Faculté des Sciences
Aix*Marseille Université

Send to : aitor dot gonzalez at univ-amu dot fr

Internship proposal - Master 2 Bioinformatics

Specialisation Software Development and Data Analysis

Internship period: Jan-Jun 2024

Internship title	**Drug Repurposing for Rare Diseases using Machine Learning on a Biomedical Knowledge Graph**
Internship supervisor(s)	Anaïs Baudot, Galadriel Brière
Laboratory acronym(s)	MMG, Marseille Medical Genetics
Laboratory city(ies)	Marseille
Web site(s)	https://www.marseille-medical-genetics.org/a-baudot/
Team(s)	Systems Biomedicine The Sytems Biomedicine team is an interdisciplinary team of computational biologists, bioinformaticians and systems biologists. The main goals of the team are to develop computational approaches to better understand human diseases, in particular rare diseases.
E-mail(s)	anais.baudot@univ-amu.fr ; marie-galadriel.briere@univ-amu.fr
Internship description	Drug repurposing emerged as a promising strategy in the field of rare diseases research. Instead of developing new drugs from scratch, this approach focuses on investigating existing drugs that have already been approved for other medical conditions, with the goal of finding new treatments for rare diseases. To address the complexity of biological systems and leverage the vast knowledge accumulated by the scientific community, Knowledge Graphs (KG) are frequently used to represent and store the intricate interactions among various biological entities, such as genes/proteins, metabolic pathways, biological processes, diseases, and drugs. Notably, KG embedding techniques have proven to be valuable for integrating such diverse and complex information, by representing the entities of a KG into low dimensional vector representations. This transformation captures semantic relationships and similarities between different entities, enabling more efficient processing, analysis, and comparison of complex information. In the field of drug repurposing, KG embedding techniques can leverage the complex interactions gathered in a Biomedical KG and propose new treatments for a disease based on the similarity of the low dimensional representations of drug and disease nodes. In this project, we aim to develop a drug repurposing strategy dedicated to rare diseases, which are particularly challenging for drug repurposing,

	as these conditions, often caused by a single genetic mutation, can manifest in a multitude of ways, affecting different tissues and and biological processes in the body. Our hypothesis is that KG alignment techniques, which aim to map embeddings from multiple KGs into a shared latent space, can effectively harness the diverse range of phenotypes associated with rare diseases during the drug repurposing process. Specifically, we aim to embed the nodes from a KG centered around drugs and a KG focused on diseases and map them into a common latent space, utilizing genes (common to both KGs) and known drug-disease associations as reference points (anchors). By selecting relevant anchors for rare diseases, such as causal genes or disease-associated phenotypes, we can guide the drug repurposing effort towards more promising candidates for rare diseases.
Expected skills	• Programming skills in Python or R are necessary. • Fluency in English is required. • Familiarity with network science and/or omics data analysis is a plus. • Soft skills: Communication, Critical Thinking and Problem Solving skills are a plus.
Confidential (yes/no)	No
Informatics resources	IFB clusters, Mesocentre
Potential PhD project (yes/no)	To be discussed
References (2 and 3)	• Pio-Lopez, L., Valdeolivas, A., Tichit, L., Remy, É., & Baudot, A. (2021). MultiVERSE: a multiplex and multiplex-heterogeneous network embedding approach. *Scientific reports*, *11*(1), 8794. • BAUMGARTNER, Matthias, DELL'AGLIO, Daniele, PAULHEIM, Heiko, *et al.* Towards the Web of Embeddings: Integrating multiple knowledge graph embedding spaces with FedCoder. *Journal of Web Semantics*, 2023, vol. 75, p. 100741. • Térézol, M., Baudot, A., & Ozisik, O. (2023). ODAMNet: a Python package to identify molecular relationships between chemicals and rare diseases using overlap, active module and random walk approaches. *bioRxiv*, 2023-07.