

Revisiting and extending probabilistic Boolean networks

Zachary ASSOUMANI

20 septembre 2023

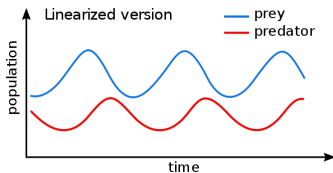
Encadré par : Elisabeth REMY & Claudine CHAOUIYA



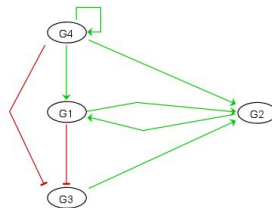
Modélisations mathématiques

Appliquées à la biologie

- Représenter l'information dans des modèles abstraits
- Pour comprendre et prédire le fonctionnement des organismes



Équations différentielles.



Réseaux de régulation booléens.

Plan

I - Présentation du sujet

Réseaux booléens, réseaux booléens probabilistes, dynamique

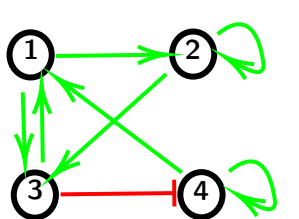
II - Mes contributions

Construction de modèles, études de cas

Réseaux booléens (Kauffman, 1969)

Interactions entre les gènes

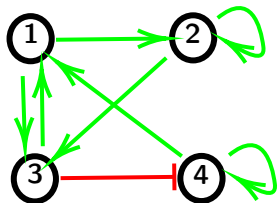
- Régulations **positives** (= activations)
- Régulations **négatives** (= inhibitions)



Réseaux booléens

Fonctions de régulation

- Une fonction booléenne $\{0, 1\}^n \rightarrow \{0, 1\}$ associée à chaque gène
- Calcule l'état suivant à partir de l'état actuel



$$f_1(x) = \begin{cases} 1 & \text{if } (x_3 = 1) \vee (x_4 = 1) \\ 0 & \text{sinon} \end{cases}$$

$$f_2(x) = \begin{cases} 1 & \text{if } (x_1 = 1) \vee (x_2 = 1) \\ 0 & \text{sinon} \end{cases}$$

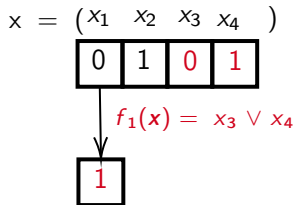
$$f_3(x) = \begin{cases} 1 & \text{if } (x_1 = 1) \wedge (x_2 = 1) \\ 0 & \text{sinon} \end{cases}$$

$$f_4(x) = \begin{cases} 1 & \text{if } (x_3 = 0) \wedge (x_4 = 1) \\ 0 & \text{sinon} \end{cases}$$

Réseaux booléens

Fonctions de régulation

- Une fonction booléenne $\{0, 1\}^n \rightarrow \{0, 1\}$ associée à chaque gène
- Calcule l'état suivant à partir de l'état actuel



$$f_1(x) = \begin{cases} 1 & \text{if } (x_3 = 1) \vee (x_4 = 1) \\ 0 & \text{sinon} \end{cases}$$

$$f_3(x) = \begin{cases} 1 & \text{if } (x_1 = 1) \wedge (x_2 = 1) \\ 0 & \text{sinon} \end{cases}$$

$$f_2(x) = \begin{cases} 1 & \text{if } (x_1 = 1) \vee (x_2 = 1) \\ 0 & \text{sinon} \end{cases}$$

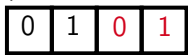
$$f_4(x) = \begin{cases} 1 & \text{if } (x_3 = 0) \wedge (x_4 = 1) \\ 0 & \text{sinon} \end{cases}$$

Réseaux booléens

Fonctions de régulation

- Une fonction booléenne $\{0, 1\}^n \rightarrow \{0, 1\}$ associée à chaque gène
- Calcule l'état suivant à partir de l'état actuel

$$x = (x_1 \ x_2 \ x_3 \ x_4)$$



$$f(x) = (f_1(x) \ f_2(x) \ f_3(x) \ f_4(x))$$

$$f_1(x) = \begin{cases} 1 & \text{if } (x_3 = 1) \vee (x_4 = 1) \\ 0 & \text{sinon} \end{cases}$$

$$f_3(x) = \begin{cases} 1 & \text{if } (x_1 = 1) \wedge (x_2 = 1) \\ 0 & \text{sinon} \end{cases}$$

$$f_2(x) = \begin{cases} 1 & \text{if } (x_1 = 1) \vee (x_2 = 1) \\ 0 & \text{sinon} \end{cases}$$

