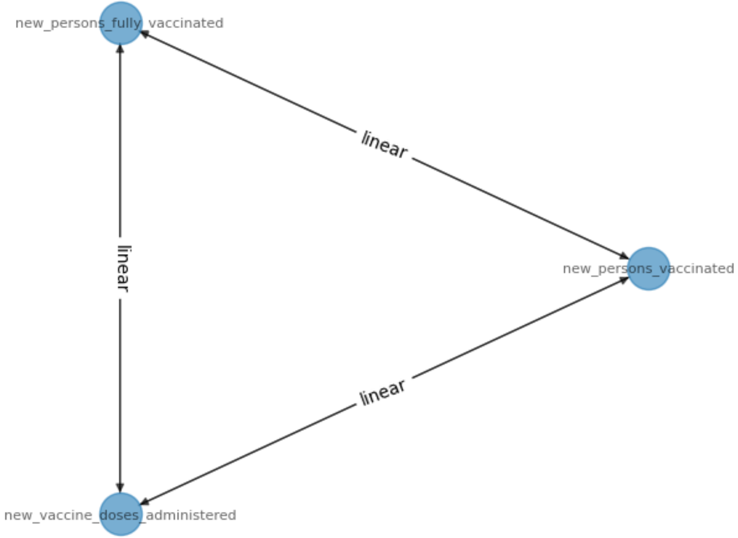
COVID dataset

Caltech

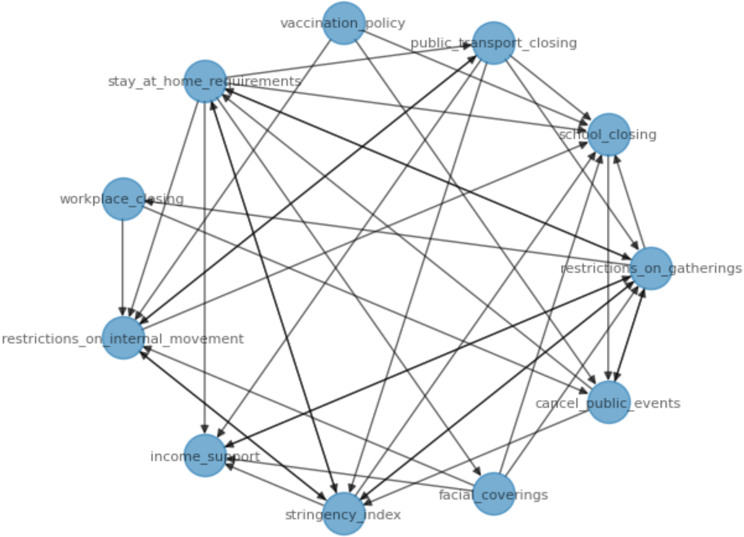


COVID dataset

Caltech

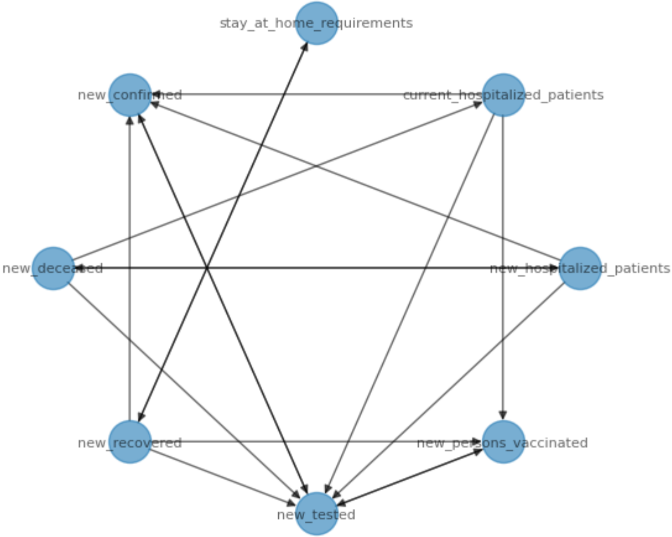


Caltech


vaccination_policy
public_transport_closing
stay_at_home_requirements
school_closing
workplace_closing
restrictions_on_gatherings
restrictions_on_internal_movement
cancel_public_events
income_support
facial_coverings
stringency_index

COVID dataset

Caltech



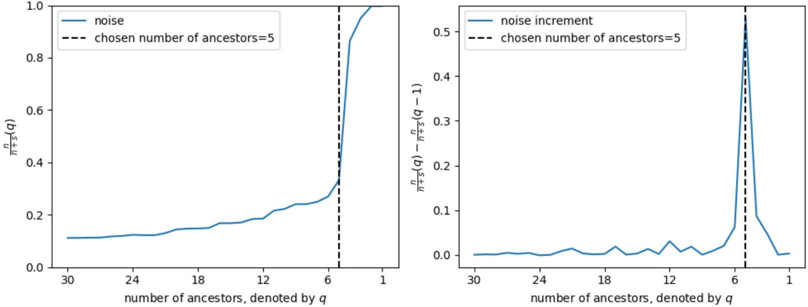


Figure 4: Evolution of the noise-to-signal ratio when pruning ancestors for the cumulative number of hospitalized patients.

