International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Drug Repurposing Using Heterogeneous Networks

M. Venkata Krishna [1], Mohammad Ausaf Ali Haqqani [2], Kallepally Sai Kiran [3], Anandeshi Prathik[4]

[1] Assistant Professor, [2,3,4] Students

Computer Science & Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad

---

Abstract - Drug discovery is a resource consuming and having high attrition rates for the development of a new drug, most of the drugs don't get pass the pre-clinical stage therefore there is a need to find and alternative purpose of existing drugs which will lead to a de-risked clinical trial. This approach is known as Drug Repurposing. There is a complex web structure between the entities resulting in the development of a new drug such as gene, drug, disease, protein and side-effects, A homogenous network will only consider relations in binary options among the entities but such complex structures require precise meaning of what relation is between the entities for which a heterogeneous network is able to model. The outcome by this will hereby significantly reduce the time and lower the attrition rates of finding new drug targets with high success possibility of passing the clinical trials.

***Key Words*:** pre-clinical, Drug Repurposing, heterogenous network.

1.INTRODUCTION

Drug repurposing is nothing but finding new medical uses of already existing drugs, Drug repurposing may allow for more systematic and substantially less expensive methods in the discovery of new treatments for diseases, when compared to traditional drug development repurposing of drug takes less time, traditional drug development takes average of 14 years to discover a new drug whereas drug repurposing take 10 years. The process of discovering new drug is a complex one which requires knowledge from numerous biological and chemical domains such as the mechanism of action by which the disease is caused must be understood so that the drug can be used to treat it, there is lots of data and information available on drugs databases.

Compared with experimental approaches that utilize actual drugs for screening, computational approaches are both time-efficient and cost-effective. Computational approaches involve machine learning, text mining, network analysis and knowledge graphs[1]. increasingly researchers are looking for new ways via which the drug discovery process can be undertaken, both in a more cost-effective manner and with a higher probability of success. Graphs have long been used in the life sciences as they are well suited to the complex interconnected systems often studied in the domain[2]. Homogeneous graphs have, for example, been used extensively to study protein-protein interaction networks, where each vertex in the graph represents a protein, and edges capture interactions between them. Knowledge graphs are heterogeneous data representations, where both the vertices and edges can be of multiple different types, allowing for more complex and nuanced relationships to be captured. In the context of drug discovery, the vertices (commonly known as entities) in a knowledge graph could represent key elements such as genes, disease or drugs – with the edge types capturing different categories of interaction between them. As an example of where having distinct edge types could be crucial, an edge between a drug and disease entity could indicate that the drug has been proven clinically successful

in treating the disease. Conversely, an edge between the same two entities could mean the drug was assessed but ultimately proved unsuccessful in alleviating the disease, thus failing the clinical trial. Homogeneous graph has been used to study PPI network and disease association not showcasing the complex relations which a heterogeneous network is able to show as a crucial distinction in the precise meaning of the relationship between the two entities cannot be captured by a simple binary option[3].

2. RELATED WORK

We have gone through various research works of the people as mentioned below:

[1] They worked on covid-19 disease, implemented three network-medicine drug-repurposing algorithms that relay on AI, network diffusion and network proximity which is used to calculate diffusion state distance to derive a similarity metric for pair of nodes. The network consists of 3 distinct types of entities drugs, proteins and diseases and with 4 types of edges. The network proximity can improve the specificity between different expressible genes. The problem in this approach is there is absence of ground truth for validation of the algorithm on human organism.

[2] Presented machine learning technique on Alzheimer's disease, it uses a specific drug associated gene list for feature selection and evaluates a predictor of Alzheimer's disease, drugs which were found to be targeting proteins associated with the disease were repurposed through connectivity maps. Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD) effort. They focused on gene expression measurements, because it provides a natural connection between preclinical cell culture platforms, here perturbation experiments can be carried out, and patient-derived tissue specimens. Gene expression also features prominently in previous drug repositioning efforts focused on aging and AD with several approaches making use of the Connectivity Map and related tools [5]. The drawback is it gives a biased association between the side effects of the drug and the stage of the disease.

