

---

# Apprentissage statistique de réseaux géniques, de l'inférence au clustering

Magali Champion\*<sup>1</sup>

<sup>1</sup>Mathématiques Appliquées Paris 5 – Institut National des Sciences Mathématiques et de leurs Interactions, Centre National de la Recherche Scientifique, Université Paris Cité – France

## Résumé

Les réseaux biologiques sont des outils puissants pour modéliser sous la forme de graphes des interactions existant entre des entités biologiques et comprendre ainsi des mécanismes biologiques sous-jacents. Les réseaux de régulation de gènes en particulier représentent les régulations (activation ou inhibition) – les arêtes – entre différents gènes – les nœuds – d'un système biologique et permettent par exemple d'identifier le processus pouvant expliquer le développement d'une maladie.

Inférer un réseau génique, c'est utiliser des données biologiques – données d'expression des gènes qui quantifient l'information contenue dans les gènes – pour apprendre les interactions inconnues et reconstruire le graphe associé. Parmi les challenges qui complexifient cette tâche, on trouve la grande dimension des données, la sparsité et la causalité. Nous étudierons les techniques traditionnellement utilisées par la communauté scientifique.

Une fois le réseau génique inféré, se pose la question de l'identification des mécanismes qui peuvent expliquer un phénomène biologique. Dans ce cadre, le clustering de réseaux permet de découvrir des groupes de gènes – clusters – aux caractéristiques communes qui vont agir de manière coordonnée dans une configuration biologique précise. Nous verrons quelles sont les techniques de clustering qui peuvent être utilisées, quelles en sont leurs limites et comment les évaluer.

**Mots-Clés:** apprentissage, inférence, clustering, réseau de gènes

---

\*Intervenant