



Offer #2024-07444

Doctorant F/H Graphes dynamiques de co-expression des gènes et inférence de réseaux de régulation

The offer description below is in French

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

Context

Des cellules ayant le même génome peuvent néanmoins exprimer leurs gènes à des niveaux très différents, y compris lorsqu'elles reçoivent les mêmes signaux de leur environnement. Cette variabilité joue un rôle fondamental dans les mécanismes dits de *prise de décision cellulaire* (comme la différenciation de cellules « souches » pluripotentes en cellules « matures » spécialisées) et peut s'expliquer par la conjonction de deux phénomènes biologiques : la stochasticité de l'expression des gènes (présence d'aléa intrinsèque dans les quantités produites d'ARNm et de protéines) et le fait que ces derniers interagissent entre eux via les molécules produites, formant un système dynamique complexe appelé *réseau de régulation*, avec des propriétés émergentes. Les états cellulaires stables sont alors interprétés comme des attracteurs de ce système dynamique, avec une certaine probabilité de passer d'un attracteur à un autre [1].

Dans ce contexte, l'inférence de graphes d'interactions entre les gènes est devenue un cas d'école pour la statistique en grande dimension, notamment depuis l'arrivée puis la massification des données transcriptomiques en cellules uniques (*single-cell*). En particulier, les données de type scRNA-seq permettent de construire de façon classique des graphes non-orientés appelés *réseaux de co-expression*, par exemple en estimant la matrice de covariance entre les gènes. Ces réseaux ont l'intérêt de représenter des données complexes sous forme simplifiée, mais ils ne constituent qu'une information partielle puisqu'ils ne décrivent pas les interactions biologiques à l'origine des motifs de co-expression observés.

Assignment

Ce projet de thèse s'intéresse à la question suivante : *comment passer des réseaux de co-expression (motifs statistiques observés) aux réseaux de régulation (interactions biologiques causales) ?*

Cette question s'avère nettement plus difficile et n'est pas encore résolue de façon satisfaisante à ce jour, notamment à cause de son caractère « mal posé » : les données étant essentiellement observationnelles, il n'est pas possible d'identifier les interactions causales de façon universelle, c'est-à-dire sans faire d'hypothèses sur le système dynamique sous-jacent.

On propose ici de se baser sur un modèle mathématique particulier pour ce système dynamique : il s'agit d'un processus markovien déterministe par morceaux (PDMP) dont la seule source d'aléa est le phénomène biologique de *bursting* transcriptionnel [2,3]. L'intérêt principal de ce modèle est de pouvoir capturer la variabilité des données *single-cell* sans nécessiter l'ajout d'un modèle de bruit *ad hoc*, contrairement aux équations différentielles ordinaires ou même stochastiques souvent utilisées dans ce contexte.

Main activities

Dans un premier temps, on se placera dans le cas d'un jeu de données scRNA-seq constitué de plusieurs *snapshots* (profils transcriptomiques simultanés d'un échantillon de cellules individuelles) avec ou sans structure temporelle. Il s'agira de développer un cadre statistique à la fois bien posé mathématiquement et réaliste d'un point de vue biologique, à partir de résultats obtenus sur le modèle PDMP. L'objectif est d'être capable d'intégrer les différents snapshots dans une phase de pre-processing standardisée, en distinguant la variabilité biologique des différentes sources de variabilité technique [4].

Dans un second temps, on s'intéressera à l'inférence bayésienne exacte de graphes dynamiques de co-expression à partir de données scRNA-seq temporelles, en adaptant un formalisme existant basé sur une famille de modèles graphiques probabilistes [5]. Ces modèles sont construits par mélange de lois, à partir d'une hyper-loi sur des champs aléatoires particuliers (qui vérifient la propriété de Markov sur un arbre, lui-même aléatoire). Un résultat algébrique de type *matrix-tree* permet alors d'interpréter la constante de normalisation comme le déterminant d'une certaine matrice, rendant l'inférence exacte réalisable en temps polynomial, et ce malgré l'explosion du nombre de graphes possibles.

