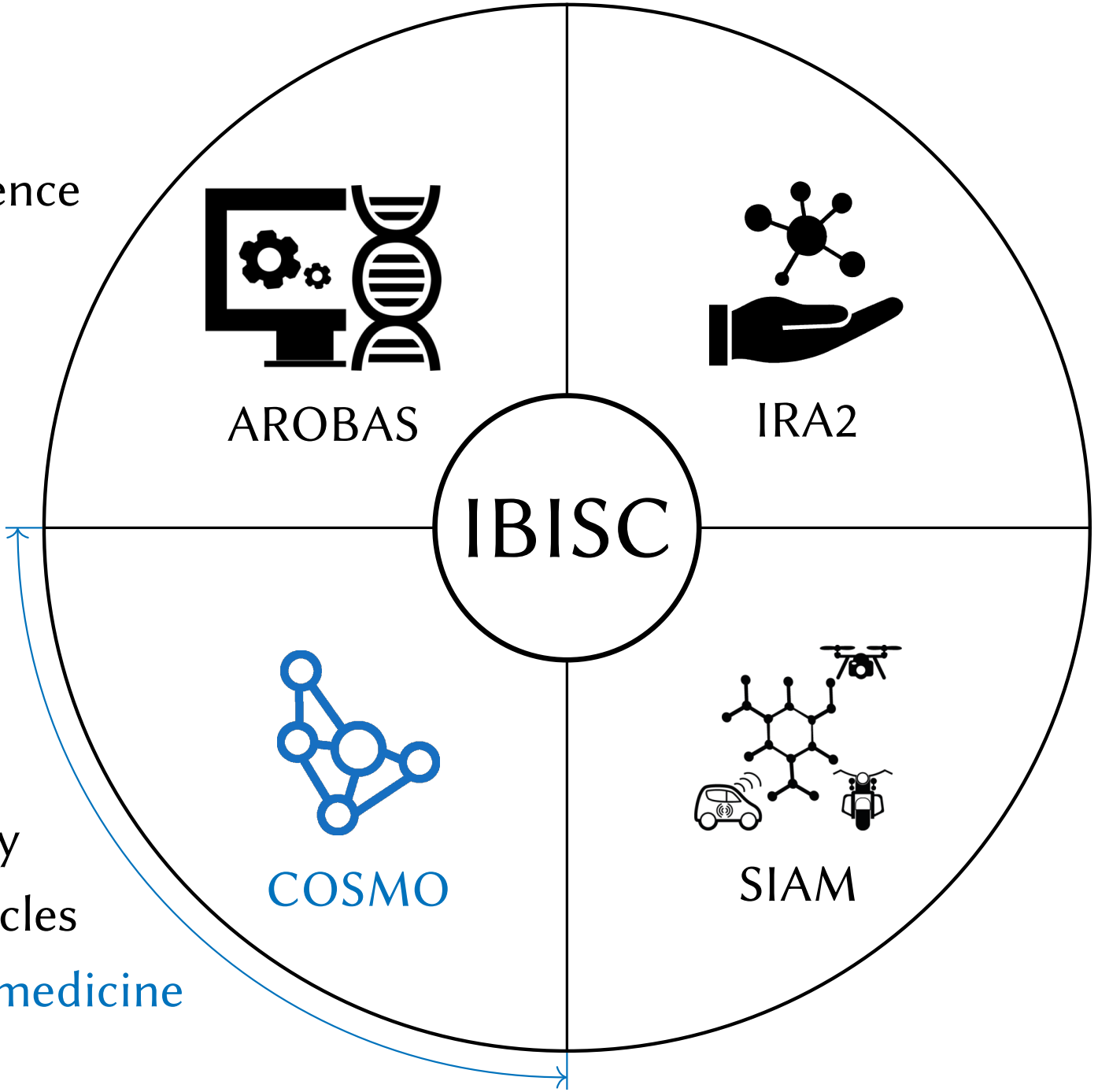
# Sequential Reprogramming of Biological Network Fate

