

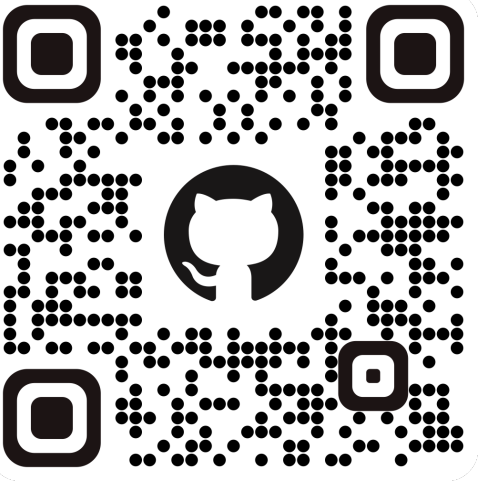
**Analysis of molecular interactions and pathways to assess the similarity of biomarkers:
a case study on autism spectrum disorders.**

Valentina Nale, Alice Chiodi, Matteo Gnocchi, Marco Moscatelli, Alessandra Mezzelani, Ettore Mosca
Institute of Biomedical Technologies, National Research Council, Segrate (Milan), Italy



<https://www.gemma-project.eu>

MOTIVATION: the knowledge of molecular interaction network can be used to get rid of the possible heterogeneity of biomarkers found by different studies, provided that the different biomarkers are in network proximity and hence associated with the same biological functions



Two lists of biomarkers (L1, L2)

The diagram illustrates the structure of the input data for the model. It shows two sets of input features, List1 and List2, each with 10 columns. List1 columns are labeled g1, g1, ::, gN, g1, ::, ::, and gM. List2 columns are labeled g1, g1, ::, gN, g1, ::, ::, and gM. The input is a matrix of 0s and 1s. The output is a matrix of 0s and 1s, with the same structure as the input. The input matrix is shown with a green border, and the output matrix is shown with a red border.

X0 matrix is built by mapping the two lists of biomarkers on the interactome

Xs is a matrix of measures of the network proximity of two biomarkers on the interactome

X0

Xs

$$N_{i,j}$$

Network proximity of two biomarkers based on network diffusion (ND)

