Using Non-Negative Matrix Tri-Factorization to find candidate drugs for the treatment of Alternating Hemiplegia of Childhood (AHC)

Ceddia G (1)† , Masseroli M (1), Piccardi C (1), Ceri S (1)

(1) Department of Electronics, Computer Science and Bioengineering, Politecnico di Milano, Milano, Italy

† Email: gaia.ceddia@polimi.it

Motivation

Alternating Hemiplegia of Childhood (AHC) is a rare neurological condition characterized by frequent episodes of temporary paralysis that often affect one side of the body and in some cases both sides at once. A recent study established mutations in the ATP1A3 gene, which codes for subunits of integral membrane proteins, as the crucial cause of AHC [1]: the dysregulated excitability due to membrane proteins malfunctioning of neurons causes the pathology [1]. The challenge is to integrate heterogeneous datasets such as drug-target interactions and target-pathway associations to reveal new candidate drugs for this rare condition. To achieve this, we collected lists of drugs "already in use" for AHC (or reference drugs), possible helpful drugs according to medical doctors' expertise target proteins of both drug classes, and their associated molecular pathways in order to analyze the complete framework. Then, we applied a generalizable network-based clustering approach to the constructed network, postulating that drugs in the same cluster as the reference drugs can be considered as candidate novel drugs for the treatment of AHC.

Methods

We used Non-negative Matrix Tri-Factorization (NMTF) as a well-established co-clustering technique in machine learning to cope with the heterogeneity of the network [2]. NMTF is a well-developed method in data integration because of its capacity to factorize any relation matrix between interconnected data types [2]. We considered three different datasets: drugs, proteins and pathways provided by Dompé Pharmaceutics. We encoded the relationships between drugs and proteins in a signed binary high-dimensional relation matrix, taking into account the activation relationships as positive connections and the inhibition relationships as negative connections. Proteins and pathways were related to each other according to a binary high-dimensional relation matrix. Since NMTF is defined for positively weighted graphs, we decided to split the drugs/proteins signed relation matrix into two matrices (respectively including activation and inhibition relationships) and to apply NMTF to independently decompose each of the relation matrices into a product of three non-negative low-dimensional matrices. The entries of such low-dimensional matrices define the cluster assignation for drugs, proteins and pathways, respectively. In particular, the drug-cluster matrix, with rows representing drugs and columns representing clusters, is used to place drug d into cluster k,

when k corresponds to the entry with the largest value in row d.

Results

Figure 1A shows the results obtained considering the activation relation matrix (positive drug/target interactions), for k=5 drug clusters (k is selected as the value for which the clustering is most stable according to the dispersion coefficient [2]). Figure 1B shows results, where each drug is within a single cluster and reference drugs are highlighted in red. Interestingly, some of the reference drugs are contained in clusters 2 and 5 for positive interactions, and cluster 2 for negative interactions, which contain drugs whose targets are highly related to ion channel transport pathways; therefore, drugs listed in the same clusters as the reference drugs can be considered candidate drugs for the pathology, including Ifenprodil and Ketamine, which therefore are examples of possible novel drug treatment for AHC.

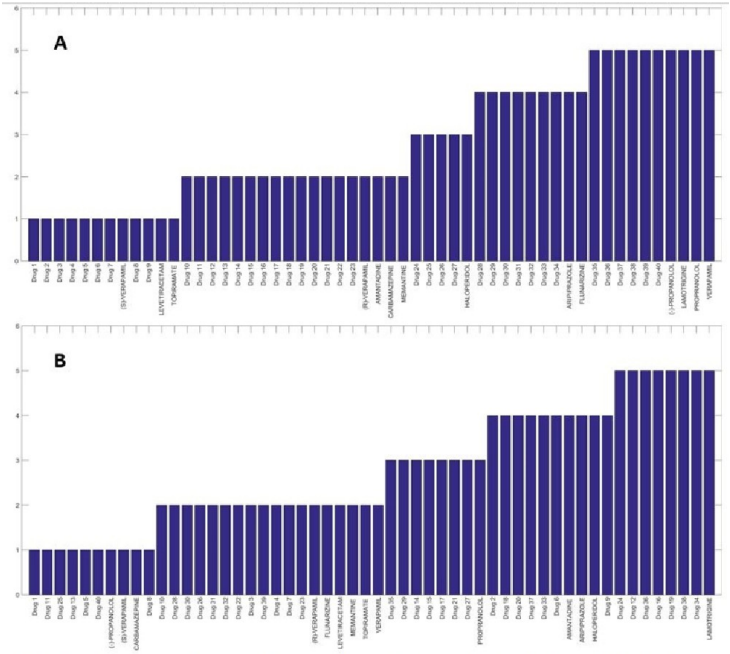

Fig.1. Drugs clustered based on their activation relation matrix (A) and inhibition relation matrix (B). X-axis represents drugs in the network and y-axis represents the cluster number