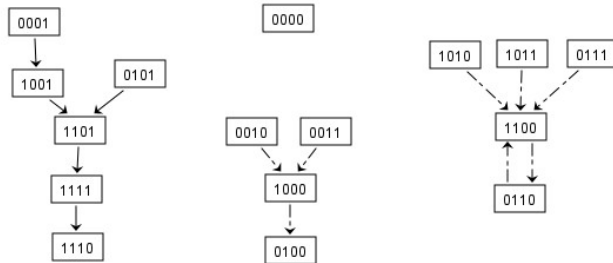
$$f_4(x) = \begin{cases} 1 & \text{if } (x_3 = 0) \wedge (x_4 = 1) \\ 0 & \text{sinon} \end{cases}$$

Réseaux booléens

Transitions d'états

0000 → 0000
 0001 → 1001
 0010 → 1000
 0011 → 1000
 0100 → 0100
 0101 → 1101
 0110 → 1100
 0111 → 1100
 1000 → 0100
 1001 → 1101
 1010 → 1100
 1011 → 1100
 1100 → 0110
 1101 → 1111
 1110 → 1110
 1111 → 1110

Graphe de transition d'états (STG) et ses attracteurs

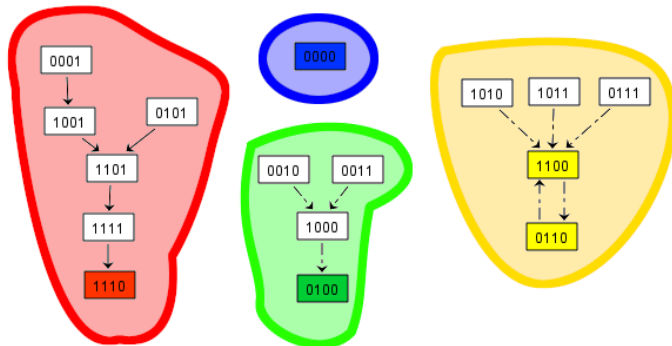


Réseaux booléens

Transitions d'états

0000 → 0000
 0001 → 1001
 0010 → 1000
 0011 → 1000
 0100 → 0100
 0101 → 1101
 0110 → 1100
 0111 → 1100
 1000 → 0100
 1001 → 1101
 1010 → 1100
 1011 → 1100
 1100 → 0110
 1101 → 1111
 1110 → 1110
 1111 → 1110

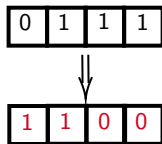
Graphe de transition d'états (STG) et ses attracteurs



Réseaux booléens : mode de mise à jour

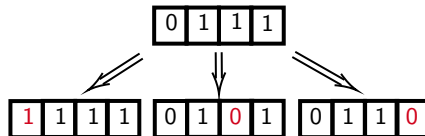
Synchrone

Tous les gènes actualisés simultanément



Asynchrone

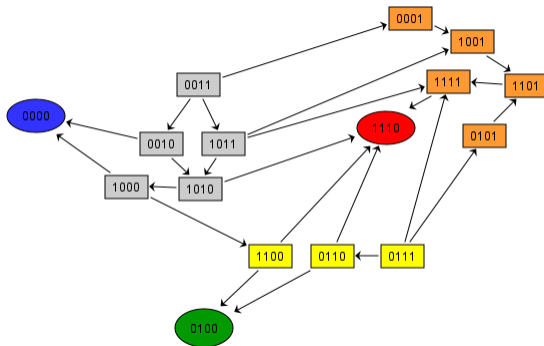
Un seul gène actualisé à la fois, choisi au hasard



→ Peut modifier les attracteurs !

