Network-based Approach To Identify New Indications For Existing Drugs In The Drugbank Database

Ceddia C[1], Pinoli P[1], Masseroli M[1]

[1]*Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano*

[§] Email:gaia.ceddia@polimi.it

Motivation

Drug repurposing, or drug repositioning, has the aim of finding new indications for already approved drugs. Compared to traditional drug discovery, it is a much less expensive process, and it does not require long and dangerous clinical trials [1]. Thus, computational approaches to relocate known chemical molecules have been increasingly used over the years. The traditional drug development attempts to design chemical compounds for a specific target, i.e. inhibiting or activating a certain protein, while drug repositioning considers the overall effect of the drug [1]. For the purpose of drug repositioning, the integration of several information, such as drug-target multiple associations and target functionality, has a key role. We propose a network-based method to infer new indications for existing drugs by adding data from heterogeneous databases to build a multi-layer network.

Methods

We built a quadripartite network composed of drugs' categories, drugs, proteins and biological pathways as nodes, and associations between such nodes as network edges. We retrieved the data about categories-drugs, drugs-proteins and proteins-pathways connections from DrugBank [2], UniProt [3] and Reactome [4] databases, respectively. Adjacency matrices of the network were factorized with the Non-negative Matrix Tri-Factorization (NMTF) method [5]; thus, the result of the factorizations depends on the entire network. We then reconstructed the adjacency matrix associating categories to drugs by using the outputs of the NMTF, i.e. the first three factor matrices, in order to predict missing categories-drugs links and, therefore, novel potential drug categorizations (drug repositioning). The reconstructed adjacency matrix has only positive elements; two nodes are considered to be connected if the corresponding element of the matrix has value greater than a threshold.

Results

Figure 1A shows the complete architecture of the constructed network; it includes 10 categories, 1,120 drugs, 1,012 proteins and 1,563 pathways. In order to test our method, we randomly removed 10% of the category-drug links (Figure 1A). The NMTF reconstruction of the category-drug relation matrix allows to restore categories-drugs links according to the overall quadripartite architecture of the network. We applied different thresholds on the elements of the reconstructed categories-drugs relation matrix to compute precision and recall measures. The precision-recall curve (Figure 1B) shows that for a threshold 0.5 the precision and recall reach 1 and 0.81 respectively. For 0.6, the precision is still high (equal to 1), but the recall decreases significantly, i.e., predicted links are always true links
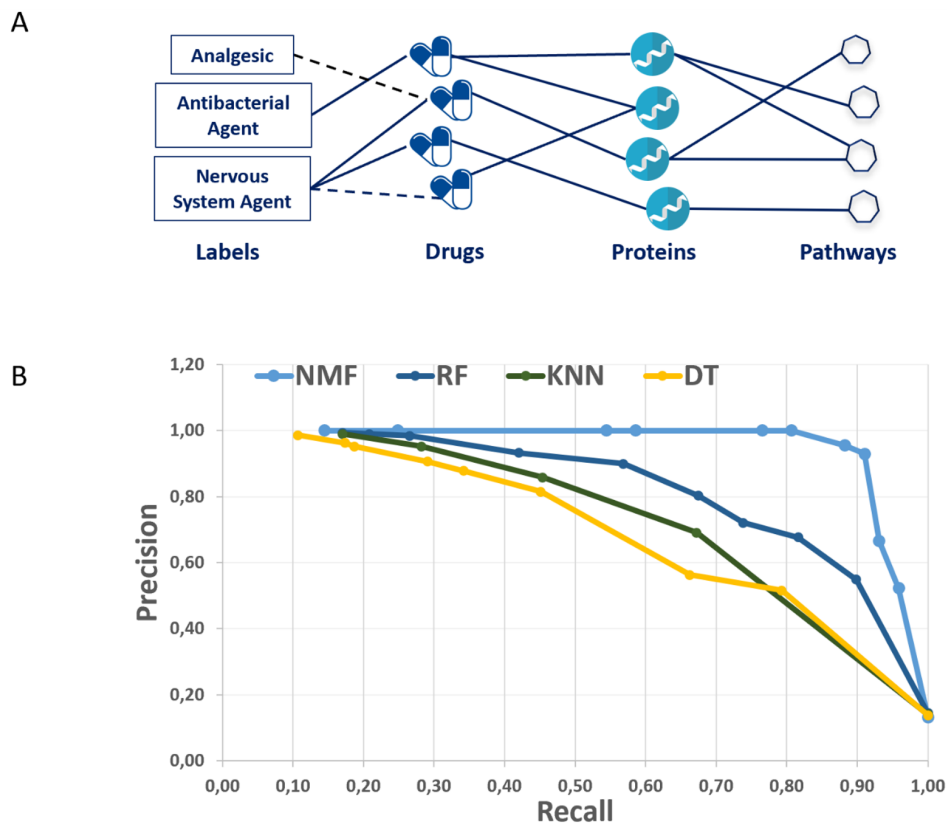
Figure 1. A: Schematic architecture of the constructed network. B: Precision-Recall curves with the NMTF, Random Forest, K-Nearest Neighbors and Decision Tree methods, applied on the partial network, i.e., only continuous links in Figure 1A were considered in the training phase.

for 0.5, but their number decreases over the total number of true links when the threshold tends to 1. For threshld values lower than 0.5, precision and recall lean towards their lowest and highest value, respectively. Figure 1B shows also that our method outperforms other traditional classifiers, such as Random Forest, Decision Tree and K-Nearest Neighbors classifiers. Furthermore, our approach was able to find new drug indications not considered in the DrugBank database, but confirmed in the literature.

References

[1] Doan TL, Pollastri M, Walters MA, Georg GI. The future of drug repositioning: Old drugs, new opportunities. Annu. Rep. Med. Chem., 2011, 46: 385-401.

[2] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res., 2018, 46(D1): D1074-D1082.

[3] The UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res., 2015, 43(D1): D204-D212.

[4] Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome pathway knowledgebase. Nucleic Acids Res., 2018, 46(D1): D649-D655.

[5] Ding C, Li T, Peng W, Park H. Orthogonal nonnegative matrix tri-factorizations for clustering. Proc. ACM SIGKDD, 2006, 126-135.