Jérémie Pardo  Sergiu Ivanov  Franck Delaplace

IBISC, Univ. Évry, Université Paris-Saclay, France

# IBISC lab

- computer science
- robotics
- HCI



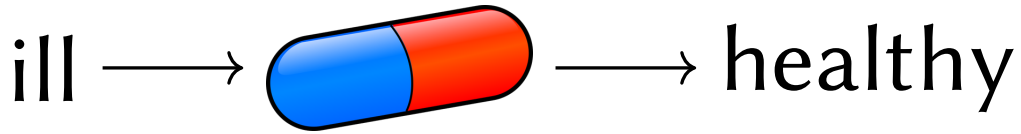

## COSMO team

- cloud
- applied theory
- next-gen vehicles
- personalised medicine

# Subcontractors

Medicine

- works in many cases



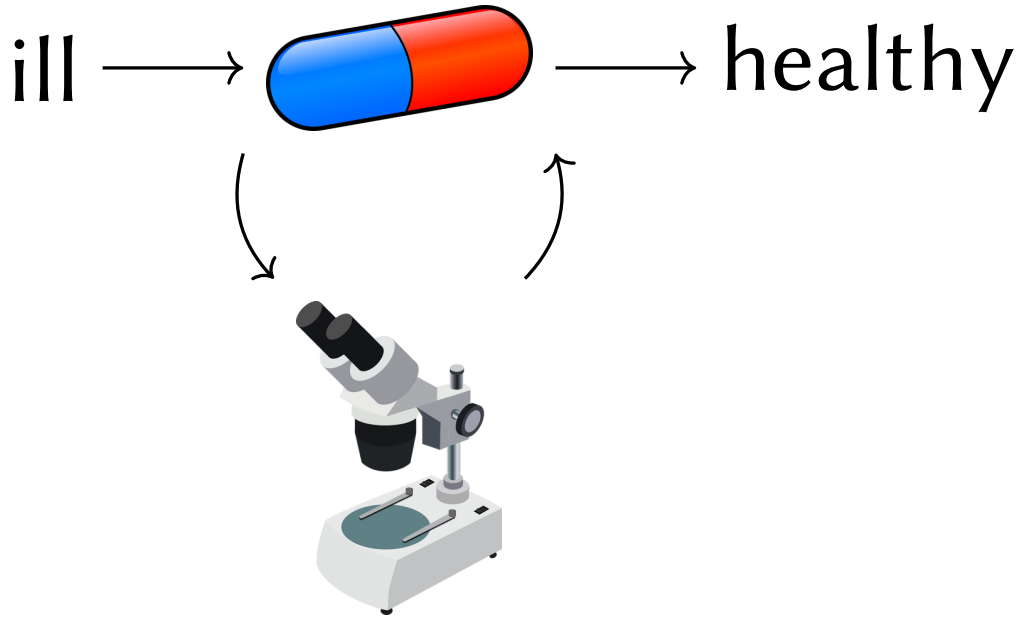
# Subcontractors

## Medicine

- works in many cases

## Biology

- needed for ill-understood diseases



