

# Contrainte pondérée d'acyclicité pour l'apprentissage de réseaux bayésiens

**Thématique :** optimisation combinatoire, programmation par contraintes, apprentissage

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

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

**Lieu :** Castanet-Tolosan (près de Toulouse), France

**Encadrants :** Simon de Givry      [simon.de-givry@inra.fr](mailto:simon.de-givry@inra.fr)      Tel : 05 61 28 50 74

George Katsirelos      [georgios.katsirelos@inra.fr](mailto:georgios.katsirelos@inra.fr)      Tel : 05 61 28 52 77

**Gratification :**  $\approx$  550 euros / mois

## Contexte

Le problème étudié est celui de l'apprentissage 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-difficile avec un espace de recherche constitué d'un nombre exponentiel de graphes dirigés sans circuit. Plusieurs approches ont été développées pour tenter de résoudre de façon optimale ce problème dont la logique propositionnelle (Cussens, 2008), la programmation dynamique (Fan et Yuan, 2015), la programmation linéaire en nombre entiers (Bartlett et Cussens, 2015), la programmation par contraintes (Hoffmann et van Beek, 2015) et leur comparaison à des méthodes de recherche locale (Lee et van Beek, 2017). L'équipe d'accueil mène des travaux en optimisation combinatoire dans les sciences du vivant et développe un outil d'optimisation qui a remporté plusieurs compétitions (toulbar2 <http://www.inra.fr/mia/T/toulbar2>). Un projet finalisé de l'équipe concerne l'apprentissage 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). L'écriture d'une contrainte globale pondérée d'acyclicité sera étudiée dans ce cadre. Les travaux menés en programmation par contraintes (<https://www.cosling.com/choco-solver/graph>) et en SAT (Gebser et al, 2014) serviront de point de départ. L'extension au cas pondéré se fera dans le solver toulbar2 à l'instar de (Givry et Katsirelos, 2017). Une comparaison avec les approches existantes sera effectuée sur des benchmarks issus de la communauté apprentissage automatique et à partir de données biologiques réelles obtenues à l'INRA (arabette et tournesol) ou simulées. Possibilité de poursuite en thèse (projet <http://www.lsis.org/demograph>).

## Bibliographie

D. Allouche and C. Cierco and S.de Givry and G. Guillermin and B. and T. Schiex and J. Vandel 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.

M. Gebser, T. Janhunen, and J. Rintanen. SAT modulo graphs: Acyclicity. In *Proc. of European Conference on Logics in Artificial Intelligence*, 2014.

S. de Givry and G. Katsirelos. Clique Cuts in Weighted Constraint Satisfaction. In *Proc. of Principles and Practice of Constraint Programming*, 2017.

H. Hoffmann, P. van Beek. Machine learning of Bayesian networks using constraint programming. In *Proc. of Principles and Practice of Constraint Programming*, 2015.

C. Lee, P. van Beek. An Experimental Analysis of Anytime Algorithms for Bayesian Network Structure Learning. In *Proc. of Machine Learning Research*, Volume 73, 2017.