



CentraleSupélec



---

Pr. Pascale Le Gall  
Laboratoire de Mathématiques et Informatique pour la Complexité des Systèmes (MICS)  
CentraleSupélec  
9 rue Joliot Curie  
91190 Gif-sur-Yvette  
e-mail: [pascale.legall@centralesupelec.fr](mailto:pascale.legall@centralesupelec.fr)

Gif-sur-Yvette, le 2 mai 2019

Rapport sur le mémoire présenté par  
Xinwei Chai  
pour l'obtention du titre de Docteur de l'Ecole Centrale Nantes  
intitulé  
*Reachability Analysis and Revision of Dynamics of Biological Regularity Networks*

Le document est rédigé en anglais et comprend 4 chapitres principaux, précédés d'un chapitre d'introduction et suivis d'un chapitre de conclusion, d'une bibliographie et de plusieurs annexes. Le travail de Xinwei Chai s'inscrit dans la promotion et l'emploi des outils informatiques associés aux méthodes formelles pour la biologie des systèmes. Dans ce cadre, il aborde deux problématiques complémentaires, à savoir l'étude des propriétés dites d'atteignabilité pour une classe particulière de modélisation des réseaux de régulation et la révision de modèles par apprentissage à partir de données biologiques données sous forme de séries temporelles. Pour chacun de ces deux thèmes, Xinwei Chai a conçu, implémenté et validé les approches discutées dans le manuscrit.

L'introduction (alors 5 pages) positionne le document dans le contexte général de l'exploitation des données biologiques, reconnues pour être à la fois nombreuses, floues et partielles. Une voie qui a déjà fait ses preuves pour la biologie des systèmes est d'analyser les données au regard d'un modèle rigoureux, de nature essentiellement mathématique ou informatique. L'utilisation de tels modèles permet alors d'abstraire le comportement des systèmes biologiques étudiés et de bénéficier des outillages informatiques dédiés à ces modèles. Dans le cadre du manuscrit, Xinwei Chai a privilégié les modèles discrets asynchrones qui ont l'avantage de modéliser de façon réaliste les phénomènes biologiques concurrents, d'être équipés d'outils d'analyse et de faciliter l'inférence des valeurs de leurs paramètres caractéristiques. Xinwei Chai met l'accent sur l'importance d'élaborer des outillages efficaces en pratique, c'est-à-dire assurant un compromis efficacité/pertinence acceptable au regard des attentes des experts en biologie.

Le chapitre 2 (17 pages) est consacré à l'état de l'art. Xinwei Chai introduit le rôle des modèles et leurs principales caractéristiques : continus ou discrets, booléens ou multivalués, probabilistes, logiques, synchrones ou asynchrones, ... Les formalismes des Frappes de

Processus (Process Hitting, PH) et des réseaux d'automates asynchrones (AAN) constituent les deux formalismes qui seront repris dans la suite du document. La dynamique de systèmes peut alors être étudiée à travers les outils de model-checking, pour peu que l'exploration exhaustive du modèle reste envisageable, ou à travers des outils d'analyse statique, dont les résultats obtenus via des approximations et heuristiques peuvent alors être inconclusifs. Parmi les propriétés temporelles d'intérêt, la question de l'atteignabilité de certaines configurations est centrale. Dès que les modèles permettent de faire des prédictions pertinentes sur les systèmes biologiques, la question duale est celle de la construction de dits modèles par apprentissage ou révision en accord avec les connaissances biologiques disponibles. Le chapitre est globalement bien structuré, mais parfois un peu succinct ou insuffisamment illustré pour que le lecteur perçoive précisément les enjeux scientifiques ciblés par Xinwei Chai.

Le chapitre 3 (25 pages) est dédié à l'analyse d'atteignabilité par le biais d'heuristiques. Xinwei Chai commence par introduire le formalisme des réseaux d'automates binaires asynchrones (acronyme ABAN), constitués d'une famille de variables booléennes, pour laquelle une variable booléenne peut changer de valeur s'il existe une transition franchissable qui l'affecte. La dynamique des ABAN est définie par toutes les séquences d'états composites qui peuvent être construites par déclenchement asynchrone des transitions. Le formalisme ABAN est une version réduite des AAN sur laquelle Xinwei Chai étudie les problèmes d'atteignabilité. Pour cela, Xinwei Chai définit une structure statique, appelée graphe simplifié de causalité locale (SLCG), reliant les variables booléennes valuées aux variables valuées présentes dans les gardes des transitions. Intuitivement l'analyse d'atteignabilité sur le graphe SLCG consiste à remonter les arcs de causalité jusqu'à atteindre les conditions initiales. Comme la présence de cycle et d'alternatives dans le graphe SLCG gêne une telle analyse, Xinwei Chai effectue une première étape de préprocessing pour les éliminer et en extraire un arbre. En partie pour éviter qu'un parcours de l'arbre particulier n'occulte des bons ordres de franchissements de transitions, Xinwei propose deux méthodes heuristiques, la première en considérant l'atteignabilité conformément à tous les entrelacements possibles de modification des états globaux, et la seconde via la programmation ASP (Answer Set Programming) qui permet d'encoder directement les arcs du graphe SLCG sous forme de règles ASP. Xinwei Chai établit à leur propos des résultats de terminaison, de correction et d'analyse de complexité. Ainsi, le chapitre contient plusieurs contributions algorithmiques pour améliorer les analyses d'atteignabilité dans le cadre des ABAN. Cependant, une rédaction plus rigoureuse des définitions, théorèmes et algorithmes aiderait à une meilleure appréhension des contributions : en particulier, les algorithmes sont déportés en annexe, et de ce fait, sont insuffisamment commentés. De même, il pourrait être appréciable d'y trouver un exemple biologique crédible et un positionnement plus fin par rapport à l'état de l'art. L'intérêt du chapitre réside dans le cheminement qui a conduit à l'élaboration des heuristiques : des éléments de sémantique des ABAN sous-tendent les arguments avancés pour élaborer les heuristiques d'analyse des structures statiques SLCG. En effet, la plupart des difficultés relevées (présence de cycles, existence de branchements, séquence des modifications d'états) ont directement trait à la sémantique des ABAN. Ne serait-il pas possible de formuler directement des heuristiques tout aussi pertinentes d'exploration du modèle opérationnel des ABAN ?

