Drug Repurposing through Graph Machine Learning for COVID-19

Gurvinder Singh
University at Buffalo
gurvind2@buffalo.edu

ABSTRACT

Coronavirus disease has created a worldwide pandemic situation where millions of people have suffered and died.Graph Machine Learning has the ability to model both structural and functional relationships between entities, whether it is drug structure or disease interaction. In this Survey paper, After establishing the key background in biomedical and Graph machine learning will move toward Drug repurposing.Here, I'm presenting the Key work done in past couple of years by scholars to apply Graph machine learning techniques to do drug repurposing for Covid-19 disease.

Keywords

Drug Repurposing, Knowledge Graph, Covid-19, Graph Machine Learning

1. INTRODUCTION

The current Covid-19 pandemic has unveiled the need for fast drug development and availability. Whereas the current process from drug discovery to drug manufacturing, on average takes 12-14 years of time and costs around 3-4 billions dollar[16]. In spite of advancements in pharmaceutical research and development, the number of new treatments approved by the FDA has remained very low. One of the ways out of this whole process is Drug Repurposing, where drugs that have been approved for other diseases, are identified for their alternative use against new diseases such as Covid-19. Drug Repurposing can reduce the whole timeline to 5-6 years because the initial work of drug discovery and safety analysis of drugs have already been done.it is less likely to fail, at least from a safety point of view, in subsequent efficacy trials. So, the candidate drug only has to go through clinical trials to measure the efficacy and the right dosage so that approval for alternative use would be done.

One of most popular example of drug repurposing is of Sildenafil[44], which was originally developed as an antihypertensive drug; later it was repurposed for the treatment of erectile dysfunction and named as Viagra. Another example is of Rituximab[49] which was originally developed for Cancer treatment but was later, after clinical analysis found to be effective for Rheumatoid arthritis patients. Global sales of rituximab topped 7 billion dollars in 2015.[1] Such successes have encouraged the Pharmaceutical companies to identify the repurposable compounds as it can bring huge profits with very less development cost.

One of the biggest Aids to drug repurposing is the presence of ever-growing databases of scientific literature, clinical trials, Genomics, protein and drug structure information, and the inter-connected nature of these data structures. There are several approaches present in drug repurposing starting from Genetic Association where Genes that are associated with a disease proves to be potential targets[23]. Another Network analysis approach is Pathway mapping, where genetic, protein or disease data combination can lead toward repurposing targets. A good comparison approach of comparing structure, transcription sites or adverse effects of one drug with another can lead to potential targets. This heterogeneous nature of interaction between same entities like interaction of different genome sites or interaction between different proteins and relationship between genome and protein are best described in form of Graph/Networks[7].

2. CURRENT PROCESS

The current process of Drug discovery and development goes through several steps which starts with Target identification[42] where disease targeted genes and protein get identified and whether the identified regions are druggable or not. As in this early stage, the number of compound which can be drug candidates are in thousand these candidate drugs are synthesize in labs during the Candidate synthesis process. [21]Binding Affinity test the strength of the binding interaction between a single biomolecule mostly (protein or DNA) to its ligand/binding partner such as drug or inhibitor. The drugs which have lower binding Affinity with target or high binding affinity with Off-target protein/DNA would be eliminated from the candidates. ADME (absorption,distribution,metabolism and Excretion) [19] properties allows drug developers to understand the safety and efficacy of a drug candidate, the candidates which are not within the required range of safety are eliminated from this pool. candidate drugs are usually changed to improve the binding