Réseaux booléens : mode de mise à jour

Exemple : STG du même réseau booléen, mais en mise à jour asynchrone

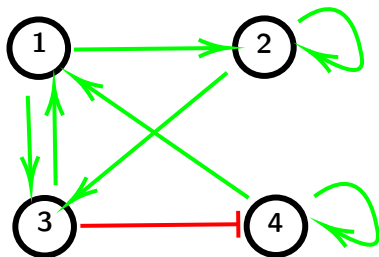


- Dans le bassin de l'attracteur rouge
- Dans les bassins des attracteurs rouge et vert
- Dans les bassins des trois attracteurs

Réseaux booléens : ordre partiel des fonctions (Cury, Monteiro, Chaouiya, 2019)

Quel choix de fonctions ?

- Fonction **positive** en x_j : $\forall x, f_i(\bar{x}^{j \leftarrow 1}) \geq f_i(\bar{x}^{j \leftarrow 0})$ (g_j activateur de g_i)
- Fonction **négative** en x_j : $\forall x, f_i(\bar{x}^{j \leftarrow 0}) \geq f_i(\bar{x}^{j \leftarrow 1})$ (g_j inhibiteur de g_i)



Pour chaque gène g_i , on considère \mathcal{F}_i l'ensemble des fonctions booléennes reflétant les régulations du gène i .

$$\begin{aligned} \text{ex : } \mathcal{F}_1 &= \{x_1 \vee x_4, x_1 \wedge x_4\} \\ \mathcal{F}_4 &= \{\neg x_3 \vee x_4, \neg x_3 \wedge x_4\}. \end{aligned}$$

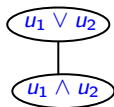
Boolean Networks : ordre partiel des fonctions (Cury, Monteiro, Chaouiya, 2019)

Quel choix de fonctions ?

- L'ensemble \mathcal{F}_i est *ordonnable* :

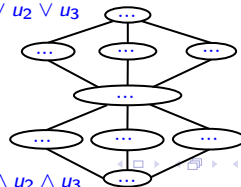
$$\forall f, f' \in \mathcal{F}_i, \quad f \preceq f' \iff \mathbb{T}(f) \subseteq \mathbb{T}(f')$$

$p = 2$ régulateurs :



$p = 3$ régulateurs :

$$\sup \mathcal{F}_i = u_1 \vee u_2 \vee u_3$$

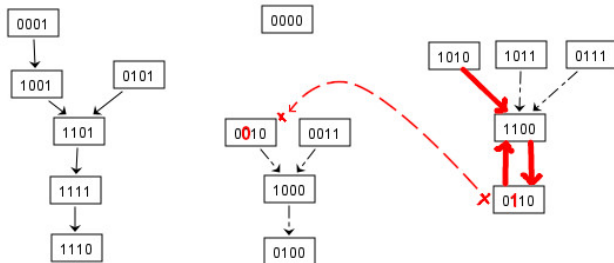


$$\inf \mathcal{F}_i = u_1 \wedge u_2 \wedge u_3$$

Réseaux booléens probabilistes

Perturbations

- À chaque itération, un gène est inversé avec une probabilité $p \ll 1$
- Proviens du bruit biologique

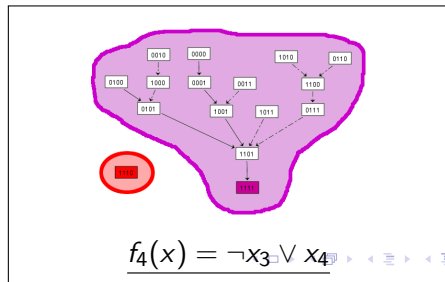
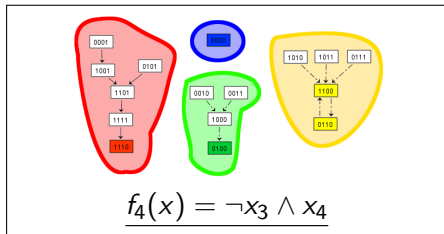


Dans cet exemple, le deuxième gène est **inversé** de '1' vers '0', et l'état change d'attracteur.