Le chapitre 4 (33 pages) concerne la mise au point des modèles des réseaux de régulation par inférence ou révision de modèles. L'enjeu est de construire ou modifier des modèles ABAN respectant des connaissances biologiques tenues pour acquises, qui peuvent être des modèles partiels, des séries temporelles, des régulations, ... Xinwei Chai aborde cette problématique en proposant plusieurs processus à mettre en œuvre de façon séquentielle ou complémentaire. Le premier processus est la révision de modèles à partir de régulations

(inhibition ou activation entre deux gènes) candidates et de contraintes d'atteignabilité. Les régulations candidates et les propriétés d'atteignabilité guident l'ajout de transitions tandis que le retrait de transitions est géré à l'aide d'ensembles de coupures, notion empruntée à la littérature. Les algorithmes proposés suivent des approches par sur ou sous approximation. Le second processus a pour objectif l'inférence de régulations candidates à partir de séries temporelles de données. Le second processus s'appuie sur le modèle sous forme d'équations différentielles ordinaires. Xinwei Chai cherche à trouver des corrélations de nature statistique à partir de séries de données temporelles, qui permettent d'émettre des hypothèses sur l'existence de régulations en lien avec des valeurs de seuil fixées. Le troisième processus est une méthode de révision d'un modèle donné sous forme d'un programme logique pour le rendre consistant avec des propriétés d'atteignabilité ou d'inatteignabilité : selon la nature de la propriété, la révision se fait par spécialisation ou généralisation du programme, guidée par les graphes SLCG associés aux propriétés considérées. Xinwei Chai propose ainsi différentes heuristiques pour inférer ou réviser des modèles ABAN en lien avec des connaissances biologiques.

Le chapitre 5 (9 pages) est un chapitre de validation des algorithmes présentés dans les chapitres 3 et 4. Les analyses d'atteignabilité du chapitre 3 sont comparées entre elles et avec d'autres outils, en particulier l'outil PINT dédié aux modèles PH, sur des exemples de la littérature, ainsi que sur des exemples générés aléatoirement. En revanche, les algorithmes du chapitre 4 sont simplement validés sur des exemples aléatoires. Ce chapitre est le bienvenu, pour illustrer et incarner les algorithmes proposés dans les précédents chapitres. En particulier, il permet de constater que les algorithmes d'analyse d'atteignabilité obtiennent des résultats et des performances tout à fait convaincants sur des modèles de grande taille. On peut regretter cependant que les données sont essentiellement analysées de façon quantitative, et qu'il n'y ait ni analyse qualitative fine en particulier au regard des gains affichés pour l'analyse des propriétés d'atteignabilité, ni réflexions méthodologiques associées à une étude de cas biologique menée de bout en bout.

Xinwei Chai conclut le manuscrit (5 pages) en en résumant les principales contributions et listant plusieurs pistes pour les poursuivre. Il a ainsi proposé un formalisme pour les réseaux de régulation, les modèles ABAN, dont les structures statiques dérivées, les graphes SLCG, sont l'élément central pour la définition d'une boîte à outils à destination des experts biologistes. Il a ainsi conçu, implémenté et validé plusieurs algorithmes d'analyse d'atteignabilité et de révision de modèles ABAN. Dans l'objectif d'assurer un compromis acceptable entre performances et correction, la conception de ces algorithmes met en jeu des heuristiques, en général formulées à partir de considérations biologiques. Même si par endroits, le manuscrit manque un peu de rigueur mathématique et d'explications, Xinwei Chai a pris soin d'illustrer les différentes notions et algorithmes introduits à l'aide d'exemples jouets afin d'aider le lecteur à se construire ses intuitions. Ainsi, les contributions de Xinwei Chai se situent principalement dans la variété des méthodes outillées proposées, s'appuyant sur différents concepts ou techniques (graphes, statistiques, programmation logique, ...), pour l'étude des réseaux de régulation. Afin d'en évaluer leur pertinence, la prochaine étape sera de les utiliser sur des études de cas, en dehors des cas d'école, et en lien avec des attentes d'utilisateurs biologistes. En conséquence, je suis favorable à ce que ce document soit présenté en vue de l'obtention du grade de docteur.

