

# Imposing constraints on network underlying logic to identify functional subgraphs

Bertrand Miannay

Equipe MeForBio, LS2N  
Equipe 11, CRCINA

Directeur de thèse : Olivier Roux (LS2N, Nantes)  
Co-encadrant : Stéphane Minvielle (CRCINA, Nantes)  
Co-encadrante : Carito Guziolowski (LS2N, Nantes)  
Co-encadrante : Florence Magrangeas (CRCINA, Nantes)

May 11, 2017

# Summary

## 1 Context

## 2 Method

- Overview
- Formalism and implementation
- Components analysis
- Space solution reduction

## 3 Application

- Data and regulatory network
- Perfect colorations generation
- Components analysis

## 4 Conclusion & perspective

## Historical

- Genes expression measure :
  - ▶ Decrease of cost/time during last 2 decades.
  - ▶ Used to compare expression profiles[LG06, PJvdR<sup>+</sup>99].
- Biological knowledge :
  - ▶ Increase of knowledge on interactions between biological entities and their roles.
  - ▶ Formalization in databases (KEGG, GO, NCI-PID, CBN, etc.).

**Accumulated regulatory knowledge and experimental observations.**

## Modelization

- Used for cellular phenomenas study [KDS<sup>+</sup>16], disease research [LLX<sup>+</sup>13, Nev01], bio-production optimization[Ate15], etc.
- Cannot work with large amounts of data
- Need pre-selection of data and network by researchers

# Summary

## 1 Context

## 2 Method

- Overview
- Formalism and implementation
- Components analysis
- Space solution reduction

## 3 Application

- Data and regulatory network
- Perfect colorations generation
- Components analysis

## 4 Conclusion & perspective

## Input/output

- From a regulatory network  $\Rightarrow$  The entities strongly associated merged in **components**
- From a set of observations  $\Rightarrow$  **Set of components** which could be used for next analysis.

## Specificity

- Method based on graph coloring approaches [TCSR<sup>+</sup>15]
- Research of the “perfect colorations”
- Implemented in ASP (Answer Set Programming)

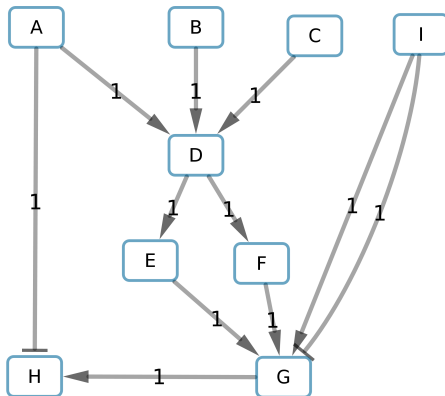
## Instanciation

**Graph** : Set of **oriented, signed** (activator or inhibitor), **weighted edges** between **nodes**

**Target** : A node with at least, one predecessor (or regulator).

## Candidate solutions generation

**Colored graph** : A graph in which each node is associated to a sign: up or down

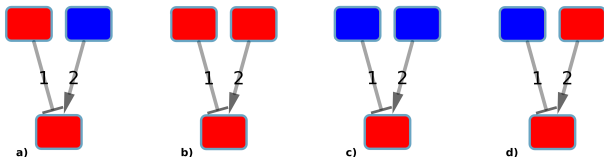


## Constraint

**Consistent target coloring** : A colored target, which is explained by at least **one** predecessor's coloration.

**Perfect target coloring** : A colored target, which is explained by **all** predecessor's coloration.

**Imperfect weighted regulator** : **Weight of the edge** between an imperfect target and its inconsistent regulator.



|                              | a | b | c | d |
|------------------------------|---|---|---|---|
| Consistent node coloring     | X | ✓ | ✓ | ✓ |
| Perfect target coloring      | X | X | X | ✓ |
| Imperfect weighted regulator | 3 | 1 | 2 | 0 |

## Optimization

- 1 : **Inconsistency minimization** : Colored graphs with the minimal number of inconsistent targets.
- 2 : **Imperfect node coloring minimization** : Colored graph with the minimal number of imperfect targets.
- 3 : **Imperfect weighted regulator minimization** : Colored graphs with the minimal sum of imperfect weighted regulator

## Components identification

**Component** : set of nodes with **correlated** (positive or negative) coloration in perfect solutions.

| node         | A    | B    | C    | D    |
|--------------|------|------|------|------|
| coloration 1 | up   | up   | up   | down |
| coloration 2 | up   | down | down | up   |
| coloration 3 | down | up   | up   | down |
| coloration 4 | down | down | down | up   |

## Example

When B is up (down), C is up (down) too  $\Rightarrow$  Positive correlation.