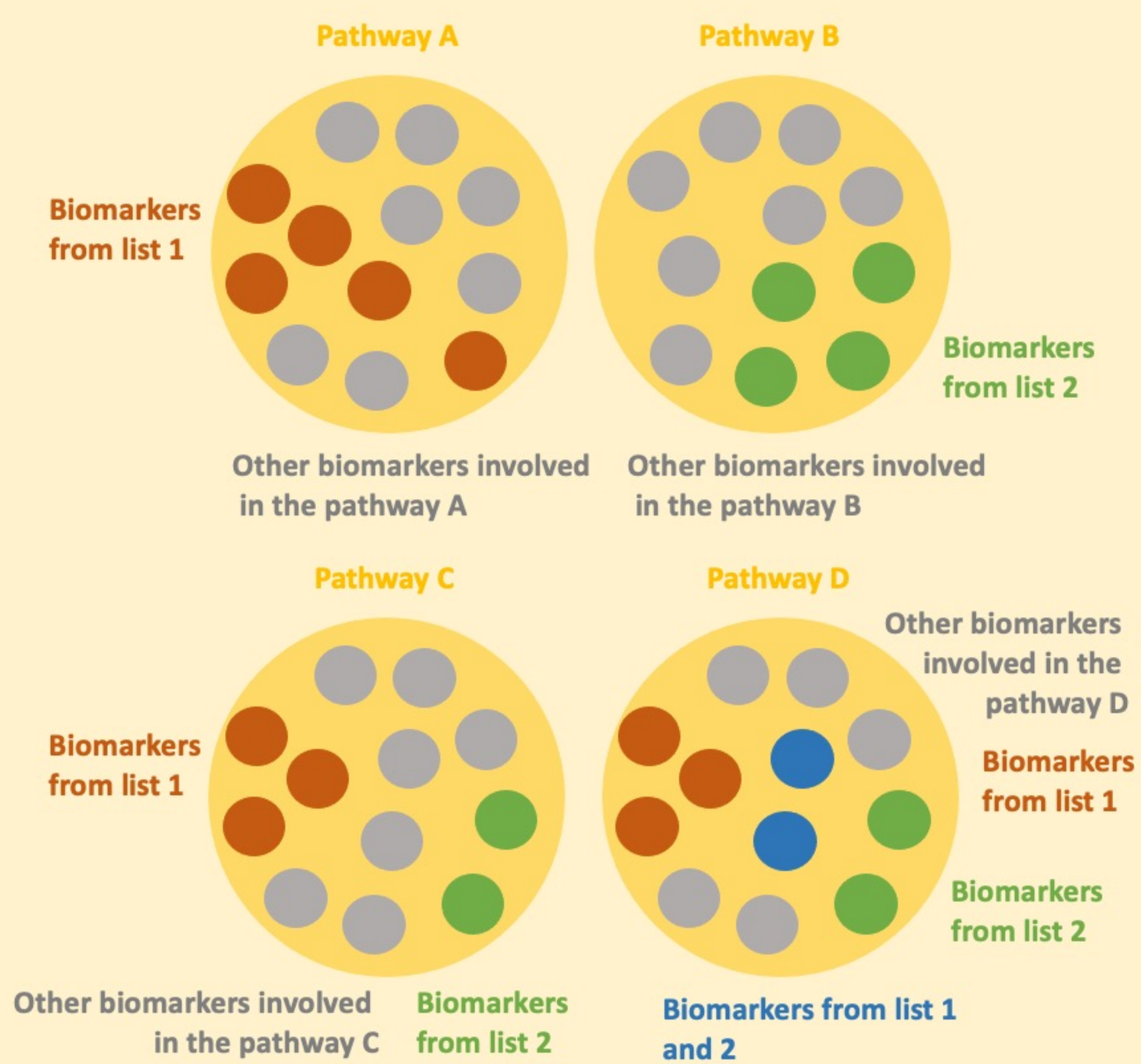
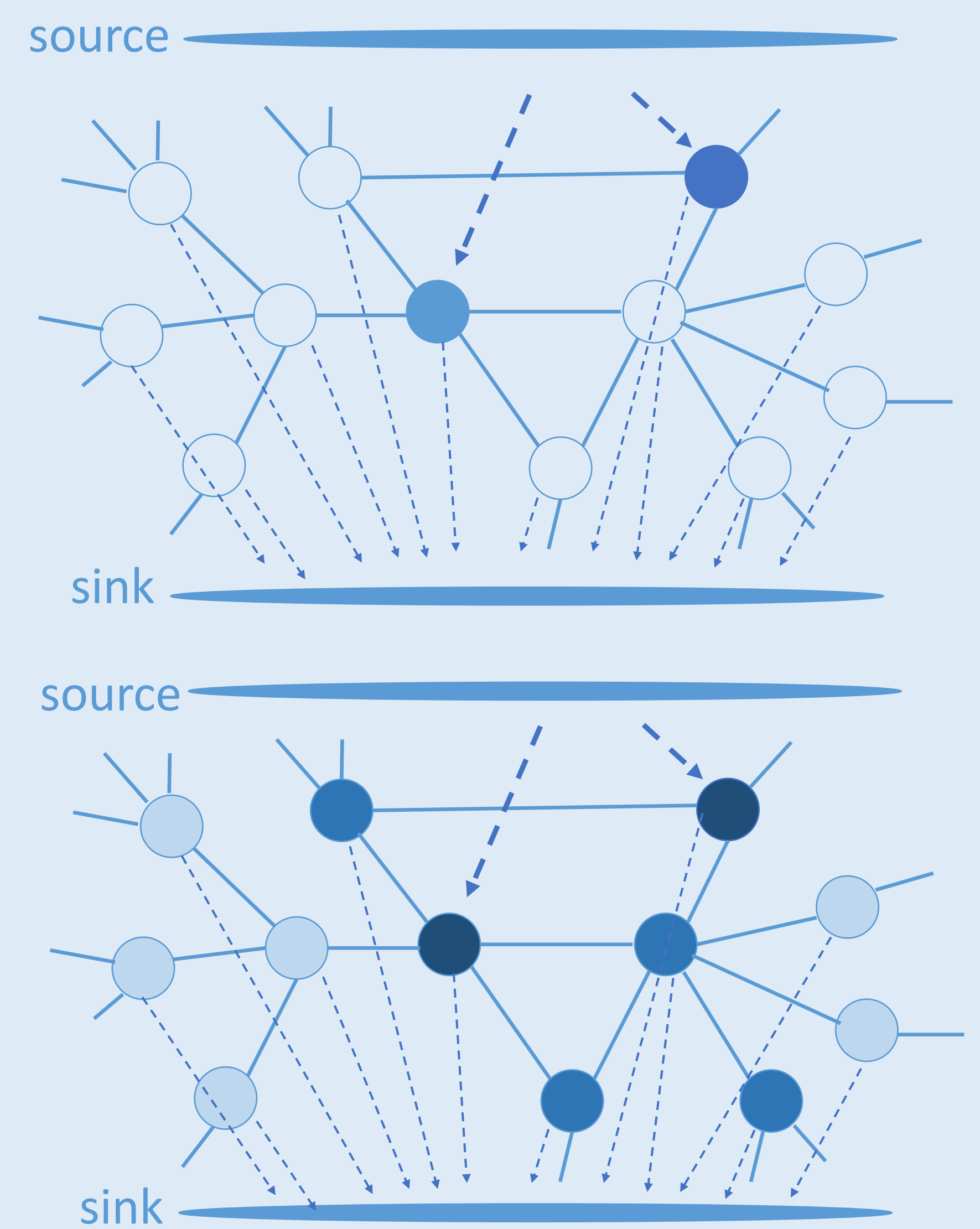
Genes from list 1