[1]Pfizer has repurposed the Sildenafil for the erectile dysfunction use and it earns near to 30 billion dollars from its sale till date[2]

affinity and Adme properties in the Lead optimization steps, After this Preclinical trials are done on the Animals to measure the effects of the drugs on alternate cell lines. Only few of initial selected drugs were able to reach to clinical trials where three separate trials are done on humans to measure efficacy and short or long term side effects. The final submission to FDA were just one or two drugs out of the initial pool of thousands of candidates. FDA approval takes long time as all the trials and test documents were checked thoroughly for accuracy and it can be possible that drug has to go through more trials or testing before the final approval. Once Approved the drug companies started to setup faculties to mass produce the drugs which further take time and resources. During the Covid-19 situation many efforts were done for the repurposing of drugs so that this long cycle of development would be avoided. Drug repurposing usually relies on identifying the novel interaction among different biological entities like genes and compounds As traditional approaches for doing that rely on costly and time consuming experimental methodologies

3. BACKGROUND

The field of computation biology is huge as it involve understanding of two different fields. Deep understanding of computation methods with machine learning and diverse field of biology from genetics to molecular biology. So, the main challenge for researcher in this field is the difference in the vocabulary. This discussion on Background is an attempt to bridge that gap.

3.1 Genomics and Transcription

The blueprint for protein creation is stored in the genes and these proteins are combined together to form cells and tissues which are the further combined together to make organs. These genes are tightly packed together in the form of a chromosome using a stabilizing protein called histone. The whole human genome is divided between these 23 pairs of chromosome and these chromosome present in all cells of the body except reproductive cells, or gametes, which carry just one copy of each chromosome[1]. The chromosome is usually present in a Free form inside the cell nucleus and only comes to a condensed form while doing replication. As the instruction to create which type of protein is embedded in the Gene in form of code like sequence of AGCT. these codes are read in the transcription process. During the transcription process, RNA Polymerase, which is a kind of protein, attaches to transcription active sites and starts opening the double helix bond and creates a mRna, a messenger Rna, which has the instructions to create the protein. The transcription of the genome is controlled through Gene Regulation where Active repressor stops and controls transcription by blocking the Active site from RNA polymerase. The other regulatory element is the Histone packaging, which can block Active sites by tight packaging and doesn't have a region for the polymerase to bind. These regulatory elements are important for the correct functioning of cells.

The mRna produced during the last step is moved out of the nucleus and picked up by the Ribosome, which starts the Translation process during which Amino Acid binds with the tRna, creating the peptides chain based upon mRna instructions. This translation happens at 3 pairs at a time and at the end of the chain protein folds into complex 3D structures from a chain. The challenge of predicting the 3D structure of the protein from its mRna is a well-known Protein Folding problem. The Deepmind team[32], which developed the AlphaFold Model which predicted the structure of the protein from its mRna instructions, predicted five understudied SARS-CoV-2 targets protein. As the accurate model of protein 3D structure helps further understanding of whether drugs will bind to the protein or not and what are the best binding sites.

3.2 Protein Drug Interactions

Protein combined with other proteins to create cells and tissues. So, protein-protein interactions are important for the defense system of the body. The drugs effect this interaction between proteins by two different modes first is competition and other is Stabilization[38]. In the Competition mode, the drug weakens the interaction by either binding to the available position for the second protein or by decreasing the attraction energy of the first protein to weaken the strength of the interaction between the two proteins. In the Stabilization mode, the drug binds to the binding site of the first protein and increases the area to the binding site so that other proteins can bind with the first. The drug can also increase the attraction of the first protein and strengthen the interaction between the two proteins. The Binding Affinity is an important property while selecting the drug out of candidate compounds as it measures the strength of the binding between the drug and the protein. The compounds which can't bind to the protein are usually eliminated from the target drug list. The drug binding can also help to fight viruses by attaching to the target binding site of the virus and blocking the virus from advancing further into the cells and starting the replication process.

3.3 Crispr and Cas9 Protein

The new technique to fight disease is the Crispr technique[27] Crispr is a family of DNA sequences found in the genomes of prokaryotic organisms such as bacteria and archaea. Crispr stands for Clustered Regularly interspaced short palindromic repeats; it is a gene editing technique where Cas9 protein forms a complex with the guide RNA in a cell then this complex attaches a matching genomic DNA sequence adjacent to a spacer. The Cas9-Rna with its cutting heads cuts the double strands of the DNA and then programed DNA can be inserted at the cut. This way, harmful mutations in genes can be fixed. as these mutations are sometimes responsible to a weak immune system of the body which can be fixed using the Crispr technique. As the New Application of the Crispr[37] is coming out where it is also used as a regulatory system which disables the Active sites for the transcription in the genome. Other applications of Crispr are to mark the area of the genome which is responsible for the creation of a particular protein. this helps to create interconnections between Proteins and Genome.