[3] In this network model, protein-protein interaction and drug target interaction were made, so that we can estimate how efficiently the drug is working on the target disease. A network is built by an inhouse program called AutoNet, it includes of a database made up of pathways, protein-protein interactions, meta pathways and drug target interactions from drugbank[6]. A proximity method with proximity distance calculated as a mean of the shortest distances between any drug and each of the above modules. Limited existing knowledge about the disease resulted in focusing on the evaluation of a small number of candidates.

[4] Transcriptome data for COVID-19 and 21 diseases were retrieved from GEO database to analyse the RNA sequence limma and deseq2 R libraries were used and non-Euclidian separation distance measures the distance between two disease specific modules. The distance between associated target and associated gene were preserved with a randomization approach [7]. Limitation is the hypothesis generated through computation need to be tested and validated experimentally.

[5] Past events of clinical trials on patients by testing multiple candidates and inferring through the large-scale medical claims database [8]. It is demonstrated on coronary artery disease. In this approach Randomized clinical trials for drugs were considered. Electronic health records have the presence of uncertainty.

[6] Korea chemical bank drug repurposing database was used to investigate potential drugs for COVID-19 through a virtual screening method with a program called GOLD (Genetic Optimization for Ligand Docking) and molecular dynamic simulation. This virtual screening has a high throughput in screening through vast number of

drugs. Binding free energy calculation is very computer intensive [9].

[7] Knowledge embeddings are an important aspect for predicting the missing links between the entities. The low dimension representation of the entities and relations to predict the possible outcome of a given triple. Multiple hyper parameters for various optimizers can be changed. They can produce results only with combination of well-tuned approaches and doing manual hyper parameter optimization.

[8] It leverages path-based learning and representation learning with logical rules through a graph traversal method and evolving deterministically with meta paths among the entities. Logical rules are written in the form of triplets for giving the relations in generic form. It does not consider relevant bar medical prediction data which hampers the link prediction process.

[9] It utilizes knowledge bases by extracting data triplet from it and then it is integrated into a knowledge graph then modelled into a graph data thereby created a drug centric model. It uses a path-based representation for the data and calculating the overall path cost in a normalized manner. It does not consider embedding base representation for the relations thereby not utilizing the full potential of triplets.

[10] The virtual ligand library consisting of 61 reported antiviral agents are studied through a molecular docking process on the proteins of COVID-19 after which repurposing is done.

The docking interactions are of high accuracy. Molecular docking is very intensive on computer simulation with many constrains to be optimized.

[11] It uses a variational graph auto encoder with graph embeddings derived from the TDD base. Multi relational graph neural networks are used to represent the biological interaction with validation done through biological means. PPI and drug adverse effects are taken into consideration for repurposing. The knowledge graph was not comprehensive and up to date with the required interactions for link prediction.

[12] It uses genetics informed network by targeting the pathways between genes and the interactions between different molecules are considered resulting in the development of the disease network with validation and experimental models. Multiple combinations of associated genes are used for the disease network. Validation is only done through experiments using animal models

[13] A molecular compounds similarity prediction model for computing four different compound-compound similarity scores based on bio activity, chemical structure, target enzyme and protein with similar biological and physiochemical properties. It helps in identifying novel drug targets and indications [16]. The model is highly dependent on the structural properties of the compounds.

[14] It uses the Connectivity Map Approach (CMAP) and structural based virtual screening to repurpose the drugs against selected targets. CMAP helps to find functional connections with mode of action of the drug. Prevalence of uncertainty in the bio activity of drugs is present [17].

[15] It uses a frozen pre trained transformer to find the relations among unstructured text between multiple entities with representation of knowledge in a distinct and predictive manner. Text mining gives us semantic representation of the scientific literature on drugs. The only metric used for validating the model is F1-Score which is not significant in real world performance [18].

3. METHODOLOGY

We aim to reposition drug that have been previously tested in clinical trials and show new indications for them.

In this project, we proposed an approach that combines literature-based discovery and knowledge graph completion for drug repurposing. Unlike similar efforts that largely focused on drug-specific knowledge, we incorporated knowledge from a wider range of biomedical literature.

