

# Séminaire de Probabilités et Statistiques

Mardi 13 octobre à 14h

Salle de Conférences

**Vincent Brault**

AgroParisTech

*Deux modèles pour l'étude des données Hi-C.*

Depuis 2009, Lieberman-Aiden et al. (2009) proposent une méthode pour mesurer les interactions entre deux emplacements du chromosome et ainsi tenter d'expliquer le comportement en trois dimensions de ces derniers. A partir de ces données, deux types de structures sont recherchées : la première consiste à retrouver des blocs sur la diagonale afin de trouver les parties du chromosomes qui forment des boucles et la seconde a pour objectif de former un quadrillage pour pouvoir résumer l'information en trois dimensions.

Pour la première démarche, Lévy-Leduc et al. (2014) proposent un modèle de segmentation bidimensionnelle d'une matrice symétrique dont l'objectif est de retrouver les instants de ruptures délimitant ces blocs. Dans cet exposé, nous montrons que la vraisemblance associée est auto-pénalisante ; c'est-à-dire qu'elle est maximale pour le bon nombre de ruptures.

Dans un second temps, nous proposons un modèle pour étudier les interactions entre les différents groupes en segmentant les lignes et les colonnes dans le but de former un quadrillage. Nous montrons que ce problème peut être ramené à un problème de moindre carré avec un vecteur sparse et proposons un algorithme efficace et rapide pour traiter des matrices de taille 10 000x10 000.