3.4 Type of Drugs

The Coronavirus like the one that causes Covid-19, is named because of their crown-like spikes on their surface, called spike proteins[3]. These spikes are ideal targets for vaccines. There are usually three main approaches to making a antiviral vaccine. First one is to use a weakened or dead virus where a drug containing a dead virus is introduced into the

body so that the body immune system learn how to identify the virus and make antibodies to that. The measles, mumps and rubella (MMR) vaccine and the chickenpox and shingles vaccine are examples of this type of vaccine. A second approach is to use a subunit of the whole virus where only very specific parts of a virus or bacterium that the immune system needs to recognize. This subunit can be protein like spike protein or sugar. Most of the vaccines on the childhood schedule are subunit vaccines, such as whooping cough, tetanus, diphtheria and meningococcal meningitis. A third and new type of vaccine uses just part of the genetic material of the virus. A nucleic acid vaccine such as some of covid-19 vaccine, uses a specific set of instructions to our cells, either as DNA or mRNA, for them to make the specific protein that we want our immune system to recognize and respond to. The vaccines, which are made of mRna, are wrapped in a coating that makes them easy to deliver inside the cell. These mRNA vaccines teach your cells how to make copies of spike proteins, so that if you are exposed to a real virus later, your body's immune system will recognize it and know how to fight it off.

4. DATASETS AND REPRESENTATION

4.1 Datasets

As we know machine learning requires huge amounts of data to model the complexities. For Graph machine learning, the source of data required is Graphs. In the Past Homogeneous Graphs have been extensively studied for the protein-protein interactions; Drug-drug interactions where each node represents either protein or drug and an edge captures the interaction between them. But Recently there is a push toward Knowledge Graphs which usually used for the Drug repurposing tasks[9]. Knowledge graphs are heterogeneous data representations where both edges and nodes can be of different types. These entities usually are genes, diseases or drugs. The edge between drug and disease indicates that the drug is able to treat the disease. Knowledge graphs are the most used tool while analyzing the COVID-19 disease as it enables recently developed graph machine learning tools to be exploited[8].

To Construct the knowledge graph, there is need to map the drugs, disease, pathways correctly. This mapping will take these homogeneous graphs and put them together using the single ID; this preprocessing needs to be done because these individual datasets are maintained by different government and private entities. which uses a different structure for the dataset. The need for metadata is required if you want to use different data sources to do the mapping as metadata has information about structure and sources So these metadata text can be parsed to get the correct hypergraph edges.

The Literature-based Knowledge discovery is also quite common where multiple Triplets of subject – predicate – object were parsed from the Text so that it can be used in a put in a form of Knowledge graph. These documents can be extracted from biomedical literature and COVID-19 papers, case statistics, experimental data. One of the biggest releases in a form of dataset is Covid-19 Open Research Dataset (CORD-19), which is a growing resource of scientific papers on Covid-19 and related historical coronavirus research. This Dataset contains over 50K scientific papers from Pubmed central, bioRxiv, medRxiv and WHO[54].