The approach proposed here is not specific to a disease and can be used to repurpose drugs for other diseases [4]. It can also be generalized to answer other clinical questions, such as discovering drug-drug interactions or identifying drug adverse effects. Innovative computational methods leveraging existing biomedical knowledge and infrastructure could help us plan for, respond to, and mitigate the effects of global health crises. Drug repurposing is a key piece of this response, and our approach provides an efficient computational method to facilitate this goal. Knowledge graph is given as the input to a Knowledge Embedding model using the DGL-KGE (Deep Graph Library – Knowledge Graph Embeddings) library which learns the best representation of the relations such that the domain specific knowledge is retained through the low dimensional vectors of each entity and relation. The embeddings are used to calculate a plausible scalar value of a new relation between the head and tail which is analogous to a drug being repurposed for a new therapeutic effect and a new phenotype. The drugs are ranked in the descending order of similarity score for a specific disease, target, gene or side effect. The modules

Include in our project are:

- Building the knowledge graph:

We use CTDbase to find drug-target interactions, pathways, and gene/drug-phenotype relationships to build the knowledge graph. This graph had four types of nodes and five types of edges which included PPI, drug target interaction, biological pathways and gene/drugphenotype interactions.

- Embedding using graph neural network:

We used deep graph neural embedding with multi-relational data, and we used variation graph auto-encoders with message passing neural networks to account for uncertainty, learn the graph distribution, and reconstruct missing interactions using node embeddings in a supervised manner

- Evaluating the knowledge graph embedding:

We used the t-SNE plot to validate the confidence of a knowledge network embedding by viewing lower dimensional projection and observing the distribution of high dimensional node embedding

- Initial drug ranking:

The model is created to choose the most effective medications after the drug embeddings have been determined, these were extracted from NIH clinical trials using a customized neural network based on Bayesian probabilistic inference.

- Population based validation:

The investigation of drugs administered to the particular disease was estimated using optum EHR database.

- Drug combination search:

Our strategy was to test the hypothesis that a "drug combination has a theuraptic impact if the targets of the individual medications hit the disease module but the target has a different neighborhood" based on the pattern.

4. RESULTS

Our strategy was to test the hypothesis that a "drug combination has a theuraptic impact if the targets of the individual medications hit the disease module but the target has a different neighborhood" based on the pattern. We took DRKG dataset, which consist of 97,238 entities and 5,874,261 triplets belonging to 107 edge-types.