Réseaux booléens probabilistes (Shmulevich et al., 2001)

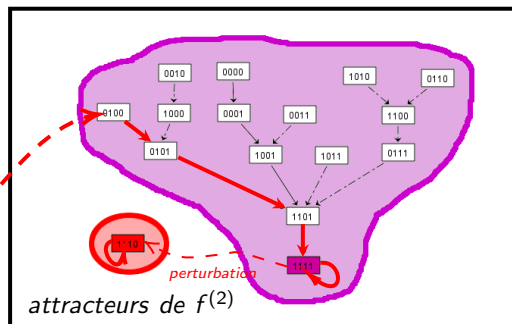
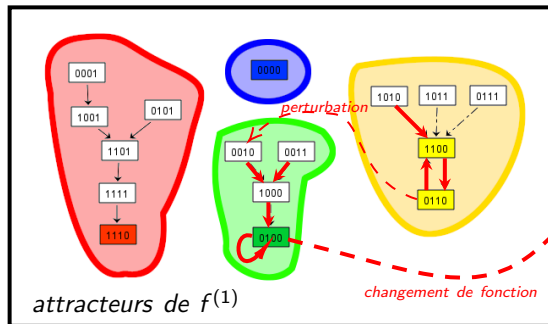
Probabilistic Boolean Networks (PBNs)

- À chaque itération avec une proba q , les fonctions de régulation sont re-sélectionnées. Contextes $F = (F^1, \dots, F^m)$, de poids respectifs $C = (c_1, \dots, c_m)$.
- Changer de fonction peut altérer le 'paysage des attracteurs' dans l'espace d'états.



Réseaux booléens probabilistes

Exemple de simulation de PBN à 2 contextes



Outil : distribution stationnaire empirique

Algorithme [Kim et al., 2003]

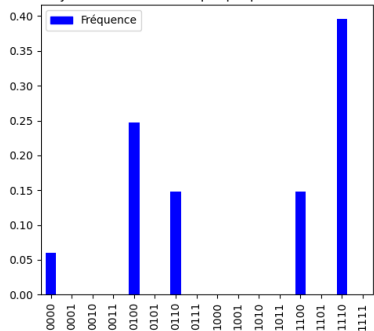
Pour i de 1 à R :

- Initialiser $x_0 \in \{0, 1\}^n$
- Appliquer T itérations
- Consigner les états des N itérations suivantes

Retourner l'histogramme des états visités : les $S_{TNR}(x)$.

→ Échantillonne les états à long terme

zhou sync - Distribution empirique pour 1000 simulations



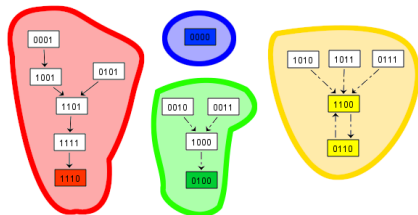
Application à l'exemple précédent,
 $N = R = 1000$, $T = 100$.

Mes contributions

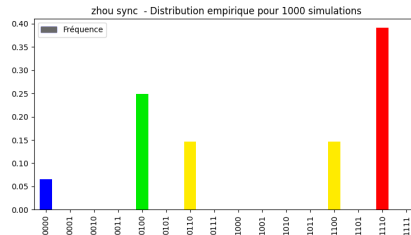
- Distribution empirique pour identifier les attracteurs des BN
 - Génération aléatoire de modèles BN et extension en PBN
- Illustration : Effet du choix d'une fonction voisine sur la dynamique (exemples synthétique + biologique)

Distribution empirique et attracteurs : BN synchrone

STG :



Loi stationnaire :



→ Les états transients ont une distribution nulle

→ Deux états d'un même attracteur ont la même fréquence

Distribution empirique et attracteurs : BN synchrone

Résultat 1

$$\exists T_0 ; \forall T \geq T_0, \quad S_{TNR}(x) > 0 \implies x \text{ appartient à un attracteur } A$$

Résultat 2

Pour A attracteur, $x \in A$,

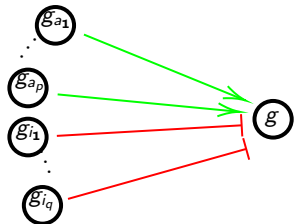
$$\lim_{N, R \rightarrow \infty} S_{TNR}(x) = \frac{|B_A|}{|A| \cdot 2^n}$$

- Les états de même hauteur \Rightarrow détectés comme un (ou plusieurs) attracteur(s).
- Prévalence de A : $\frac{B_A}{2^n} \simeq |A| \cdot S_{TNR}(x)$
marge d'erreur en $\mathcal{O}(\frac{1}{\sqrt{R}} + \frac{1}{N})$