# Subcontractors

## Medicine

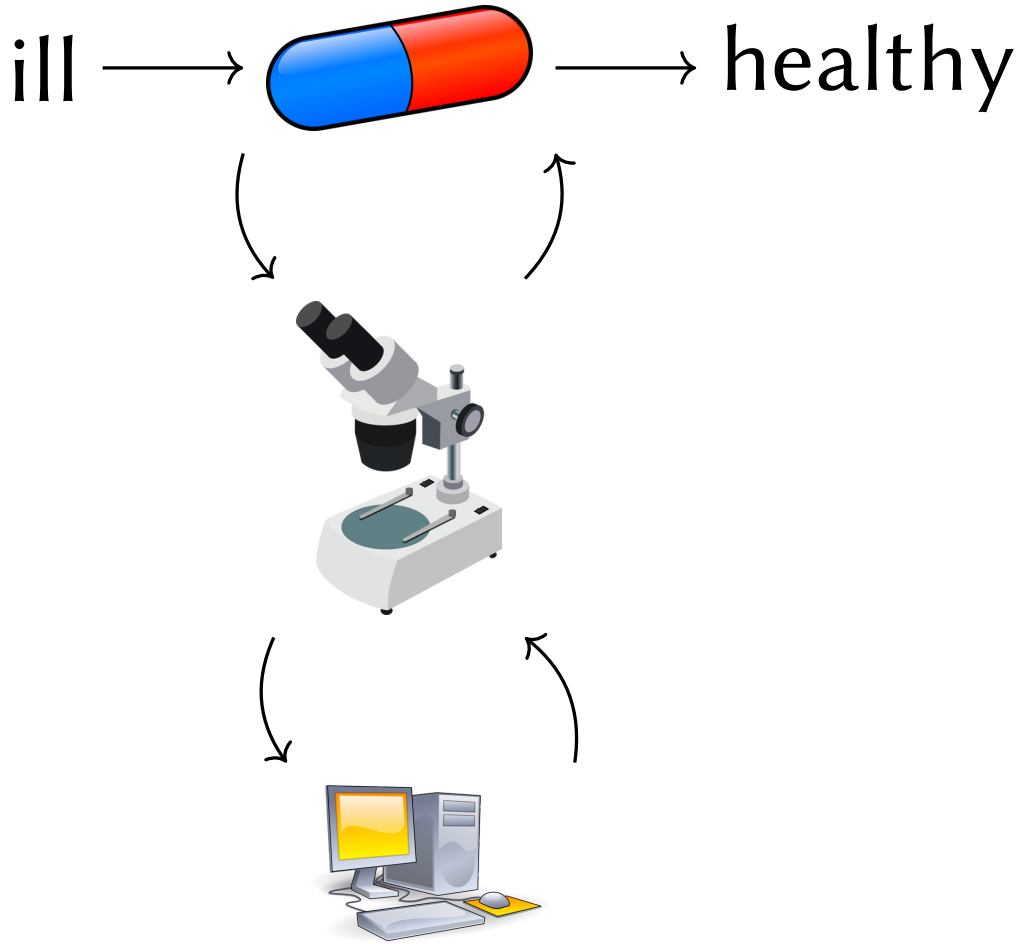
- works in many cases

## Biology

- needed for ill-understood diseases

## Computer Science

- analyse and model



# Data analysis and modelling

$\longrightarrow$ data

What does it mean?

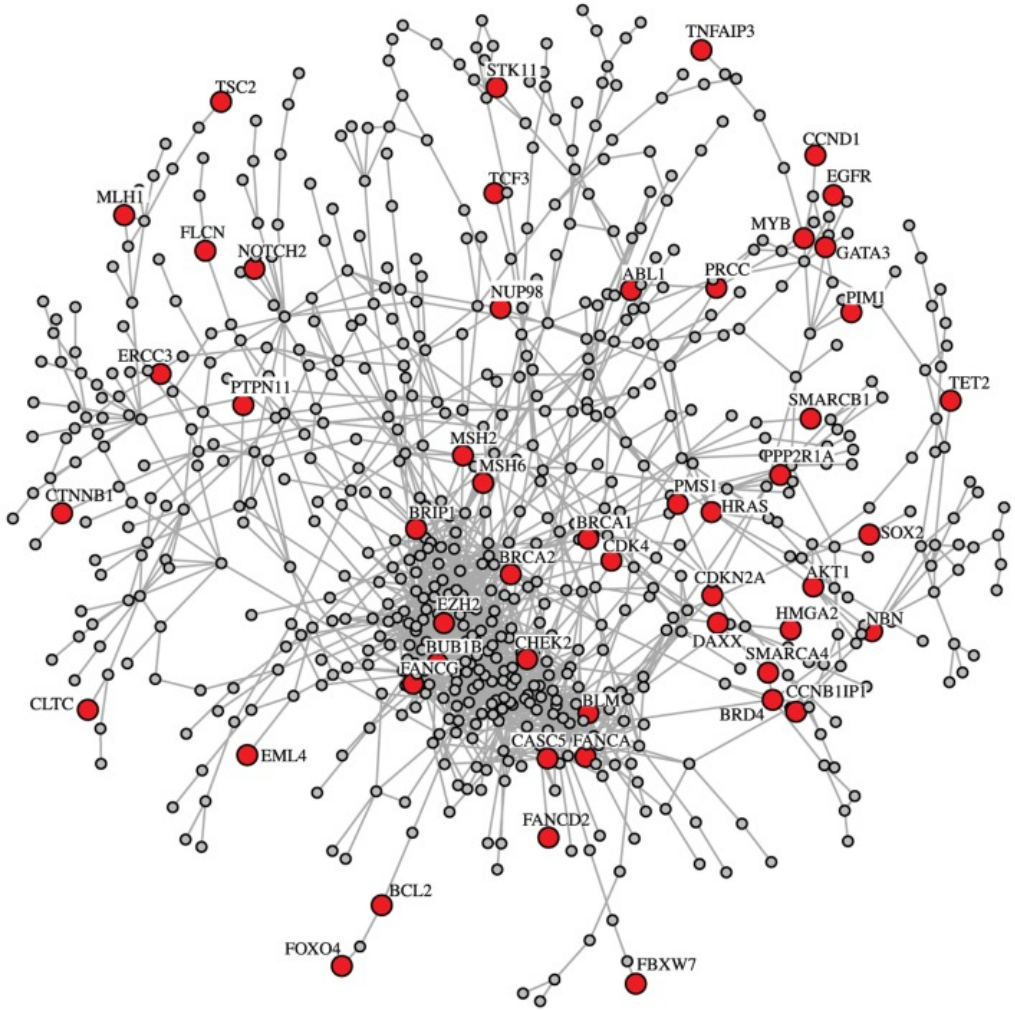
Model $=$ a simplified formal representation of a phenomenon

$\Rightarrow$ a (possible) consistent explanation of the causes

---

We analyse models.