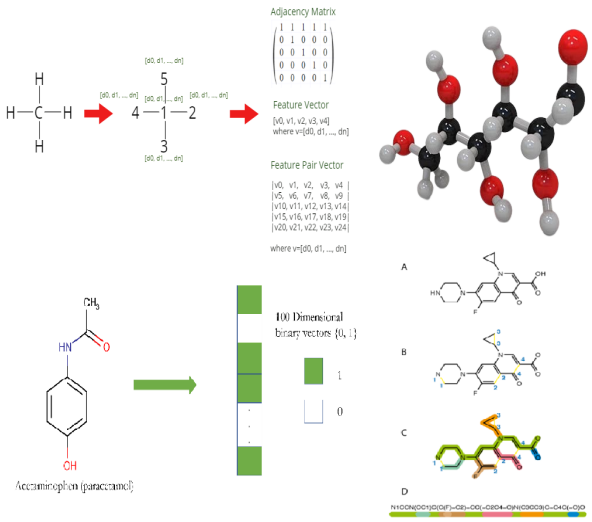

Figure 1: Top part is showing how 2D/3D structure of molecules can be represent in form of Adjacency matrix. Bottom part of image is showing the Fingerprint and Smile representation.

4.2 Representation

Structure and relationships can be represented in the form of graphs. Usually chemical structures are three-dimensional and this 3D object consists of atoms with well-defined locations within a molecule. Relative distance to each molecule, energy present in structure, shape of electron probability cloud are some of the information which would be lost if we represented the structure in two dimensions instead of three. The weighted, unweighted, undirectional and directional structure can be represented in form of Adjacency Matrix.

Another way of representing the molecular structure is a bit-like pattern generation for the absence or presence of certain substructures within a larger structure. This is called fingerprinting where a bit-like pattern can represent anything from Absolute/Atomic charge to presence or absence of different type of bond. Smile is a text-like representation of the molecules which take this two-dimensional structure and convert it into the one-dimensional sequences but features like length of atomic bonds and 3D coordinates position of molecules lost in this conversion.

4.3 Knowledge Graphs

Knowledge graphs[2] more often built using resources covering primary information on the various relevant entities and relations. One of the first attempts to build a knowledge graph on drug discovery was Hetionet. The data used for this knowledge graph is from Gene,Drugbank, DisGeNet, Reactome and Gene ontology[24] dataset. The next is Drug Repurposing Knowledge graph which is built upon HetioNet. DRKG was first built to tackle drug repurposing for Covid-19. DRKG was enriched using data from STRING, Drug-

[2]https://het.io/
https://github.com/gnn4dr/DRKG
https://github.com/dsi-bdi/biokg
https://github.com/MindRank-Biotech/PharmKG[58]

Table 1: Biomedical Datasets

Dataset Type	Dataset	Comments
\Pathway	Reactome,WikiPathways,KEGG pathways	For Pathways,Reactions
\Compounds	Drugbank,DrugCentral,RepoDB,CHEMBL,PubChem	Drugs and Chemicals
\Disease	KEGG disease,DisGeNET,OMIM,GWAS	Link Genes, Disease
\Interactions	Omnipath,BIOGRID,STRING,PPI	Interactions Gene and proteins
\Genes	RNACentral,UniProtKB,Ensembl,BioSNAP	Transcripts,Protein
\Drug discovery	Pharos,Open Target Platform	Focused on Target discovery
\Anatomy	ATLAS,BIOGPS	Relate Anatomy with Disease
\Covid RNA-SEQ	ImmGen,NCBI,Broad	SARS-COV2 Sequence,Literature

Bank, and GNBR resources. BIOKG is another Knowledge graph which used various biomedical resources including UniProt[4], Reactome[17], OMIM and Gene Ontology. Other Knowledge graphs which combine multiple resources are PharmKG, OpenBioLink, Clinical KG. one of the earlier work in Covid-19 drug repurposing is Knowledge graph completion, where task is of predicting unseen relations between two existing entities or to predict the tail entity given the head entity and the relation. Here, both classical Network based approaches and newly developed Graph machine learning approaches were applied to find the correct relationships.