Génération aléatoire de modèles BN

Algorithme

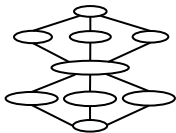
- Paramètres : n (nombre de gènes), k (nombre de régulateurs par gène), p_{neg} (probabilité qu'un régulateur soit négatif)
- **Étape 1** : construire un graphe
- **Étape 2** : pour chaque gène, déterminer la fonction de régulation "par défaut"



Fonction par défaut de g : aucun inhibiteur, au moins un activateur.

$$f_g(x) = \left(\bigvee_{j=1}^p x_{a_j} \right) \wedge \neg \left(\bigvee_{j=1}^p x_{i_j} \right)$$

Construction du PBN des fonctions voisines



Objectif : dans un BN, chaque gène g_i a une fonction de référence f_i .
On souhaite construire un PBN avec leurs fonctions voisines dans l'ordre partiel, appartenant à \mathcal{F}_i .

→ Le module Java `FUNCTIONHOOD` (Cury, Monteiro, Chaouiya, 2019) aide au calcul des enfants ou des parents d'une fonction dans le diagramme de Hasse.

Algorithme

- Paramètres : bn (réseau booléen), i_{modifs} (indices des gènes dont les fonctions sont à modifier), $dist$ (direction à explorer parmi les voisins (= parents, enfants...))
- **Pour chaque** $i \in i_{modifs}$: calcul des voisins de f_i dans \mathcal{F}_i en direction $dist$
- Construction du PBN admettant pour fonctions de régulation des g_i : $\{f_i + \text{voisins de } f_i\}$

Études de cas

Objectif : étudier l'effet du choix de fonctions voisines sur les attracteurs

- À partir d'un BN initial
- Étendre les fonctions de certains gènes pour former un PBN
- Analyser la dynamique à long terme du BN et du PBN

Étude de cas I : modèle synthétique

Recherche de modèles

- **Entrée :** \mathcal{C}_1 et \mathcal{C}_2 (deux conditions), *dist* (direction à explorer parmi les voisins)
- **Sortie :** *bn* (réseau booléen dont les attracteurs vérifient \mathcal{C}_1 ,
pbn (son PBN aux fonctions étendues vers *dist*, dont les attracteurs vérifient \mathcal{C}_2)

Les conditions portent sur le nombre et la taille des attracteurs.

Exemples : $\mathcal{C}_1(a) = "a \text{ ne contient que des points fixes}"$

$\mathcal{C}_2(a) = "a \text{ contient un attracteur de taille 2}"$

Étude de cas I : modèle synthétique

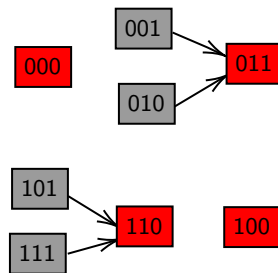
On recherche :

$C_1(a) = "a \text{ possède 4 points fixes}"$

$C_2(a) = "a \text{ possède 3 points fixes}"$

$dist = "enfants"$

$$\begin{cases} f_1(x) = x_1 \\ f_2(x) = x_2 \vee x_3 \\ f_3(x) = (\neg x_1 \wedge x_2) \vee (\neg x_1 \wedge x_3) \end{cases}$$



STG du réseau.

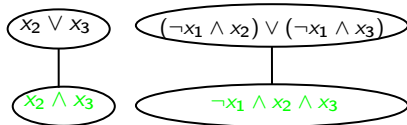
Étude de cas I : modèle synthétique

$C_1(a) = "a \text{ possède 4 points fixes}"$

$C_2(a) = "a \text{ possède 3 points fixes}"$

$dist = "enfants"$

Enfants dans les diagrammes de Hasse :

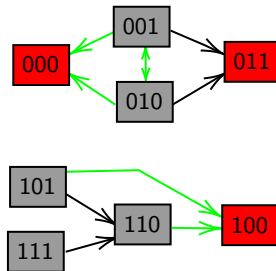


Fonctions admises pour chaque gène :

$$F_1 = \{x_1\}$$

$$F_2 = \{x_2 \vee x_3, x_2 \wedge x_3\}$$

$$F_3 = \{(\neg x_1 \wedge x_2) \vee (\neg x_1 \wedge x_3), \\ \neg x_1 \wedge x_2 \wedge x_3\}$$



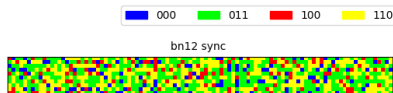
STG du réseau.

