



**Laboratoire de Mathématiques et Informatique pour la Complexité et les
Systèmes
MICS**

Présente

L'AVIS DE SOUTENANCE

De Madame Lily MONNIER

Équipe Biomathématiques, Laboratoire MICS, CentraleSupélec, Université Paris-Saclay, soutiendra publiquement ses travaux de thèse de doctorat intitulés :

“Statistical Learning for Single-Cell Data Analysis in Oncology.”

Sous la Direction Monsieur Paul-Henry COURNÈDE

Le Mercredi 5 Mars 2025 à 9h

À CentraleSupélec, dans le **Théâtre Rousseau** (Bâtiment Bouygues).

Membres du jury :

Frédéric CHALMEL, Rapporteur & Examineur, INSERM

Nathalie VIALANEIX, Rapporteur & Examinatrice, INRAE

Elsa BERNARD, Examinatrice, Gustave Roussy

Céline HUDELLOT, Examinatrice, Université Paris-Saclay

Fragkiskos MALLIAROS, Examineur, Université Paris-Saclay

Franck PICARD, Examineur, CNRS – ENS Lyon

Résumé :

Le séquençage de l'ARN à l'échelle d'une unique cellule (single-cell) permet d'étudier les niveaux d'expression des gènes dans chacune des milliers ou des millions de cellules individuelles. Il joue un rôle clé en oncologie pour étudier l'hétérogénéité des tumeurs et identifier les différents types de cellules, même rares. Cependant, l'analyse de ces données soulève plusieurs questions statistiques d'intérêt, en particulier, la correction de l'effet de batch. Nous proposons de résoudre ce problème en exploitant la puissance des méthodes d'apprentissage profond. En effet, la grande quantité de cellules disponibles permet d'entraîner des architectures complexes et d'apprendre des représentations latentes adaptées aux tâches d'intérêt. Notre méthode sera évaluée sur plusieurs cas tests en oncologie et comparée à l'état de l'art diverses tâches telles que le clustering et la détection de gènes différentiellement exprimés. Pour répondre à cette problématique, nous nous sommes heurtés au fardeau computationnel de combiner et équilibrer plusieurs objectifs de diverses natures. Ainsi, nous avons développé une méthode basée sur de l'apprentissage par renforcement, afin d'adapter automatiquement les poids des différents sous-objectifs dans un objectif global. Nous comparerons notre méthode à l'état de l'art sur plusieurs cas tests, comprenant des jeux de données issus du Multi-Task Learning et de la correction d'effet batch. La haute résolution générée par le séquençage single-cell permet l'étude de trajectoires cellulaires, présentes lors des processus de différenciation des cellules sanguines ou de l'évolution clonale d'une tumeur. Pour cela, nous envisageons de mettre à profit les avantages des Graph Neural Networks (GNNs), afin répondre au problème d'inférence de trajectoires en remplaçant les étapes de réduction de dimension et de clustering par des blocks GNNs. Nous évaluerons notre méthode sur un jeu de données simulé et la comparerons à l'état de l'art.

Abstract:

Single-cell RNA sequencing reveals genome-wide gene expression levels within each of thousands to millions of individual cells. It plays a key role in oncology to investigate tumor heterogeneity and identify different cell types, even rare subtypes. However, analyzing these data raises several statistical issues of interest, notably batch-effect correction. We addressed this issue by leveraging the power of deep learning methods. Indeed, the large number of cells available makes it possible to train complex architectures and learn latent representations that are adapted to the tasks of interest. Our method will be implemented in a dedicated analysis pipeline and applied to several test cases of interest in oncology. It will be evaluated through diverse tasks, such as clustering or differential gene expression detection, and compared to the state of the art. While addressing the batch-effect-correction problem, we faced the computational burden of combining and balancing several losses of diverse natures. Thus, we developed a Reinforcement-Learning-based method to automatically weigh the different sub-objectives into a global objective. We will compare our method to the state of the art on several use cases, encompassing Multi-Task Learning and batch-effect-correction scenarios. The fine-grained resolution produced by single-cell sequencing enables the study of cellular trajectories, such as blood differentiation or tumor clonal evolution processes. We harnessed Graph Neural Networks' (GNNs) graph formulation benefits to answer the trajectory inference problem by substituting the classic dimensionality reduction and clustering steps with GNN blocks. Our method will be evaluated on a simulated dataset and compared to the state of the art.