5. GRAPH MACHINE LEARNING

Most graph machine learning methods can be divided into two parts. general-purpose encoder and task-specific decoder[12].The encoder embeds graph nodes, into low dimensional feature space, then applies permutation invariant pooling functions (like Sum, max, or mean) to produce graph level embeddings. The decoder takes the encoder produced embedding and outputs task-specific predictions. The end-tasks can be Graph Classification or regression, node classification, link prediction, clustering, or community detection and Generative tasks like graph generation etc. These tasks can be performed in a Inductive or Transductive way[33]. Transductive tasks expect that all data points are available when learning a encoder function, including unlabelled data points. whereas inductive correspond to classical supervised learning algorithm and is more general than transductive. Variating the graph along the temporal dimension also results in changes to composition, structure and attributes. Diverse graph centrality measures different aspects of graph connectivity; Like closeness centrality measures how closely a node is connected to all other nodes. Other centrality like betweenness centrality measures how many shortest paths between pairs of other nodes, a given node is part of. Motifs[39] and graphlets[43] corresponds to local wiring patterns of nodes. These centrality features with other node-level features can also be used as input to the machine learning model.

Geometric approaches posit each relation type between two nodes as a geometric transformation from source to target in the embedding space. one of early work is TransE[10], which is a purely translational approach where translation between head and tail entity happens through relation in continuous vector space. TransE employs distance-based score functions which are mostly either L1 or L2 norms. One downside is its ability to account for symmetric relationships or one-to-many interactions. One counter part of TransE is RotatE[50], which represents relations as rotations in a complex latent space. such that rotated vectors lie close to the target node vector in terms of Manhattan distances. [3]

5.1 Message passing networks

A message passing system[20] has three functions; first, a message-passing function that permits information exchange between nodes over edges. The second is Aggregation function that collects the message from the neighbouring nodes and combines it in a permutation invariant way. third, is an update function that produces the node-level representation given the previous representation and the aggregated messages. although it is not uncommon for the message or update function to be absent or reduced to an activation function only.

$$msg_{ji} = Msg(h_j^t, h_i^t, x_{j,i}^e)$$

$$h_{ji}^{t+1} = Update(h_i^t, Agg(msg_{ji}, j \in N_i))$$

Here h_t are node representations after layer t. initial node representations are typically set to node feature x_i^v. Here Aggregate function is a permutation invariant which can be any of element wise min-max Pooling or a LSTM function. For the Update function choice, it is usually Multilayer preceptrons with a Relu or Sigmoid activation.

5.2 Graph convolutional network

Graph convolution network [33] are a slight variation on the neighborhood aggregation idea. It also takes inspiration from CNN, where CNN's are specially built to operate on regular or Euclidean structured data, while GCN can also work where the numbers of nodes connections vary and the nodes are unordered, basically irregular or non-Euclidean

[3]Some of the Graph Machine Learning Libraries:
github.com/pyg-team/pytorchgeometric
github.com/deepchem/deepchem
github.com/rdkit/rdkit
github.com/elix-tech/kmol
torchdrug.ai
github.com/divelab/DIG
github.com/dmlc/dgl

Table 2: Knowledge Graph Generation[9]

KG Dataset	Gene-Gene	Gene-Disease	Gene-Drug	Drug-Drug	Drug-Disease
\HetioNet[25]	HID,LINCS	DisGeNET,GWAS C	LINCS,DrugCentral	DrugBank	MEDI,LabeledIn,
\DRKG[28]	STRING,GNBR,IntAct	GNBR	GNBR,IntAct,DGIdb	Drugbank	GNBR,DrugBank
\BIOKG[53]	IntActUniProt	KEGG,OMIM	KEGG,IntAct	DrugBank	DrugBank
\OBL[11]	STRING	DisGeNET	STITCH	-	DrugCentral

structured data. GCN in a neighborhood aggregation setting can be represented as:

$$h_v k = \sigma\left(W_k \sum_{u \in N(v) \cup v} \frac{h_u^{k-1}}{\sqrt{|N_u||N_v|}}\right)$$

Here W_k applies the same transformation matrix for self and neighbor embeddings. instead of a simple average, normalization varies across neighbors, this helps Down-weight high degree neighbors.