# Our focus: cancer and GRN




Emmert-Streib, Frank et al. The gene regulatory network for breast cancer: integrated regulatory landscape of cancer hallmarks. Frontiers in genetics vol. 5 15. 3 Feb. 2014, doi:10.3389/fgene.2014.00015

# Our central dogma

"Cancer cells acquire pathological phenotypes through accumulation of mutations that perturb signaling networks."

Creixell, Pau et al. Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. Cell vol. 163,1 (2015): 202-17. doi:10.1016/j.cell.2015.08.056

# Other diseases

- Duchenne muscular dystrophy
  - muscle loss
  - expectancy: ~25–26 years

  Team “Muscular diseases”, I-STEM, Évry

- Rheumatoid arthritis
  - autoimmune
  - chronic inflammation of joints

  Team “Rheumatoid arthritis”, GenHotel, Évry
  - work on a fully annotated molecular map

- ...

# Modelling advantages of GRN

1. Networks are “easy” to analyse.
2. Networks lend themselves well to fully discrete modelling.
   - capturing non-linearity

First use of Boolean networks:

 Kauffman, Stuart. Homeostasis and Differentiation in Random Genetic Control Networks. Nature vol 224, pages 177–178 (1969).

# Continuous models

+ capture the “continuous” expression levels
  - average values
- difficult to analyse in general
  - often a linearisation is necessary
  - math is hard
- no behaviours far from average

# Outline

# Discrete/qualitative modelling

Gene expression is modelled by discrete levels.

- 0, 1, 2, ...

\+ easier to analyse

- still hard and often intractable

− difficult to construct

- what does level 1 stand for?

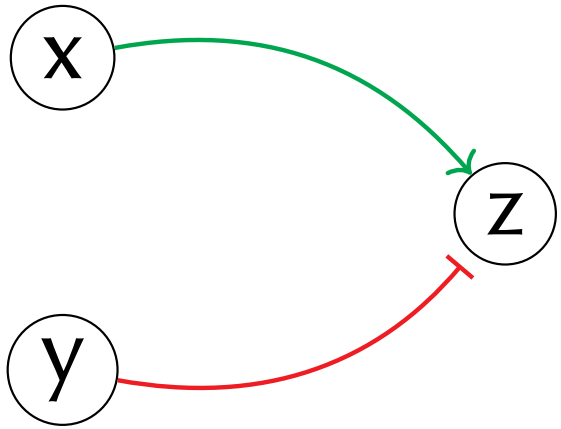
− difficult to validate

# Boolean modelling

Gene x expressed $\longrightarrow$ $x = 1$

Gene x inhibited $\longrightarrow$ $x = 0$

Gene z: promoted by x, inhibited by y:



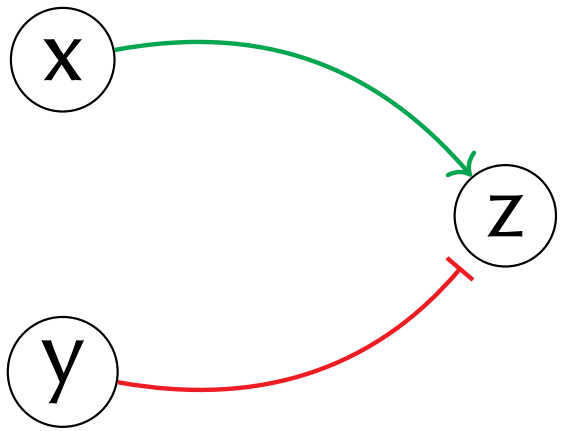
# Boolean network: definition

1. a set of Boolean variables x, y, z
2. list of Boolean update formulas for every variable