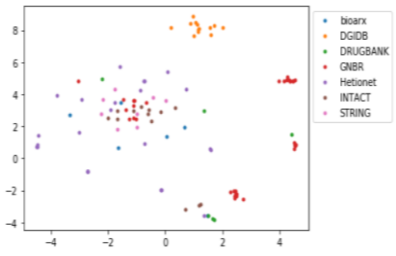

Fig 1: ANALYZING THE GENERATE RELATION EMBEDDING SIMILARITY

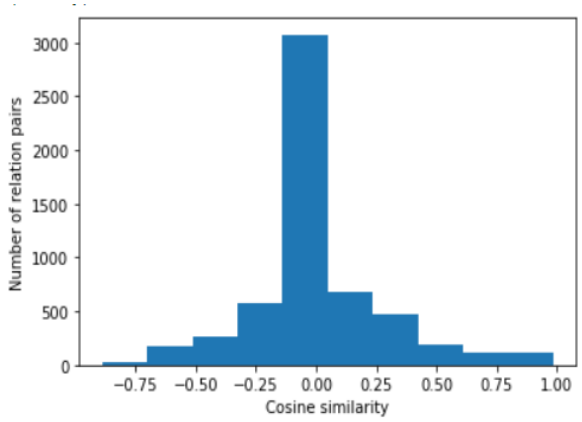

Fig 2: PAIR-WISE RELATION EMBEDDING COSINE SIMILARITY

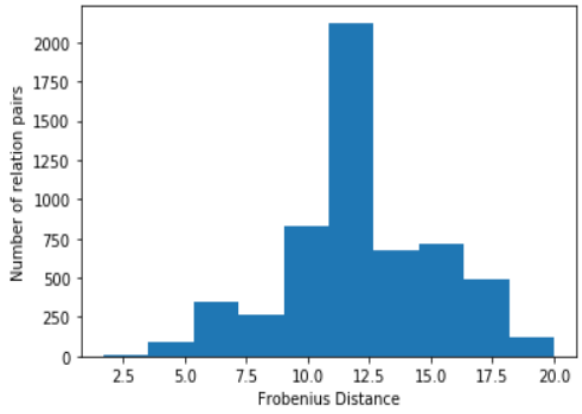

Fig 3: ANALYZING THE GENERATE ENTITY EMBEDDING SIMILARITY

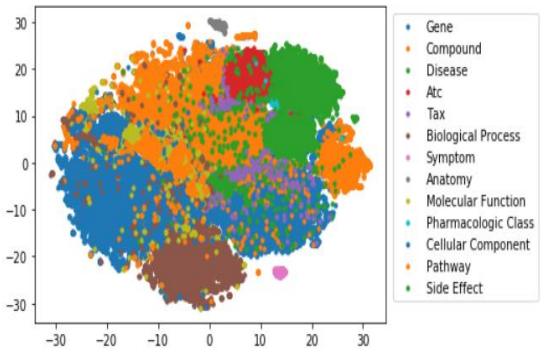

Fig 4: EVALUATING WHETHER THE LEARNED KGE MODEL CAN PREDICT THE EDGES OF DRKG

```
Out[9]: [('Molecular Function::GO:0016817',
          'Molecular Function::GO:0016818',
          0.9998758),
         ('Biological Process::GO:0045936',
          'Biological Process::GO:0010563',
          0.9998726),
         ('Cellular Component::GO:0099512',
          'Cellular Component::GO:0099513',
          0.99986345),
         ('Biological Process::GO:0048870',
          'Biological Process::GO:0051674',
          0.999861),
         ('Cellular Component::GO:0000323',
          'Cellular Component::GO:0005764',
          0.99985904),
         ('Biological Process::GO:0043207',
          'Biological Process::GO:0051707',
          0.999855),
         ('Biological Process::GO:0044403',
          'Biological Process::GO:0044419',
          0.9998515),
         ('Biological Process::GO:0007272',
          'Biological Process::GO:0008366',
          0.99984413),
         ('Biological Process::GO:0006935',
```

Fig 5: EVALUATING HOW SIMILAR ARE THE PREDICTED LINKS AMONG DIFFERENT RELATION TYPES

After building the model we have given covid -19 clinical trials drugs as input. It will display the output as shown below.

```
[0]   Ribavirin        -0.21416784822940826
[4]   Dexamethasone    -0.9984006881713867
[8]   Colchicine       -1.080674648284912
[16]  Methylprednisolone      -1.1618402004241943
[49]  Oseltamivir      -1.3885014057159424
[87]  Deferoxamine     -1.513066053390503
```

Fig 6: REPURPOSED DRUGS FOR COVID-19

5. CONCLUSION

Repositioning the drug which has been previously tested in clinical trials and show new indications for them. Find the multiple combinations of drugs either as beneficial or having toxic consequences in interaction with the biological system. Bias evident in different data sources

is removed in this heterogeneous graph. Meaningful evaluation of multiple relation types as a single metric is to done. Common metric is used considering the influence on real world experimentation.

6. FUTURE SCOPE

To circumvent most of the expensive drug discovery processes and increase their productivity by reducing the discovery and development timeline. The pharmacokinetic profiles of the future candidates are to be predicted accurately. The molecular dynamics simulation of the drug with the protein at the binding pocket with high binding affinity is to be integrated. The clinical trials of the new drug must be predicted beforehand to reduce the attrition rate of the new drugs.