An efficient implementation of GCN is from sparse batch operations which uses Spectral method. Spectral GCNs make use of the Eigen-decomposition of graph Laplacian matrix to implement this method of information propagation. Eigen-decomposition helps to understand graph structure. In this implementation, Adjacency Matrix takes into account in addition to the node features. The use of Adjacency matrix in the forward pass enables the model to learn feature representations based on nodes connectivity. The resulting GCN model is a first-order approximation of Spectral Graph Convolution, which behaves like a message-passing network where the information is propagated along neighboring nodes within the graph. In this equation , a normalized version of the Adjacency matrix is being used.

$$H^{k+1} = \sigma\left(D^{-\frac{1}{2}}\widetilde{A}D^{-\frac{1}{2}}H^k W_k\right)$$

$$\widetilde{A} = A + I$$

$$D_{ii} = \sum_j A_{i,j}$$

This implementation gave the time complexity of $\mathcal{O}(|E|)$

5.3 Graph attention network

Graph attention networks (GAT)[52] weights incoming messages with an attention mechanism. This signifies the different importance score to nodes of the same neighborhood and this works in both transductive and inductive settings. The attention weights, α_{ij} are softmax normalized, that is

$$\alpha_{ij} = softmax_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k=1}^{N}\exp(e_{ik})}$$

where e_{ij} is output of a single layer NN without a bias with LeakyReLU activations, that takes the concatenation of transformed source and target node features as input.

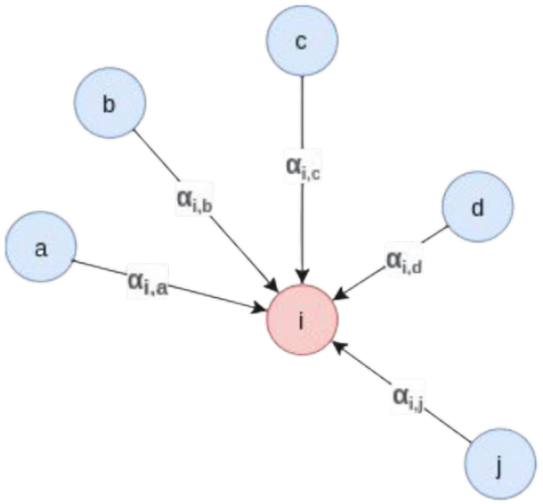

Figure 2: Here the weightage score $\alpha_{i,j}$ is assigned to the messages coming from the neighboring nodes.

$$e_{ij} = MLP\left([W^t h_i^t || W^t h_j^t]\right)$$

Computationally, it is highly efficient as the operation of the self-attentional layer can be parallelized across all edges, and the computation of output features can be parallelized across all nodes.

5.4 Variational Graph AutoEncoder

variational graph autoencoders (VGAE) [34] is a framework for unsupervised learning on graph-structured data. This model makes use of latent variables and is capable of learning interpretable latent representations for undirected graphs. It has been shown to be of utmost accuracy, to predict links in networks that, although missing, are highly likely to exist. Traditional Variational Autoencoder could generate new data from the original source distribution. As the Architecture of the VAE form encoder and decoder part. Encoder embeds the input X to a distribution which is usually parameterized as a multivariate Gaussian rather than a point And then a random sample Z is taken from the distribution rather than generated from the encoder directly.VGAE that applies the idea of VAE to graph-structured data extend it's capability to generate new graphs or reason about graphs.

The encoder or inference model of VGAE consists of GCN and takes an adjacency matrix A and a feature matrix X as inputs and generates the latent variable Z as output. The first GCN layer generates a feature matrix to be passed to

the second layer.

$$X^{'} = GCN(X, A) = ReLU(\widetilde{A}XW_0)$$

$$\widetilde{A} = D^{-\frac{1}{2}}\widetilde{A}D^{-\frac{1}{2}}$$

Then the Second GCN layer generates the μ and $log\sigma^2$ and we calculate the Z using the parameterization trick.