Étude de cas I : modèle synthétique

Observations

- 110 n'est plus un état stable.
- Le bassin strict de 011 se partage avec le bassin large de 000.

Résultats de $R = 1000$ simulations. Couleur = attracteur atteint (000, 011, 100, 110).



(a) BN initial :

11.5% / 41.2% / 12.0% / 35.3%

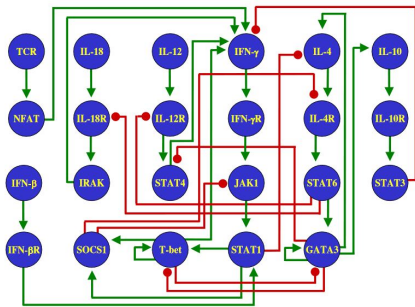


(b) PBN étendu aux enfants :

17.2% / 32.3% / 50.5% / 0%

Étude de cas II : modèle biologique Th_23

On étudie un réseau booléen à $n = 23$ gènes, modélisant la différenciation des lymphocytes T auxiliaires ('Th').



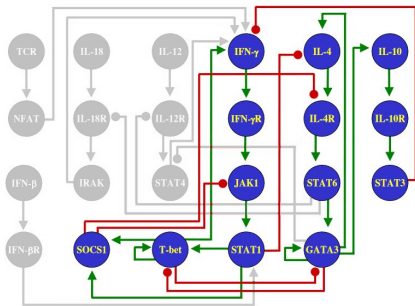
Grappe de régulation du réseau Th_23
(Mendoza & Xenarios, 2006)

Il possède 3 états stables :

- **Th0** : aucune activation
- **Th1** : activation de [IFN γ , IFN γ R, SOCS1, **Tbet**]
- **Th2** : activation de [**GATA3**, IL-10, IL-10R, **IL4**, **IL4R**, STAT3, STAT6]

Chacun correspond à un type cellulaire, dont les marqueurs sont écrits en couleur.

Étude de cas II : modèle biologique Th_23

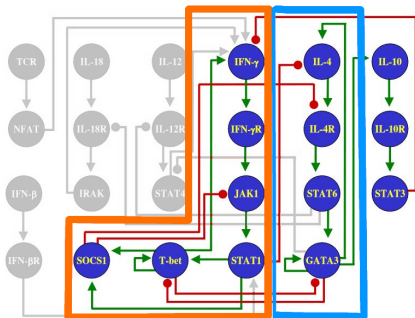


Partant de $x_0 = [IFNg]$, quels attracteurs atteints pour :

- Réseau initial / extension des fonctions de **Tbet / GATA3 / IL4 / IL4R**
- Mise à jour **synchrone / asynchrone**

De cet état de départ, il existe des gènes *inactivables* (ici en gris).

Étude de cas II : modèle biologique Th_23



Deux boucles d'activation → :

- celle de **Tbet / Th1**
- celle de **GATA3 / Th2**

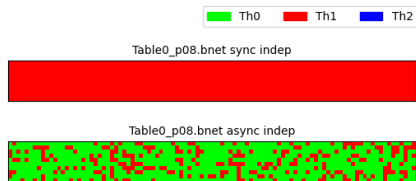
Et une inhibition mutuelle de Tbet ↔ GATA3.

BN initial : seule la boucle **orange** est activable.

Synchrone : déterministe, en sept étapes ⇒ Th1.

Asynchrone : désactivation totale (⇒ Th0)
ou activation de Tbet (⇒ Th1)

Étude de cas II : modèle biologique Th_23



Résultats : $R = 1000$ simulations
0% / 100% / 0% - 74.3% / 25.7% / 0%

Deux boucles d'activation \rightarrow :

- celle de Tbet / Th1
- celle de GATA3 / Th2

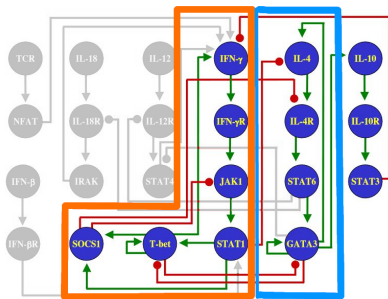
Et une inhibition mutuelle de Tbet \leftrightarrow GATA3.

