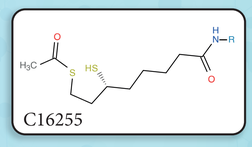
P29

Comparison of neural network-based methods for similarity prediction in compounds with unknown structure

Eugenio Borzone, Leandro Di Persia, Matías Gerard
Research Institute for Signals, Systems and Computational Intelligence -UNL -Conicet

Background

Similarity between compounds is widely used in chemoinformatics.



Usually It is calculated using structural information of the compounds, so it is only available for compounds with known structure.

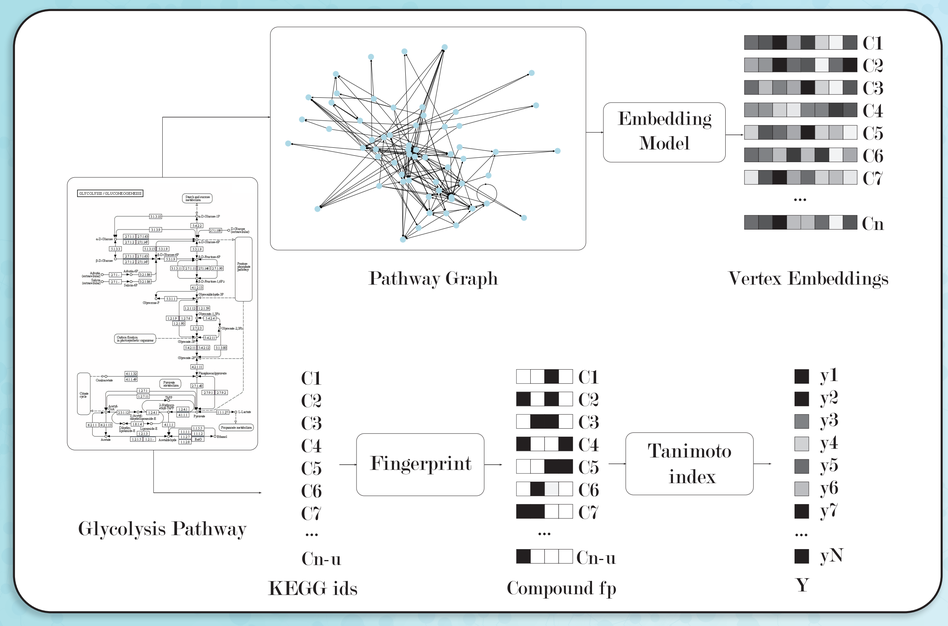
This compound contains a generic substituent R, which means that there is another bonded structure at that point.

Unknown Structure
C00138

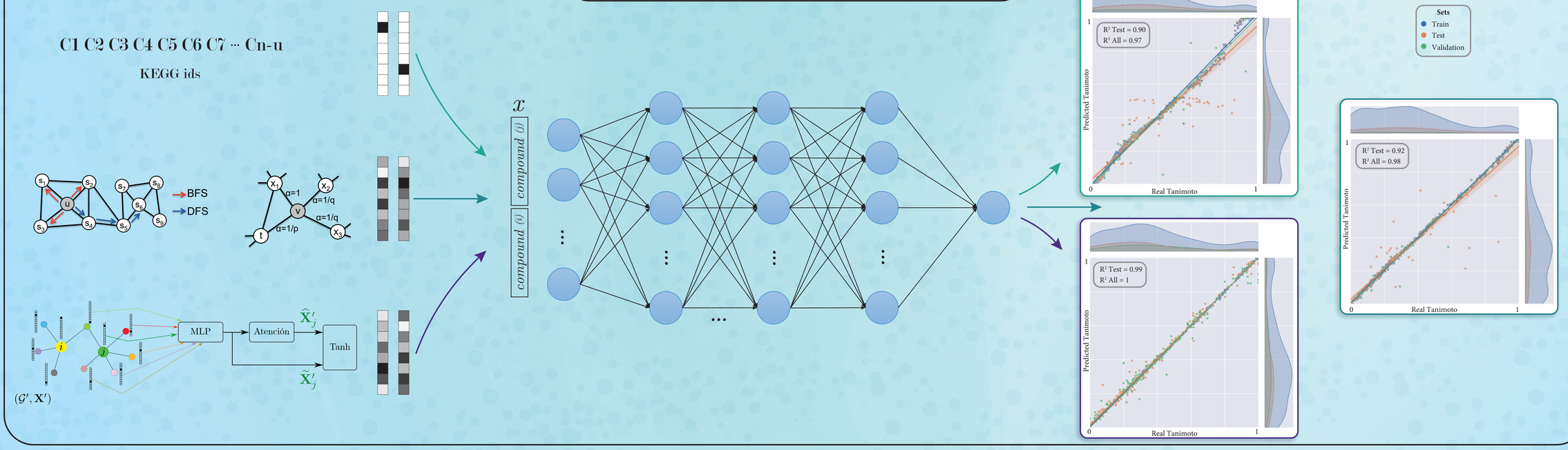
To address this constraint, we use the information of the metabolic pathways topology, in order to infer similarity between compounds with unknown structure.