$$\mu = GCN_{\mu}(X, A) = \widetilde{A}X^{'}W_1$$

$$log\sigma^2 = GCN_{\sigma}(X, A) = \widetilde{A}X^{'}W_1$$

$$Z = \mu + \sigma * \epsilon$$

The decoder is defined by an inner product between latent variable Z and the output of our decoder is a reconstructed adjacency matrix $\hat{A}$

$$\hat{A} = \sigma\left(zz^T\right)$$

In general, the encoder represented as

$$q\left(z_i|X, A\right) = N\left(z_i|\mu_i, diag(\sigma_i^2)\right)$$

and the decoder as

$$p\left(A_{ij} = 1|z_i, z_j\right) = \sigma\left(z_i^T, z_j\right)$$

Inner product can calculate the cosine similarity of two vectors, which is useful when we want distance measure which is invariant to the magnitude of the vectors. Therefore, by applying the inner product on the latent variable Z and Z^T, we can learn the similarity of each node inside Z to predict our adjacency matrix.

6. APPLICATIONS

Graph Machine Learning has lots of applications in drug discovery and development processes. Before Covid-19, GML were applied in range of tasks Starting from Early stages like target identification where search for a molecular target with a significant functional role in the pathophysiology of a disease or targeting other classes of bio-molecules is also possible, such as nucleic acids [42]. GML role in finding chromatic interaction between different genome sequences where an algorithm that uses graph embedding followed by clustering to predict sub-compartments using Hi-C chromatin interaction data [6]. GML methods are already State of the art on molecular property prediction tasks such as well-established benchmarks like QM9 and MD17. Methods like SchNet[47] and DimeNet[35] uses the message passing framework where MLP transformation done to each atom representation input and update for the atom embeddings based on the embeddings of the other atoms of the molecule. where SchNet uses 2D structure of the molecules, while DimeNet relies on 3D structure of the molecule and the message from neighbors to current node is updated based on neighbors incoming messages as well as the distances between atoms and the angles between atomic bonds. Graph machine learning has also been applied successfully in predicting the binding Affinity of the protein and drug. Paper by Mingjian Jiang et al. [30] shows the implementation of DGraphDTA. where for drug molecules uses the SMILES representation for molecule graph and for protein graph construction, the contact map is constructed based on the protein sequence. After getting two graphs, they applies two GNNs to extract the representations, then the final representations are concatenated for affinity prediction.

GML uses some of the learning from the previous stages of drug development and discovery to do repurposing of drugs. Here, work on Off-Target repurposing to find missing drug–target interactions has been done with or without the requirement of the structure of the protein. Like Torng et al.[51] worked on the task of associating drugs with protein pockets they can bind to. Here, drugs' structure is based on their atomic structures and protein pockets are represented based upon their Euclidean distance between a set of key amino acids. For obtaining the embeddings of protein pockets, autoencoders were used and in the last, MLP layers were applied on the concatenated embeddings of protein pockets and drug representation to get score for drug protein binding. In contrast, Nascimento et al.[41] proposed KronRLS-MKL, a method that casts drug–target interaction prediction as a link prediction task on a bi-partite graph capturing drug–protein binding. The author used Kronecker operator to obtain a kernel between drug and protein pairs. The kernel is then used to predict a drug–protein association based on their similarity to existing drug–target links.

Alternatively, On-target repurposing takes the full picture into account and uses known targets of a drug to infer new unknown indications based on diverse data.Knowledge graph approaches have been particularly used as an on-target repurposing task. Here one can identify functional relationships between a drug's targets and genes associated with a disease. For instance, Fang et al.[18] finds missing protein-protein interactions between a genetically validated target and a drug's primary target to repurpose drug candidates.

6.1 Modelling knowledge graph

Knowledge Graphs provides a high level overview of the association between diseases/symptoms, biological targets like genes, proteins, protein complexes, nucleic acids and chemical entities like clinical drugs and compounds. These associations are usually extracted from the structure data, but it's also common to use Nlp to extract important entities from the Text data. Entity recognition can help to extract relevant subject – predicate – object triplets from the text. One of the most applied method of graph machine learning for Covid-19 is purposing the problem in a form of knowledge graph completion.