BN initial : seule la boucle orange est activable.

Synchrone : déterministe, en sept étapes \Rightarrow Th1.

Asynchrone : désactivation totale (\Rightarrow Th0)
ou activation de Tbet (\Rightarrow Th1)

Étude de cas II : modèle biologique Th_23



PBN - Extension IL4 :

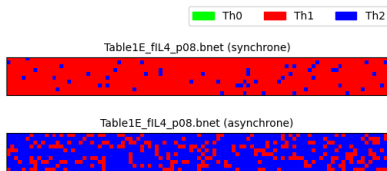
$$\{GATA3 \wedge \neg STAT1\} \rightarrow \{GATA3 \wedge \neg STAT1, GATA3 \vee \neg STAT1\}$$

GATA3 $\vee \neg$ **STAT1** peut activer IL4, donc la boucle **bleue**.
Les deux boucles d'activation fonctionnent en parallèle.

Synchrone : si GATA3 activé en moins de 7 étapes \Rightarrow Th2
sinon \Rightarrow Th1

Asynchrone : si désactivation de la boucle **orange** \Rightarrow Th2
si Tbet activé en 1^{er} \Rightarrow Th1
si GATA3 activé en 1^{er} \Rightarrow Th2

Étude de cas II : modèle biologique Th_23



Résultats : $R = 1000$ simulations
0% / 94.7% / 5.3% - 0% / 24.2% / 75.8%

PBN - Extension IL4 :

$$\{GATA3 \wedge \neg STAT1\} \rightarrow \{GATA3 \wedge \neg STAT1, GATA3 \vee \neg STAT1\}$$

GATA3 $\vee \neg STAT1$ peut activer IL4, donc la boucle bleue.
Les deux boucles d'activation fonctionnent en parallèle.

Synchrone : si GATA3 activé en moins de 7 étapes \Rightarrow Th2
sinon \Rightarrow Th1

Asynchrone : si désactivation de la boucle orange \Rightarrow Th2
si Tbet activé en 1^{er} \Rightarrow Th1
si GATA3 activé en 1^{er} \Rightarrow Th2

Conclusions

Conclusions

- Estimation des attracteurs et de leurs bassins pour les BN
 - Construction et simulation de modèles BN et de PBN
- Modification d'une fonction ou du mode de mise à jour \Rightarrow significatif sur l'atteignabilité des attracteurs

Pistes de continuation

- Distribution empirique pour les PBN
- Autres cas de figure (attracteurs cycliques...)
 - Perturbations p
- Autres générations de fonctions monotones

Merci pour votre attention.

Sources :

- Ilya Shmulevich and Edward R. Dougherty. Probabilistic Boolean Networks - The Modeling and Control of Gene Regulatory Networks. SIAM, 2010. ISBN 978-0-89871-692-4. doi :10.5555/1734075.
- W. Abou-Jaoudé, P. Traynard, P.T. Monteiro, J. Saez-Rodriguez, T. Helikar, D. Thieffry, and C. Chaouiya. Logical Modeling and Dynamical Analysis of Cellular Networks. Frontiers in Genetics, 7(94), 2016. doi :10.3389/fgene.2016.00094.
- José E. R. Cury, Pedro T. Monteiro, and Claudine Chaouiya. Partial Order on the set of Boolean Regulatory Functions. arXiv, 2019. doi :10.48550/arXiv.1901.07623
- Luis Mendoza and Ioannis Xenarios. A method for the generation of standardized qualitative dynamical systems of regulatory networks. Theoretical Biology and Medical Modelling, 3, 2006. doi :10.1186/1742-4682-3-13.
- Seungchan Kim, Edward Dougherty, Yidong Chen, Michael Bittner, and Edward Suh. Can Markov Chain Models Mimic Biological Regulation ? Journal of Biological Systems, 10(4) :337–357, 03 2003. doi :10.1142/S0218339002000676.