Genes from list 2

Genes from list 1

X

Network diffusion: it defines the network proximity of each vertex to the «source» vertices

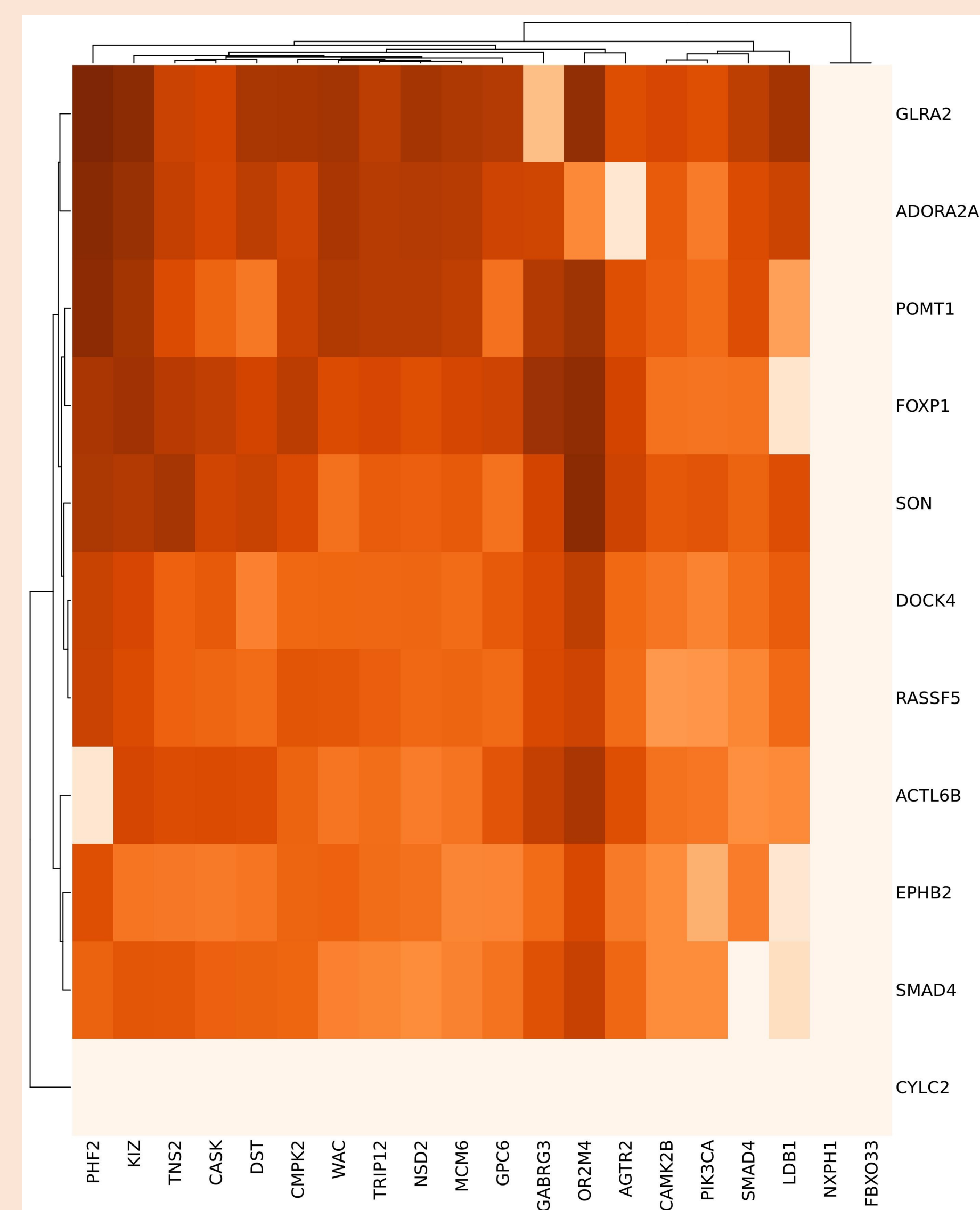


$P_{i,j}$ = Semantic similarity e.g., Jaccard Index

$$P_{i,j} = J(Ptw_i, Ptw_j) = \frac{|Ptw_i \cap Ptw_j|}{|Ptw_i \cup Ptw_j|}$$

where P_{tw_i} and P_{tw_j} are the list of **pathways** in which P_i and P_j occurs respectively

Functional similarity $S_{i,j} = N_{i,j} + P_{i,j}$



Biomarkers for this case study are sampled from SFARI Gene database. The interactome on which genes are mapped is STRING v11.5 with confidence score > 0.7 and the top 3 neighbour with confidence score > 0.4, without text mining.

The heatmap shows the level of functional similarity between pairs of biomarkers based on distance on the network (ND) and semantic similarity.