$$z = x \wedge \bar{y} \quad \leftarrow \text{negated y}$$

$$x = x$$

$$y = y$$

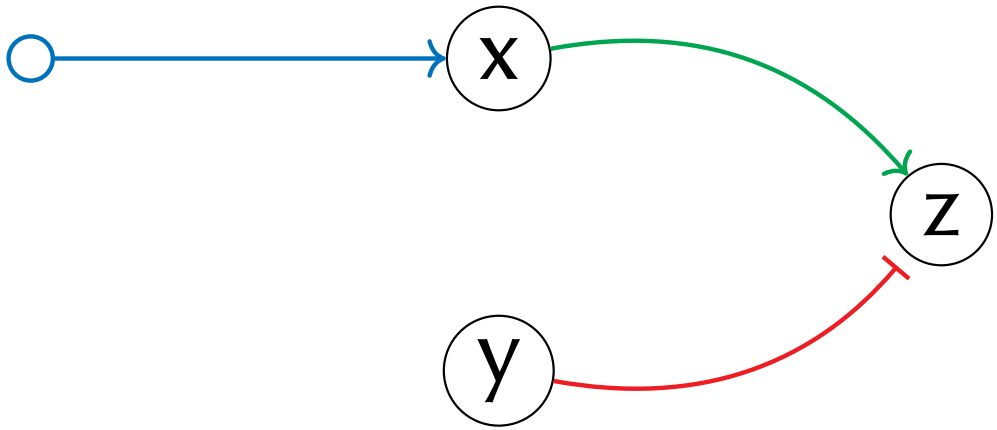



## Note

Finding the Boolean operators ($\wedge$, $\vee$) is difficult.

# Boolean network reprogramming

Equip Boolean networks with control inputs.




Control actions:

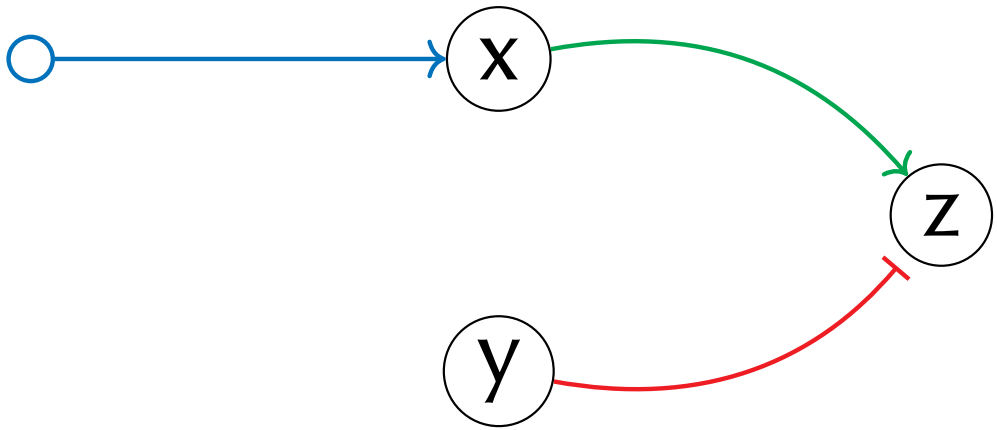
- freeze x to 0
- freeze x to 1
- unfreeze x

# Outline

# The control inference problem

Initial conditions: $x = 0, y = 0, z = 0$

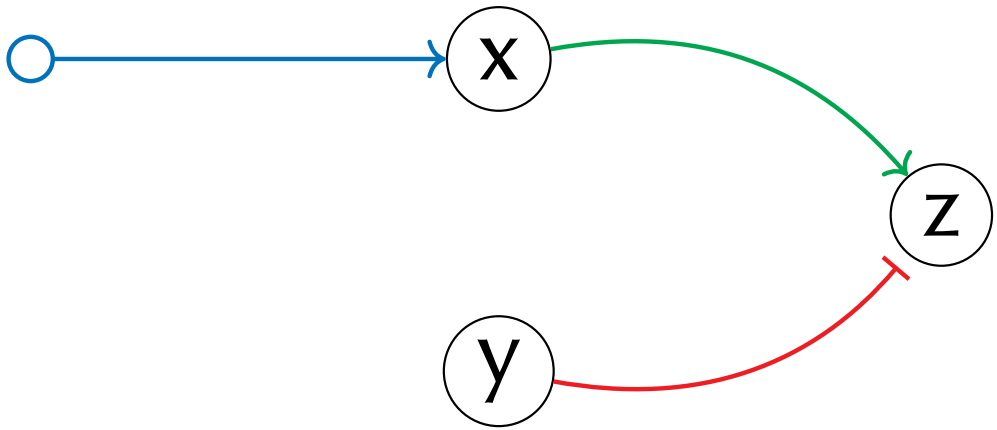
How to ensure that $z = 1$?



