Graph-based label propagation methods for gene function prediction

Context: A better understanding of gene-phenotype relationships requires an integration of biological information of various kinds. However, this information is often dispersed in several databases on the Internet, having heterogeneous way of access. For biologists, it is difficult to search these data as the mass of information is hard to manage.

Objective: The current challenges are related to the development of methods for functional analysis of genes and in particular to methods for prioritization of candidate genes. The data integrated from databases are incomplete, heterogeneous, insufficient to infer with certainty the function of genes.
One of the first objectives will be the development data integration methods to extract functional information on genes in scientific documents. A second objective will be to develop Graph-based methods to propagate gene functions across the graph and predict functions to unlabeled genes. Finally, the methods will be evaluated and validated on published data.

Program:
- development of data integration methods based on a corpus of scientific datasets and documents identified by the partners.
- Development of Graph-based label propagation methods
- Validation of the methods through published use cases (Arabidopsis and rice data) in international journals.

Prerequisites:
- Python programming (numpy, pandas, scikit-learn, nltk, gensim, networkX, etc...)
- Knowledge in Graph theory
-

References:
- Li H, Zhang R, Zhao Z, Liu X. LPA-MNI: An Improved Label Propagation Algorithm Based on Modularity and Node Importance for Community Detection. Entropy (Basel). 2021 Apr 21;23(5):497. doi: 10.3390/e23050497. PMID: 33919470; PMCID: PMC8143565.
- Liu M, Yang J, Guo J, Chen J, Zhang Y. An improved two-stage label propagation algorithm based on LeaderRank. PeerJ Comput Sci. 2022 May 18;8:e981. doi: 10.7717/peerj-cs.981. PMID: 36091993; PMCID: PMC9454888.

Application: Applications for this position (CV, Motivation Letter, last grade report, References) will be received EXCLUSIVELY in a single PDF document accessible for download via email sent to Jérome AZE (jerome.aze_at_irmm.fr) and Pierre LARMANDE (pierre.larmande_at_ird.fr).

Supervision:
- UMR 5506 LIRMM Equipe ADVANSE : AZE Jerome
- UMR 232 DIADE Equipe CERES: LARMANDE Pierre

Location : LIRMM et IRD-Occitanie
Gratification : 6 months