

Apprentissage de dynamiques multiéchelles par réseaux d neurones profonds pour l'étude du développement embryonnaire

Stage de master 2 Recherche

Encadrants

- Thierry Artières
QARMA@LIS, ECM
<https://pageperso.lis-lab.fr/thierry.artieres/thierry.artieres@lis-lab.fr>
- Paul Villoutreix
QARMA@LIS et Turing Center for Living Systems
<https://paulcinq.wordpress.com/paul.villoutreix@lis-lab.fr>

Description du projet (1 à 2 pages max) :

Contexte scientifique:

Key words: Geometric deep learning, evolving multigraph, developmental biology

Les organismes multicellulaires se développent à partir d'un ovule fécondé. La séquence d'événements conduisant au positionnement précis des cellules individuelles avec l'identité requise est orchestrée par la régulation des gènes à l'intérieur et entre les cellules, la prolifération et les réarrangements cellulaires ainsi que les changements morphologiques globaux. Même si plusieurs composantes de ces processus ont été élucidées, il manque encore un cadre générique qui permettrait de saisir les interactions entre eux. Nous proposons de décrire un organisme en développement comme un graphe évolutif où chaque cellule correspond à un nœud, les relations spatiales sont encodées par les arrêtes et l'état transcriptomique est un label sur les nœuds (vecteur de dimension 20000, correspondant à chacun des gènes étudiés). Nous explorerons comment l'action de la prolifération cellulaire, le réarrangement spatial et la dynamique de différenciation affectent les caractéristiques de ces graphes en utilisant des données publiées du développement de l'organisme modèle *C. Elegans* (Li, X., et al., 2019 & Packer J. et al., 2019).

Sujet

L'objet de l'étude vise de manière générale à comprendre les règles locales d'évolution de graphes représentant des organismes en développement. D'un point de vue méthodologique, des approches récentes ont généralisé l'apprentissage profond à toutes sortes de domaines comme les graphes et sont des modélisation

pertinentes pour l'étude (Kipf et al., 2017). Les graphes permettent en effet de fournir un modèle pour des espaces non-euclidien (comme la forme d'un embryon, ou bien la distribution des cellules dans l'espace transcriptionnel). Une des idées envisagées ici est de modéliser la dynamique du développement d'organismes à l'aide de modèles de type Graph Convolutional Networks pour pouvoir interpréter après apprentissage les convolutions apprises comme des règles locales d'évolution. Cela pose plusieurs problèmes en particulier liés au nombre de données d'apprentissage et à la nature des convolutions envisagées. Des solutions à ces deux problèmes seront envisagées via des connaissances a priori sur les données considérées.

Ce projet couvre donc une dimension théorique, établir un cadre théorique pertinent, et une dimension computationnelle et appliquée, développer les implémentations nécessaires à l'analyse de données biologiques issues de la littérature. En particulier nous chercherons à rendre compte de la variabilité transitoire et de son impact sur les modèles d'apprentissage.

Références bibliographiques

Li, X., Zhao, Z., Xu, W., Fan, R., Xiao, L., Ma, X., & Du, Z. Systems Properties and Spatiotemporal Regulation of Cell Position Variability during Embryogenesis. *Cell reports*, 2019.

Packer, J.S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan, K., Trapnell, C., Kim, J. and Waterston, R.H., 2019. A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single cell resolution. *Science*, 2019.

T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations, 2017.