When B is up (down), D is down (up)  $\Rightarrow$  Negative correlation.

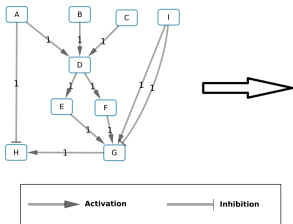
**Component syntax** : "B +, C +, D -"



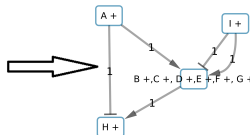
## Maximal similarity computing

**Component configurations** : Two possibilities of colorations for a component.

**Maximal similarity (MS)** : For a set of observation (nodes associated to signs) and a component : the maximal percentage of observed nodes in the component with the same sign as in the configurations.



|               | A    | B    | C    | D    | E    | F    | G    | H    | I    |
|---------------|------|------|------|------|------|------|------|------|------|
| Coloration 1  | down | down | down | down | down | down | down | up   | up   |
| Coloration 2  | down | down | down | down | down | down | down | down | up   |
| Coloration 3  | down | up   | up   | up   | up   | up   | up   | up   | up   |
| Coloration 4  | up   | up   | up   | up   | up   | up   | up   | down | up   |
| Coloration 5  | up   | up   | up   | up   | up   | up   | up   | up   | up   |
| Coloration 6  | up   | down | down | down | down | down | down | down | down |
| Coloration 7  | down | down | down | down | down | down | down | up   | down |
| Coloration 8  | down | down | down | down | down | down | down | down | down |
| Coloration 9  | down | up   | up   | up   | up   | up   | up   | up   | down |
| Coloration 10 | up   | up   | up   | up   | up   | up   | up   | down | down |
| Coloration 11 | up   | up   | up   | up   | up   | up   | up   | up   | down |
| Coloration 12 | up   | down | down | down | down | down | down | down | down |



## Toy example

**Configuration** : Component : "B +, C +, D +, E +, F +, G +"

- (B,up), (C,up), (D,up), (E,up), (F,up), (G,up)
- (B,down), (C,down), (D,down), (E,down), (F,down), (G,down)

**Observations** : (B,up), (C,up), (D,up), (E, down)

$$MS = \frac{3}{4}$$

## coloring property

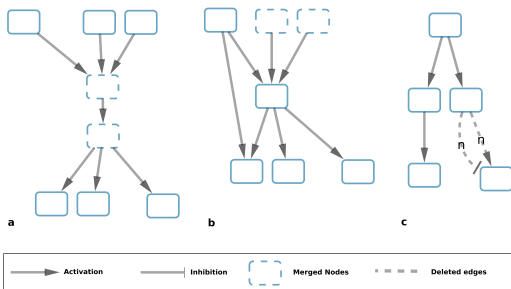
**Symmetric reduction** : A colored graph and its reverse coloring ( up  $\leftrightarrow$  down) have the same optimization scores.

## Topological property

**Consistent coloring** : Identification of nodes which will have a sign correlation in consistent solutions (Figure 1-a)

**Imperfect coloring** : Identification of nodes which will have a sign correlation in candidate solutions with minimized imperfect colorations (Figure 1-b)

**Edges balance** : Deletion of balanced edges (Figure 1-c)



**Figure:** Patterns searched by the 3 reductions methods used in this study. a : nodes correlated in consistent solution. b : nodes correlated sharing the same target. c : edges with same weight, root, target and opposite signs.

# Summary

## 1 Context

## 2 Method

- Overview
- Formalism and implementation
- Components analysis
- Space solution reduction

## 3 Application

- Data and regulatory network
- Perfect colorations generation
- Components analysis

## 4 Conclusion & perspective

## Context

- Analysis of genes expression data from Multiple myeloma myeloma patients with regulatory network.

## Data

Gene expression profiles (GEP) from :

- 602 multiple myeloma patients (myeloma cells : **MM**)
- 9 healthy donors (normal plasma cells : **NPC**).

Identification for each GEP of the over-expressed (up) and under-expressed (down) genes.

## Regulatory network

From Pathway Interaction database (PID)

- Extraction of the downstream events from three signaling pathways (IL6/IL6-R, IGF1/IGF1-R and CD40) [Kle10] to the variant genes

Generation of an induced subgraph from NCI-PID, containing 2269 nodes, 2683 edges and connecting 529 variant genes.