7. REFERENCES

[1] Deisy Morselli Gysi, Ítalo do Valle, Marinka Zitnik, Asher Ameli, Xiao Gan, Onur Varol, Susan Dina Ghiassian, J. J. Patten, Robert A. Davey , Joseph Loscalzo, and Albert-László Barabási" Network medicine framework for identifying drug-repurposing opportunities for COVID-19"

[2] Steve Rodriguez, Clemens Hug, Petar Todorov, Nienke Moret, Sarah A. Boswell, Kyle Evans, George Zhou, Nathan T. Johnson, Bradley T. Hyman, Peter K. Sorger, Mark W. Albers & Artem Sokolov" Machine learning identifies candidates for drug repurposing in Alzheimer's disease"

[3] Xu Li, JinchaoYu, Zhiming Zhang, Jing Ren, Alex E.Peluffo, Wen Zhang, Yujie Zhao, Jiawei Wu, Kaijing Yan, Daniel Cohen, WenjiaWanga" Network bioinformatics analysis provides insight into drug repurposing for COVID-19."

[4] Pasquale Sibilio, Simone Bini, Giulia Fiscon, Marialuisa Sponziello, Federica Conte, Valeria Pecce, Cosimo Durante, Paola Paci, Rosa Falcone, Giuseppe Danilo Norata, Lorenzo Farina, Antonella Verrienti" In silico drug repurposing in COVID-19: A network-based analysis"

[5] Ruoqi Liu, Lai Wei & Ping Zhang" A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data."

[6] Raj Kumar, Vikas Kumar, Keun Woo Lee." A computational drug repurposing approach in identifying the cephalosporin antibiotic and anti-hepatitis C drug derivatives for COVID-19 treatment."

[7] Stephen Bonner, Ian P Barrett, Cheng Ye, Rowan Swiers, Ola Engkvist, Charles Tapley Hoyt, William L Hamilton" Understanding the Performance of Knowledge Graph Embeddings in Drug Discovery"

[8] Yushan Liu, Marcel Hildebrandt, Mitchell Joblin, Martin Ringsquandl, and Volker Tresp." Integrating logical rules into neural multi-hop reasoning for drug repurposing."

[9] Yongjun Zhu, Chao Che, Bo Jin, Ningrui Zhang, Chang Su, and Fei Wang." Knowledge-driven drug repurposing using a comprehensive drug knowledge graph."

[10] Bhumi Shah, Palmi Modi, Sneha R. Sagar" In silico studies on therapeutic agents for COVID-19: Drug repurposing approach."

[11] Kanglin Hsieh, Yinyin Wang, Luyao Chen, Zhongming Zhao, Sean Savitz, Xiaoqian Jiang, Jing Tang, and Yejin Kim." Drug repurposing for covid-19 using graph neural network with genetic, mechanistic, and epidemiological validation."

[12] Serguei Nabirotchkin, Alex E Peluffo, Philippe Rinaudo, Jinchao Yu, Rodolphe Hajj, Daniel Cohen." Next-generation drug repurposing using human genetics and network biology."

[13] Suganya Chandrababu, Dhundy Bastola." A Novel Prediction Model for Discovering Beneficial Effects of Natural Compounds in Drug Repurposing"

[14] Harshita Tiwari, Manas Ranjan Barik, Amit Nargotra." Application of Computer-Aided Drug Design for Lead Identification and Drug Repurposing."

[15] Yongjun Zhu, Woojin Jung, Fei Wa, Chao Che" Drug repurposing against Parkinson's disease by text mining the scientific literature"

[16] David Mendez, Anna Gaulton, A Patr´ıcia Bento, Jon Chambers, Marleen De Veij, Eloy Felix, Mar ´ ´ıa Paula Magarinos, Juan F Mosquera, Prudence Mutowo, Michał Nowotka, et al. ˜ Chembl: towards direct deposition of bioassay data. Nucleic acids research, 47(D1):D930– D940, 2019.

[17] Sunghwan Kim, Paul A Thiessen, Evan E Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A Shoemaker, et al. Pubchem substance and compound databases. Nucleic acids research, 44(D1):D1202–D1213, 2016.

[18] David S Wishart, Craig Knox, An Chi Guo, Dean Cheng, Savita Shrivastava, Dan Tzur, Bijaya Gautam, and Murtaza Hassanali. Drugbank: a knowledgebase for drugs, drug actions and drug targets. Nucleic acids research, 36(suppl 1):D901–D906, 2008.