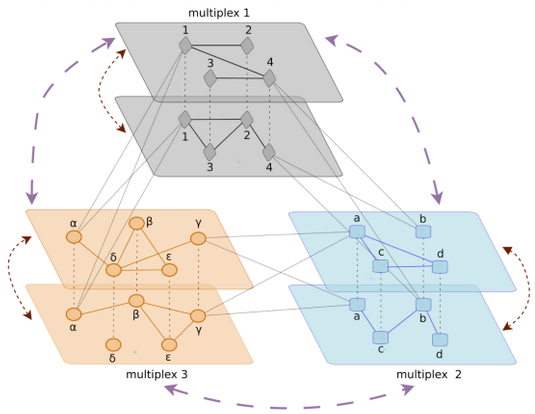
Random Walk with Restart on Multilayer Networks:
From Node Prioritization to Supervised Link Prediction, and Beyond

Anthony Baptista[1,2,3], Galadriel Brière[4], Anaïs Baudot[1]

1 - Aix Marseille Univ, INSERM, MMG, 13385, Marseille, France ; 2 - The Alan Turing Institute, The British Library, London, NW1 2DB, United Kingdom ;
3 - School of Mathematical Sciences, Queen Mary University of London, London, E1 4NS, United Kingdom ; 4 - Aix Marseille Univ, CNRS, I2M, Marseille, France

RWR on multilayer networks with MultiXrank

Universal multilayer networks provide a natural way to integrate diverse and multi-scale data sources into a common framework.



MultiXrank is a Random Walk with Restart (RWR) algorithm able to explore universal multilayer networks [1].

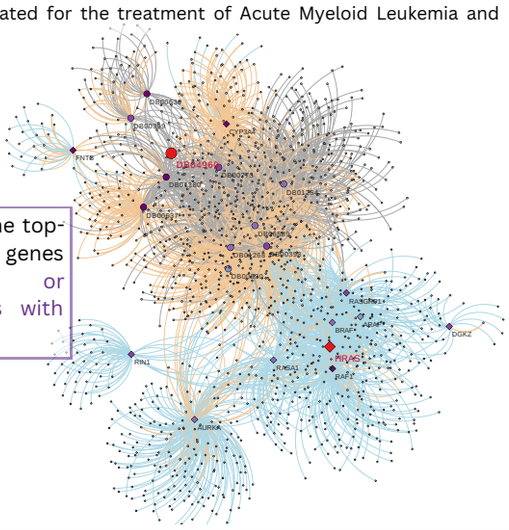
Starting from one or more seed nodes, the random walker navigates the different network layers and generates scores that reflect node's relevance with respect to the seed(s). These scores can then be used in a wide variety of downstream analyses. We aim here to highlight these versatile usages in various bioinformatics tasks.

Node Prioritization in Leukemia

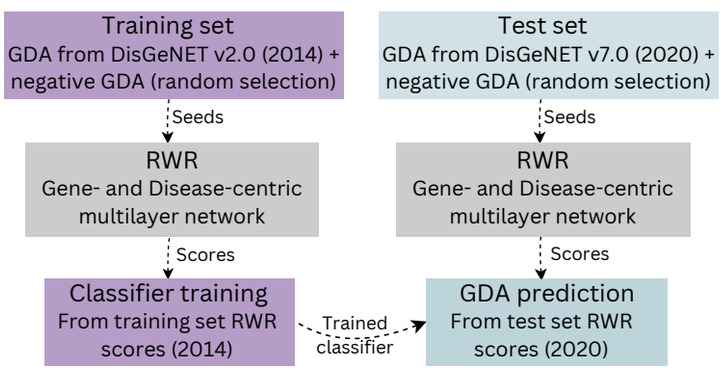
We applied MultiXrank on a multilayer network composed of a **gene multiplex**, a **drug multiplex** and their **bipartite interactions**, starting from two seeds nodes (gene HRAS, drug Tipifarnib DB04960), to prioritize genes and drugs of interest in Leukemia.

HRAS is recognized as an oncogene that plays a role in the development of specific types of leukemia [2].
Tipifarnib is a drug investigated for the treatment of Acute Myeloid Leukemia and other types of cancer [3].

A literature review of the top-10 prioritized drugs and genes establishes known or suspected connections with leukemia.



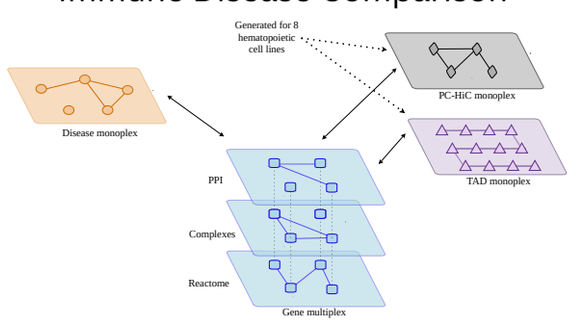
Supervised Prediction of Gene-Disease Associations (GDA)



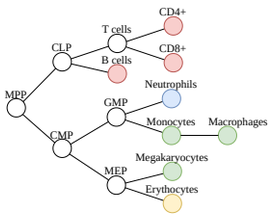
Performance of the model to predict GDA from DisGeNET v2.0 (2014) and DisGeNET v7.0 (2020):

	Model	Accuracy	F1 score	True Positives	False Positives	True Negatives	False Negatives	# Positives	# Negatives
Training set	XGB	0.85	0.79	508	197	948	70	578	1145
Test set	XGB	0.64	0.61	4038	2023	5195	3180	7218	7218

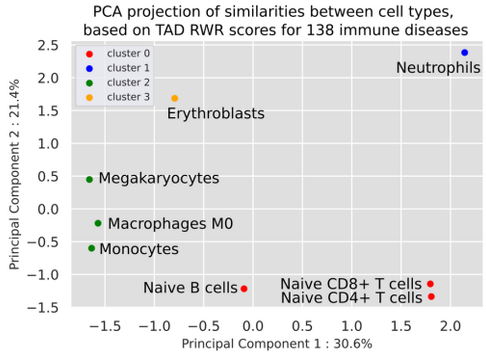
Diffusion Profiles in Hematopoietic Cells for Immune Disease Comparison



We selected 138 immune diseases, iteratively used as seeds for the exploration of each hematopoietic cell-type specific multilayer network. Then, we integrated the RWR scores obtained for each disease on each network, and for each node type.
Preliminary results show that RWR scores for PC-HiC and TAD nodes capture the tree lineage of hematopoietic cells.



MPP: Multi-Potent Progenitor
CLP: Common Lymphoid Progenitor
CMP: Common Myeloid Progenitor
MEP: Megakaryocyte/Erythrocyte Progenitor
GMP: Granulocyte/Macrophage Progenitor



Ongoing work: **Comparing immune disease diffusion profiles to identify comorbid immune diseases.**

[1] Baptista, A., Gonzalez, A., & Baudot, A. (2022). Universal multilayer network exploration by random walk with restart. Communications Physics, 5(1), 170.
[2] Parikh, C., Subrahmanyam, R., & Ren, R. (2007). Oncogenic NRAS, KRAS, and HRAS exhibit different leukemogenic potentials in mice. Cancer research, 67(15), 7139-7146.
[3] Luger, S., et al (2015). Tipifarnib as maintenance therapy in acute myeloid leukemia (AML) improves survival in a subgroup of patients with high risk disease. Results of the phase III intergroup trial E2902. Blood, 126(23), 1308.