On s'attaquera enfin à l'inférence de réseaux de régulation, en construisant un modèle statistique capable d'intégrer une séquence temporelle de snapshots scRNA-seq. L'objectif fondamental sera de relier de façon quantitative les paramètres du modèle statistique à ceux du modèle dynamique PDMP, tout en assurant l'identifiabilité des interactions grâce à la structure temporelle reliant les snapshots : on pourra partir des résultats obtenus par deux approches existantes, qui n'utilisent que partiellement l'information temporelle [6,7]. Une piste prometteuse consistera à étendre le formalisme bayésien précédent en introduisant un processus markovien (de sauts ou gaussien) à valeurs dans l'espace des matrices laplaciennes de graphes orientés pondérés, puis à estimer la loi a posteriori de ce processus par un algorithme de type EM variationnel [8].

On pourra également s'intéresser à l'adaptation de ce formalisme aux données transcriptomiques spatiales, en plein essor, qui contiennent une information potentiellement très riche (interactions possibles entre les cellules) tout en soulevant des problématiques différentes des données single-cell.

Bibliographie

1. E. Ventre, T. Espinasse, C.-E. Bréhier, V. Calvez, T. Lepoutre, and O. Gandrillon, "Reduction of a stochastic model of gene expression: Lagrangian dynamics gives access to basins of attraction as cell types and metastability," *Journal of Mathematical Biology*, vol. 83, no. 5, p. 59, 2021.
2. U. Herbach, A. Bonnaffoux, T. Espinasse, and O. Gandrillon, "Inferring gene regulatory networks from single-cell data: a mechanistic approach," *BMC Systems Biology*, vol. 11, no. 1, p. 105, 2017.
3. E. Ventre, U. Herbach, T. Espinasse, G. Benoit, and O. Gandrillon, "One model fits all: Combining inference and simulation of gene regulatory networks," *PLOS Computational Biology*, vol. 19, no. 3, p. e1010962, 2023.
4. A. Sarkar and M. Stephens, "Separating measurement and expression models clarifies confusion in single-cell RNA sequencing analysis," *Nature Genetics*, vol. 53, no. 6, pp. 770–777, 2021.
5. M. Meilã and T. Jaakkola, "Tractable Bayesian learning of tree belief networks," *Statistics and Computing*, vol. 16, no. 1, pp. 77–92, 2006.
6. E. Ventre, "Reverse engineering of a mechanistic model of gene expression using metastability and temporal dynamics," *In Silico Biology*, vol. 14, no. 3-4, pp. 89–113, 2021.
7. U. Herbach, "Harissa: stochastic simulation and inference of gene regulatory networks based on transcriptional bursting" in *Lecture Notes in Computer Science*, vol. 14137 of Lecture Notes in Bioinformatics, pp. 97–105, 2023.
8. R. Momal, S. Robin, and C. Ambroise, "Tree-based inference of species interaction networks from abundance data," *Methods in Ecology and Evolution*, vol. 11, no. 5, pp. 621–632, 2020.

Skills

Modélisation statistique : modèles graphiques probabilistes, inférence bayésienne, méthodes variationnelles

Langues : français, anglais scientifique

Benefits package

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Sécurité sociale

Remuneration

2100€ brut/mois la 1ère année

General Information

- **Theme/Domain** : Modeling and Control for Life Sciences
Biologie et santé, Sciences de la vie et de la terre (BAP A)
- **Town/city** : Vandœuvre-lès-Nancy
- **Inria Center** : [Centre Inria de l'Université de Lorraine](#)
- **Starting date** : 2024-10-01
- **Duration of contract** : 3 years
- **Deadline to apply** : 2024-04-28

Contacts

- **Inria Team** : [SIMBA](#)
- **PhD Supervisor** :
Herbach Ulysse / ulysse.herbach@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

Date limite de candidature : 29 avril 2024

How to apply

Upload your file on jobs.inria.fr in a single pdf or zip file. Your file should contain the following documents:

- Your CV
- A cover/motivation letter describing your interest in this topic
- A short (max one page) description of your Master thesis (or equivalent) or of the work in progress if not yet completed
- Your degree certificates and transcripts for Bachelor and Master (or the last 5 years)
- Master thesis (or equivalent) if it is already completed and publications if any (it is not expected that you have any). Only the web links to these documents are preferable, if possible.

In addition, one recommendation letter from the person who supervises(d) your Master thesis (or research project or internship) should be sent directly by his/her author to ulyse.herbach@inria.fr.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security :

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :

As part of its diversity policy, all Inria positions are accessible to people with disabilities.