## Graph reduction

- Consistent coloring, Imperfect coloring, Edges balance reductions
- New graph with 193 nodes 389 edges

Table: Computation time.

| Graph           | number of nodes | number of edges | time    |
|-----------------|-----------------|-----------------|---------|
| Generated graph | 2269            | 2683            | 72',12" |
| Reduced graph   | 193             | 389             | 14"     |

## Perfect solutions and components identification

- **16834** coloring model
- 15 components identified from the regulatory network
- Only **2 components (2 and 6)** include more than one node.

## Components validation and specification

- Computing then comparison between  $MS^{MC}$  and  $MS^{NPC}$  for each component
- Only **component 2** is statically different between MC and NPC.

## Biological analysis : gene ontology

- Genes in the component 2 are strongly associated to cancer pathways.
- Genes in the component 6 are not associated to specific pathway

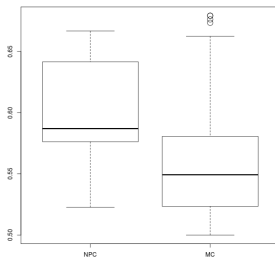


Figure: MS comparison between Normal Plasma Cells and Myeloma Cells

# Summary

## 1 Context

## 2 Method

- Overview
- Formalism and implementation
- Components analysis
- Space solution reduction

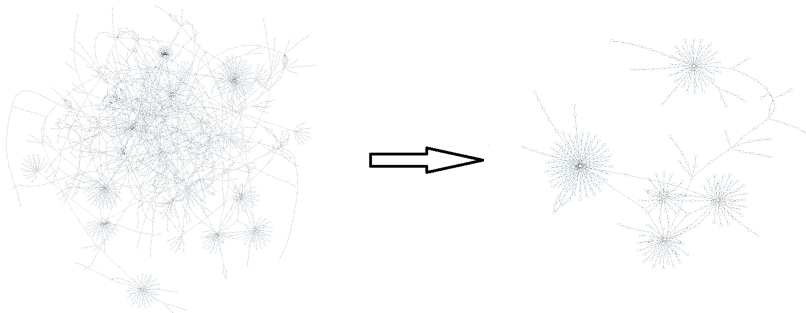
## 3 Application

- Data and regulatory network
- Perfect colorations generation
- Components analysis

## 4 Conclusion & perspective

## Conclusion

- Identification of functional subgraphs/components from a regulatory network.
- Identification of specific components from dataset



## Perspective

- Use other database (trrust, causal bionet, etc.)
- Improve topological reduction
- Identification of specific components to sub-type of MM patient (poor/good prognosis)



**Merci de votre attention**

# Références I



**Ozlem Ates.**

Systems Biology of Microbial Exopolysaccharides Production.  
*Frontiers in bioengineering and biotechnology*, 3:200, 2015.



**Aristotelis Kittas, Aurélien Delobelle, Sabrina Schmitt, Kai Breuhahn, Carito Guziolowski, and Niels Grabe.**

Directed random walks and constraint programming reveal active pathways in hepatocyte growth factor signaling.  
*FEBS Journal*, 283(2):350–360, jan 2016.



**Bernard Klein.**

Positioning NK-kappaB in multiple myeloma.  
*Blood*, 115(17):3422–4, April 2010.



**Tim Lenoir and Eric Giannella.**

The emergence and diffusion of DNA microarray technology.  
*Journal of biomedical discovery and collaboration*, 1(1):11, jan 2006.



**Wei Liu, Chunquan Li, Yanjun Xu, Haixiu Yang, Qianlan Yao, Junwei Han, Desi Shang, Chunlong Zhang, Fei Su, Xiaoxi Li, Yun Xiao, Fan Zhang, Meng Dai, and Xia Li.**

Topologically inferring risk-active pathways toward precise cancer classification by directed random walk.  
*Bioinformatics (Oxford, England)*, 29(17):2169–77, September 2013.



**J R Nevins.**

The Rb/E2F pathway and cancer.  
*Human molecular genetics*, 10(7):699–703, apr 2001.



**C M Perou, S S Jeffrey, M van de Rijn, C A Rees, M B Eisen, D T Ross, A Pergamenschikov, C F Williams, S X Zhu, J C Lee, D Lashkari, D Shalon, P O Brown, and D Botstein.**

Distinctive gene expression patterns in human mammary epithelial cells and breast cancers.  
*Proceedings of the National Academy of Sciences of the United States of America*, 96(16):9212–7, aug 1999.



**Sven Thiele, Luca Cerone, Julio Saez-Rodriguez, Anne Siegel, Carito Guziolowski, and Steffen Klamt.**

Extended notions of sign consistency to relate experimental data to signaling and regulatory network topologies.  
*BMC bioinformatics*, 16(1):345, jan 2015.