# The control inference problem

Initial conditions: $x = 0, y = 0, z = 0$

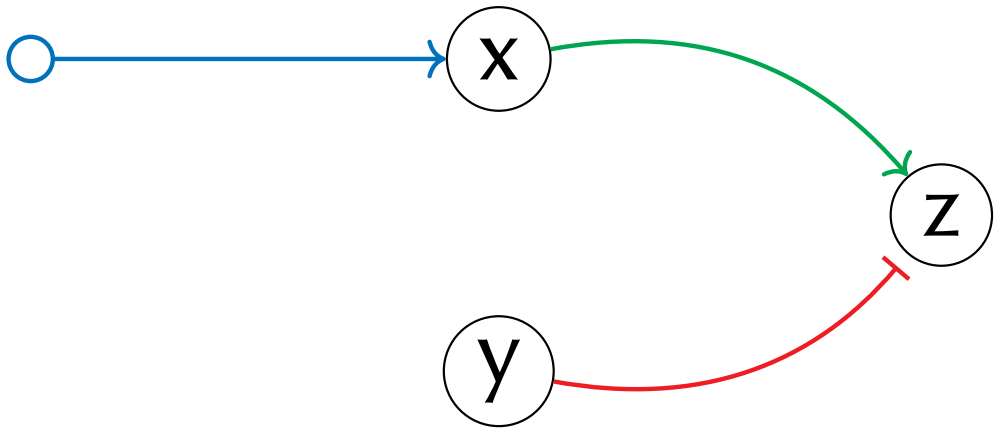
How to ensure that $z = 1$? Control $x \leftarrow 1$.



# The control inference problem

Initial conditions: $x = 0, y = 0, z = 0$

How to ensure that $z = 1$? Control $x \leftarrow 1$.




How about $x = 0, y = 1, z = 0$ ?

- assuming $z = x \wedge \bar{y}$

# Control inference is NP-hard

NP-hard $\Rightarrow$ no efficient and exact solution.

The behaviour of Boolean networks can be very complex.

Clever reductions and heuristics help a lot in practice.


Célia Biane, Franck Delaplace. Causal reasoning on Boolean control networks based on abduction: theory and application to Cancer drug discovery. IEEE/ACM Trans. Comput. Biol. Bioinform. (2018).

# Control inference for breast cancer

Cited contribution:

1. builds a Boolean network modelling breast cancer (13 nodes)
2. retrieves the main driver genes
3. segregates tumour suppressors from oncogenes
4. predicts mutations (freezing actions) leading to cancer
   - validated by literature

Célia Biane, Franck Delaplace. Causal reasoning on Boolean control networks based on abduction: theory and application to Cancer drug discovery. IEEE/ACM Trans. Comput. Biol. Bioinform. (2018).

# Outline

# Colorectal cancer: sequential tumorigenesis

"Study of the process of tumorigenesis in experimental models has suggested that at least three steps can be defined [...]"

 Eric R Fearon and Bert Vogelstein. A genetic model for colorectal tumorigenesis. Cell, 61(5):759–767, 1990.

**Note**

"[...] the total accumulation of changes, rather than their order with respect to one another, is responsible for determining the tumor's biologic properties."

# Sequential therapies: better effect

"Sequential application of anti-cancer drugs enhances cell death by re-wiring apoptotic signaling networks."

Michael Lee, Albert S. Ye, Alexandra K. Gardino, Anne Heijink, Peter Sorger, Gavin Macbeath, and Michael Yaffe. Sequential application of anti-cancer drugs enhances cell death by re-wiring apoptotic signaling networks. Cell, 149:780–794, 05 2012.

# Sequential therapies: fewer interventions

"[...] the application of perturbations in particular moments in time, and in a particular ordering, brings new reprogramming strategies, potentially requiring fewer interventions."

 Hugues Mandon, Stefan Haar, Loïc Paulevé: Temporal Reprogramming of Boolean Networks. CMSB 2017: 179-195

# Meaning of control sequences

Single control, problem statement:

1. set the values of the control inputs
2. analyse the dynamics (stable states, attractors, ...)

Typical question: reach a stable state with a given property?

- property = Boolean formula

---

Sequence of controls: when is the control changed?

# CoFaSe: change the control at any moment

| Computer science | Biology |
| --- | --- |
| Natural and OK | What does it even mean? |

Control sequences $\Rightarrow$ need to better model time.

---

"The day we understand the time evolution of subcellular elements at a level of detail comparable to physical systems governed by Newton's laws of motion seems far away."

S. Ronquist, G. Patterson, M. Brown, H. Chen, A. Bloch, L. Muir, R. Brockett, and I. Rajapakse. An Algorithm for Cellular Reprogramming. Proceedings of the National Academy of Sciences Oct 2017, 201712350; DOI: 10.1073/pnas.1712350114

# Little previous work

Hugues Mandon, Stefan Haar, Loïc Paulevé: Temporal Reprogramming of Boolean Networks. CMSB 2017: 179-195

- general solution to CoFaSe using Petri nets
- model directly derived from Boolean networks

---

S. Ronquist, G. Patterson, M. Brown, H. Chen, A. Bloch, L. Muir, R. Brockett, and I. Rajapakse. An Algorithm for Cellular Reprogramming. Proceedings of the National Academy of Sciences Oct 2017, 201712350; DOI: 10.1073/pnas.1712350114

- specific continuous model
- difficult to reinterpret for qualitative modelling
- assumes some knowledge of the biological time-dependent dynamics

# Outline

# CoFaSe is PSPACE-hard ☹

Even approximate solutions may be very hard computationally.

