

## Méthodes à noyaux pour l'inférence de réseaux de régulation génique

### Type of talk

PhD defense rehearsal

### Speaker

Arnaud Fouchet, IBISC, Arob@s Team, Evry, France

### Abstract

De nouvelles technologies, notamment les puces à ADN, multiplient la quantité de données disponibles pour la biologie moléculaire. Dans ce contexte, des méthodes informatiques et mathématiques sont activement développées pour extraire le plus d'information d'un grand nombre de données. En particulier, le problème d'inférence de réseaux de régulation génique a été abordé au moyen de multiples modèles mathématiques et statistiques, des plus basiques (corrélation, modèle booléen ou linéaire) aux plus sophistiqués (arbre de régression, modèles bayésiens avec variables cachées).

La modélisation de niveaux d'expression de gènes soulève de multiples difficultés: non-linéarité, données bruitées, faible nombre de données. Les méthodes à noyaux offrent des qualités pertinentes pour ce problème. En effet, elles sont capables d'identifier des modèles non-linéaires d'une façon robuste au bruit. Pourtant, ces méthodes ont été peu utilisées pour le problème d'inférence de réseaux de régulation génique, notamment car elles étaient difficilement interprétables. Dans cette thèse, nous proposons deux façons d'interpréter des méthodes à noyaux.

Dans un premier temps, d'un point de vue théorique, nous montrons que les méthodes à noyaux permettent d'estimer, à partir d'un ensemble d'apprentissage, une fonction de transition et la moyenne de ses dérivées partielles de façon consistante. Ces estimations de dérivées partielles permettent, sur des exemples réalistes, de mieux identifier le réseau de régulation génique que des méthodes standards.

Dans un deuxième temps, nous développons une méthode à noyaux interprétable grâce à l'apprentissage à noyaux multiples. Ce modèle fournit des résultats du niveau de l'état de l'art sur des réseaux réels et des réseaux simulés réalistes. Par ailleurs, nous montrons que cette méthode à noyaux interprétable peut intégrer des connaissances à priori. Nous montrons que des connaissances à priori raisonnables augmentent substantiellement les performances de l'algorithme.