

Leveraging hierarchical knowledge as guide for representation learning on heterogeneous network for drug repurposing

Dongmin Bang¹, Sangsoo Lim² and Sun Kim^{1,3,4*}

¹ Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea.

² Bioinformatics institute, Seoul National University, Seoul, Republic of Korea

³ Department of Computer Science and Engineering, Seoul National University, Seoul, Republic of Korea.

⁴ AigenDrug, Co., Ltd, Seoul, Republic of Korea.

* Correspondence: sunkim.bioinfo@snu.ac.kr

Over the past decade, new drug discovery and development processes have become increasingly difficult due to rising costs and slowing progress. Drug repurposing, an emerging paradigm, is a new strategy focused on using ‘old’ drugs to treat both common and rare diseases outside the scope of the original indication. By utilizing rapidly accumulating bioassay datasets, computational drug repurposing methods have achieved significant progress. Integrating different databases creates a biomedical heterogeneous network with multiple node and edge type in one network. Various tools, from traditional machine learning model to Graph Convolutional Neural Network, have been proposed to leverage this biomedical network for drug repurposing.

However, representation learning on heterogeneous biomedical network is a difficult and unresolved problem, due to its two major challenges: first, most of the existing methods do not consider multiple node-type and edge-type features of the network, and secondly, biomedical networks are heavily biased towards genes and their interactions, which covers 81% of nodes and 89% of edges of the whole network. To tackle this problem, we propose ‘Random Teleport’ model which uses edge-type transition probability and biological hierarchy data to guide the node representation learning process. In the random walk phase, edge-type transition matrix is trained to represent both node-type and edge-type distribution of the network. In addition to the edge-type transition probability, the guidance of drug/diseases hierarchical similarity as teleport probability enables random walk to generate node sequences that are not heavily biased to the vast-and-dense gene interaction network. Node sequences are then passed on to the skip-gram model for generating node embeddings. After training, given a drug-disease pair, model predicts the probability of drug treating the disease. The key contribution of our work lies in transforming tree-structured hierarchy data into a transition probability for guiding machine learning tasks.

By using three datasets (Hetio network, Multiscale interactome, The Comparative Toxicogenomics Database) for evaluating the performance of ‘Random Teleport’, we showed that guiding random walk process with biological knowledge as teleport probability enables generation of unbiased node sequences, enabling effective node embedding and, in the end, showing better performance than baseline methods in drug-disease link prediction task.