The Sachs dataset

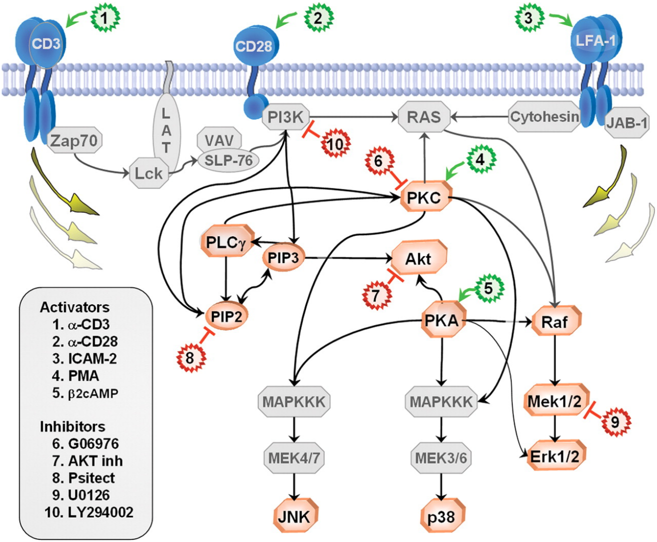
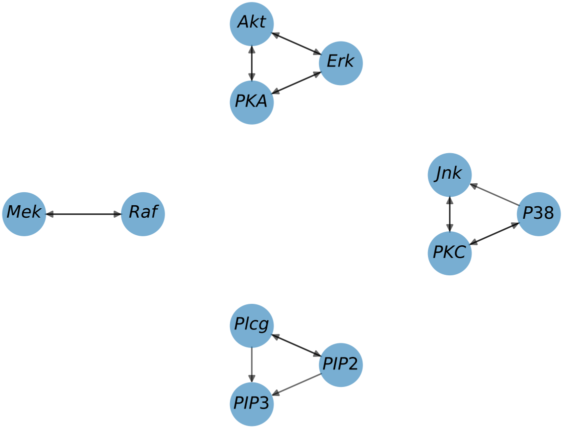


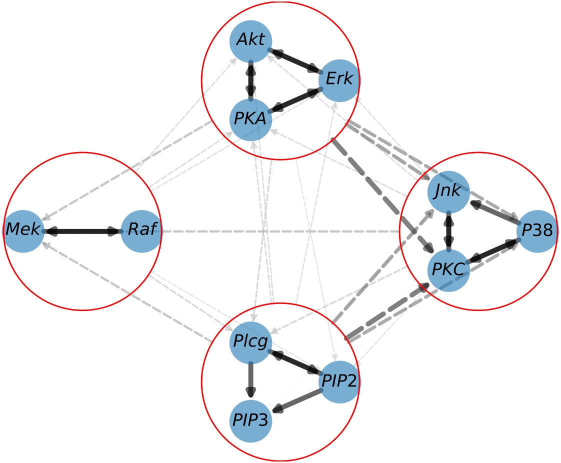
Figure 5: Sachs et al, 2005

In this dataset, some variables are strongly dependent, while other dependencies are weaker. To tackle this disparity, we cluster the nodes:



Caltech

In this dataset, some variables are strongly dependent, while other dependencies are weaker. To tackle this disparity, we cluster the nodes:



The Sachs dataset

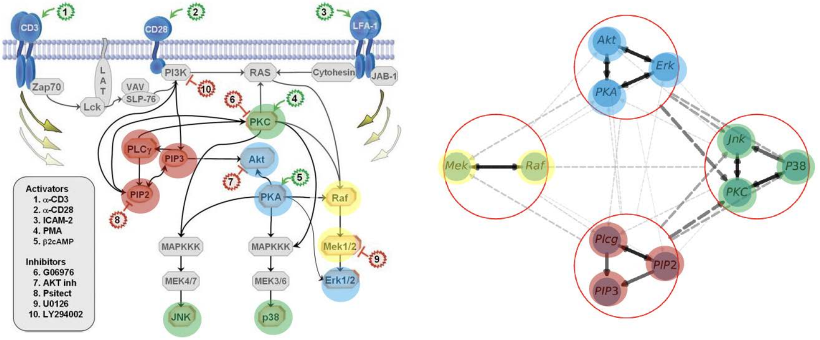


Figure 8: Comparison of recovered graph and protein signaling network

Contributions

We developed a Gaussian Process-based framework to recover functional dependencies between variables

- Works for any unlabelled dataset, with few assumptions
- interpretable
- Recovers known equations in toy examples
- Yields plausible results for real datasets

Computational Hypergraph Discovery, a Gaussian Process Framework for Connecting the Dots

Théo Bourdais, Pau Batlle, Xianjin Yang, Ricardo Baptista, Nicolas Rouquette, and Houman Owhadi

ArXiv, (2023). /abs/2311.17007

ComputationalHypergraphDiscovery

`pip install ComputationalHypergraphDiscovery`

Blog post on my website