One-step case is as hard as reachability $\in$ PSPACE-complete.

- Can a given state be reached from another given state?

Full CoFaSe may be even harder.

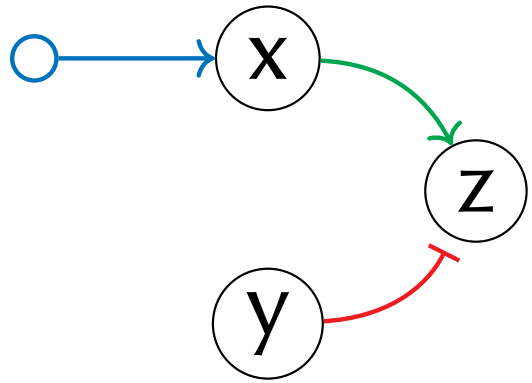
# Upper bound on the sequence length ☺

Minimal sequence = one among the shortest sequences solving a given CoFaSe problem

---

Theoretical tool:

- controlled variables (CV): $\{x\}$
- uncontrolled variables (UCV): $\{y, z\}$



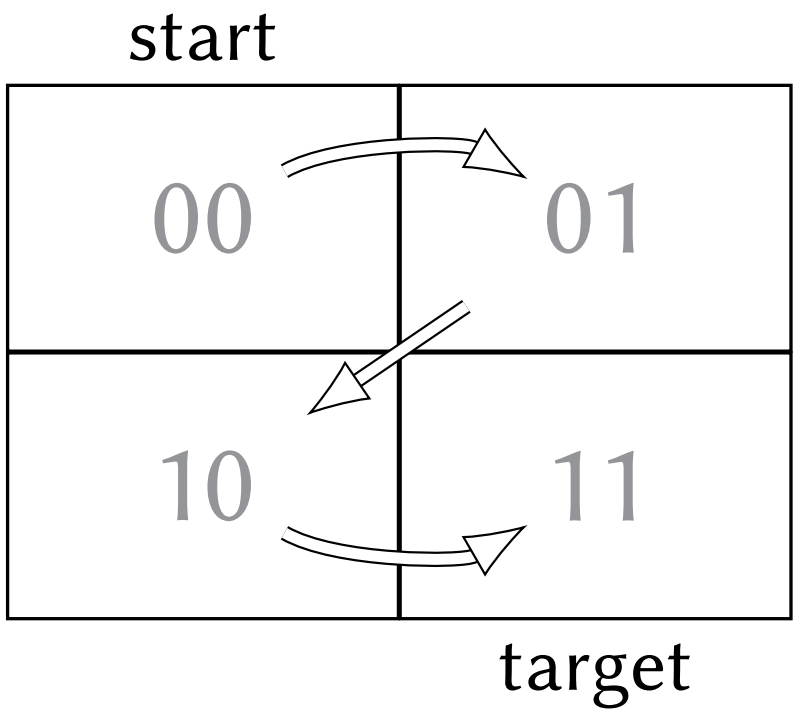

---

The length of a minimal sequence is at most $2^{|\text{UCV}|}$.

# Proof idea

Take UCV $= \{a, b\}$. Possible UCV profiles: a b $\in \{00, 01, 10, 11\}$.

Partition the state space according to the UCV profiles.




We assume the CV to be fully controllable.

Going back to a visited partition adds no value.

The formulae of UCV may forbid directly going to the target.

# UCV model biomarkers

Biomarker = a gene whose expession level helps identify the target property

Biomarkers are uncontrolled variables (UCV).

- controlling a biomarker makes no sense

The number of biomarkers is often small (< 10).

$$\Downarrow$$

Minimal control sequences are rather short!

# Outline

# Biological interpretation of CoFaSe?..

Control can be changed at any moment of time...

...of the Boolean network!

The discrete time of the network has a loose correspondence with the biological time of the modelled system.

