

# Reconstruction exacte de réseaux bayésiens par une approche contraintes

**Thématique :** apprentissage, optimisation combinatoire, réseau bayésien, satisfaction de contraintes pondérées

**Équipe d'accueil :** Statistique et Algorithmique pour la Biologie

**Laboratoire d'accueil :** Mathématiques et Informatique Appliquées de Toulouse, INRA

**Lieu :** Auzeville-Tolosane (près de Toulouse), France

**Encadrant :** Simon de Givry (degivry@toulouse.inra.fr    Tel : 05 61 28 50 74)

**Gratification :** environ 400 euros / mois

## Contexte

Le problème étudié est celui de la reconstruction automatique de la structure d'un réseau bayésien à partir d'observations. On suppose les variables aléatoires entièrement observées pour un ensemble de réalisations et la difficulté est de retrouver la structure qui maximise la vraisemblance pénalisée des observations. Il s'agit d'un problème d'optimisation combinatoire NP-dur avec un espace de recherche constitué d'un nombre exponentiel de graphes dirigés sans circuit. Récemment d'important gains de performance ont été obtenus en limitant le nombre de parents potentiels et en exploitant la taille de l'échantillon pour *pré-évaluer* un ensemble de configurations possibles de parents pour chaque variable aléatoire qui soit de taille raisonnable. Cette technique dite de *cache* a ouvert la voie à plusieurs approches complètes d'optimisation dont la logique propositionnelle (Cussens, 2008), la programmation dynamique (Fan et Yuan, 2015), la programmation linéaire en nombre entiers (Bartlett et Cussens, 2015) et la programmation par contraintes (Hoffmann et Van Beek, 2015). A voir un tutoriel sur le sujet.

L'équipe d'accueil mène des travaux sur la reconstruction de réseaux d'interactions de gènes à l'aide de différentes techniques dont les réseaux bayésiens (Allouche et al, 2014).

## Sujet

L'objectif du stage est d'étudier une nouvelle approche fondée sur le formalisme des problèmes de satisfaction de contraintes pondérées (Cooper *et al.*, 2010). Le travail s'appuiera sur un précédent stage ayant identifié un outil d'optimisation, MaxHS, combinant la logique propositionnelle et la programmation linéaire, relativement efficace pour ce type de problème. Il s'agira de remplacer la composante exploitant la logique propositionnelle par un autre composant utilisant la programmation par contraintes. Une proposition de contrainte globale d'acyclicité sera menée, pouvant s'appuyer sur la relaxation lagrangienne. Par ailleurs, la technique dite de cache donne la possibilité de décomposer le problème en sous-problèmes indépendants que l'on cherchera à exploiter. Les méthodes seront comparées sur des *benchmarks* et sur des données de réseaux de gènes (DREAM5, arabette).

Possibilité de poursuite en thèse avec un financement INRA sur les mêmes thématiques d'apprentissage et d'optimisation avec une application à des données biologiques de grande taille pour l'étude de la résistance au stress chez le tournesol.

## Bibliographie

D. Allouche and C. Cierco and S.de Givry and G. Guillermin and B. and T. Schiex and J. Vandell and M. Vignes. A Panel of Learning Methods for the Reconstruction of Gene Regulatory Networks in a Systems Genetics Context. Gene Network Inference, Springer, 2014.

M. Bartlett, J. Cussens.

Integer Linear Programming for the Bayesian network structure learning problem. In journal of Artificial Intelligence, 2015.

M.C. Cooper, S. de Givry, M. Sanchez, T. Schiex, M. Zytnicky, T. Werner.

Soft Arc-consistency revisited. In journal of Artificial Intelligence, 2010.

J. Cussens.

Bayesian network learning by compiling to weighted MAX-SAT. In Proc. of Uncertainty in Artificial Intelligence, 2008.

X. Fan, C. Yuan.

An Improved Lower Bound for Bayesian Network Structure Learning. In Proc. of the American Association for Artificial Intelligence, 2015.

H. Hoffmann, P. Van Beek.

Machine learning of Bayesian networks using constraint programming. In Proc. of Principles and Practice of Constraint Programming, 2015.