In this case, the compound has no known structure.

Dataset

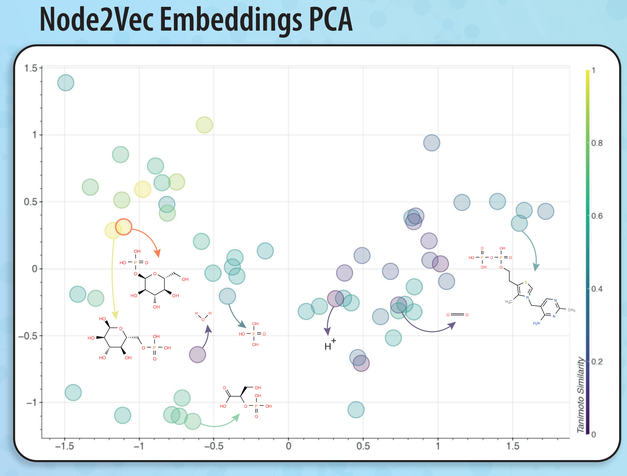


Models and Results



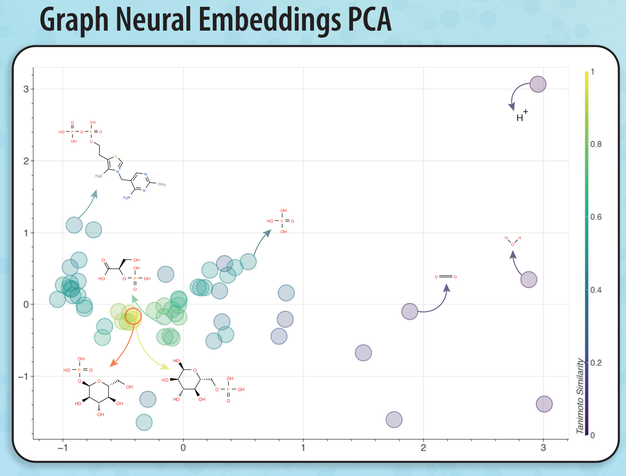
Results Analysis

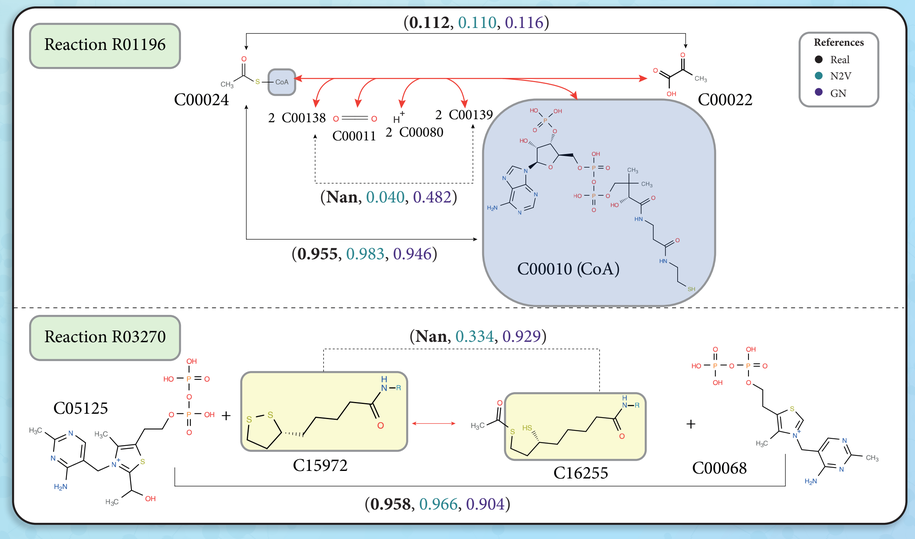
Node2Vec Embeddings PCA



Colors represent relative similarity to reference compound C00103 (Glucose-1-p) highlighted with red. It can be seen that the embedding space is capturing the similarity, more similar compounds appear closer together.

Graph Neural Embeddings PCA





sinc(i) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
E. Borzone, L. Di Persia & M. Gerard, "Comparison of neural network-based methods for similarity prediction in compounds with unknown structure"
XII Argentine Congress of Bioinformatics and Computational Biology (ARCABIO), 2022