## Flashback

| Computer science | Biology |
| --- | --- |
| Natural and OK | What does it even mean? |

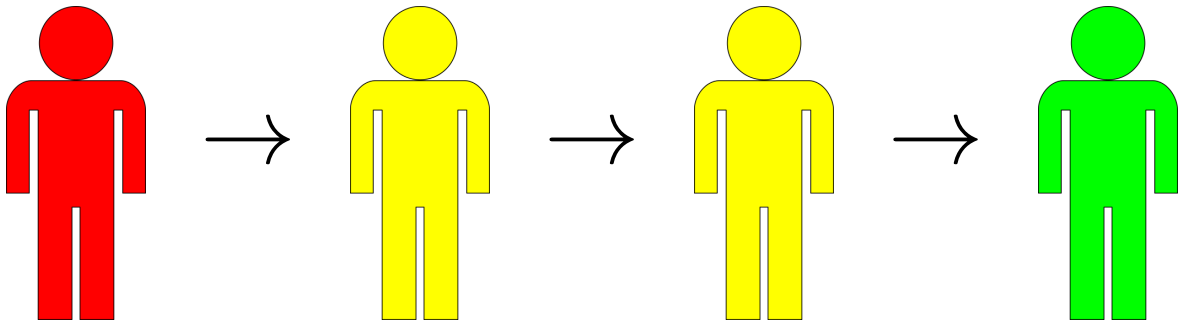
# ConEvs: change control at stable states

Stable states correspond to particular phenotypes.

Changing the control at a stable state

$\approx$

driving the system from phenotype to phenotype



# ConEvs: a bound on the sequence length

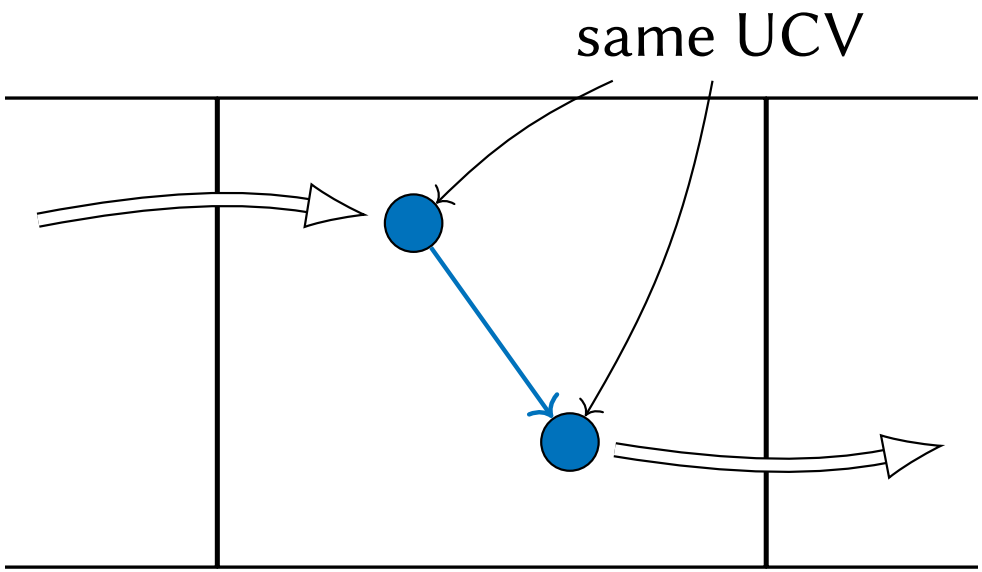
The bound for CoFaSe does not hold for ConEvs

- CoFaSe is less restricted

---

Fortunately:

The length of a minimal sequence for ConEvs is at most $2^{|\text{UCV}|+1}$.




An extra stop may be needed to prepare the CV profiles.

- control takes effect in the next step

# ConEvs: our algorithm

Input: starting states, target property.

1. Try reaching the target using one-shot control inference.

   Célia Biane, Franck Delaplace. Causal reasoning on Boolean control networks based on abduction: theory and application to Cancer drug discovery. IEEE/ACM Trans. Comput. Biol. Bioinform. (2018).

2. Enumerate new UCV profiles and try to reach them.

3. Restart from the newly reachable UCV profiles.

---

- Finds minimal parsimonious sequences.
  - control as few variables as possible per step
- Handles staying in a partition for an extra step.

# Outline

Sequential control is PSPACE-hard...

Sequential control is PSPACE-hard...

...but we have a nice upper bound on sequence length

and an algorithm exploiting it.

# Future work: asynchronous mode

Update mode = the strategy of choosing the variables to update at each step.

- Synchronous mode = all of them
- Asynchronous mode = one of them
  - non-deterministic choice

Our results work for the synchronous mode.

Asynchronous mode may offer different insight.

- stable states are the same

# Future work: non-parsimonious sequences

Sometimes minimal parsimonious sequences are not the shortest.

Changing the control inference strategy may improve the algorithm.

# Future work: biological validation

1. Infer sequences of mutations leading to certain diseases.

2. Predict sequential therapies.
   - what to administer
   - when to administer