One of the most notable knowledge graphs for drug repurposing for COVID-19 is done by BenevolentAI. [48] They used vast repository of structured medical data and also included numerous connections extracted from scientific literature using NLP. After building the Knowledge graph; A

subgraph relating to coronavirus extracted and checked by the experts. Experts analyzed this knowledge graph and revealed that the virus binds the host cells via the ACE2 receptor expressed on the surface of lung AT2 alveolar epithelial cells[40]. This work doesn't use any graph machine learning but shows the importance of knowledge graphs as a powerful way to represent input. Another prominent work done to build the Knowledge graph for Covid-19 is by Xiangxiang Zeng et al. where they built a comprehensive knowledge graph that includes 15 million edges across 39 types of relationships connecting drugs, diseases, proteins/genes, pathways, and expression from a large scientific corpus of 24 million PubMed publications. Further they used the AWS resources to apply GML algorithm such as RotatE from DGL-KE to identified 41 repurposable drugs (including dexamethasone, indomethacin, niclosamide, and toremifene) which were high-confidence candidate drugs for COVID-19 and they further validated by transcriptomic and proteomics data in SARS-CoV-2 infected human cells and data from ongoing clinical trials [56]. Another paper by RuiZhang et al. used the RotatE with TransE, DistMult, ComplEx techniques to predict drug repurposing candidates but they also Created a new Knowledge graph based on literature-based discovery (LBD) approach from PubMed and other COVID-19-focused research literature. They relied on semantic triples which were extracted using SemRep from SemMedDB.These triplets were further filtered out by rule-based and using an accuracy classifier developed on a BERT variant named PubMedBERT. They integrate the literature-based discovery and knowledge graph completion together [57].

AWS has generated a similar biological knowledge graph which extends HetioNet, called DRKG (Drug relation knowledge graph), to fight COVID-19. It included information from different databases as described in Table 2. and also data collected from recent publications particularly related to COVID-19, containing nearly 6 million edges between 100 thousand entities of 13 entity types[28]. Further work done to Apply Few-shot link prediction task using Graph machine learning on DRKG by Vassilis et al. where they used an inductive Relational GCN model [46] for learning informative relation embeddings. Their results corroborate that several drugs used in clinical trials were identified as possible drug candidates[29]. Kanglin Hsieh et al. built a COVID-19 knowledge representation from curated COVID-19 literature and on top of a comprehensive biomedical knowledge graph such as DRKG. They further derived the node's embedding of this graph using the multi-relational and variational graph autoencoder. They take the drug's embedding as features and build a drug ranking model where clinical trials are taken as silver-standard labels. There drug ranking candidates were validated using drug's gene profiles, in vitro drug screening efficacy, and large-scale electronic health records. Using these highly ranked drug candidates, they have searched for drug combinations that satisfy complementary exposure patterns[26].

Other open source COVID-19 knowledge graphs include the extension of ROBOKOP to COVID-KOP by Daniel Korn and team; This ROBOKOP biomedical knowledge graph is enriched with information from recent biomedical literature on COVID-19 annotated data. Sentence-by-sentence co-occurrence analysis added 800K new edges to the COVID-KOP graph, and co-occurrence counts at the paper level led to 4.5 million new edges. Gene ontology data for viral proteins and symptom data was also added to this Knowledge graph[36]. Another Knowledge graph created and open sourced by researchers from Germany called CovidGraph, which was built using knowledge from COVID-19 papers, case statistics, genes and functions, and molecular data [22]. In summary, recent efforts in knowledge graph construction and literature-based discovery illustrate the immense amount of research already performed on COVID-19. As more research and information about the disease was made public, the Knowledge Graph became larger and richer.

6.2 Generative models

A most advanced Graph generative Approach for drug repurposing is to use Variational graph autoencoder. VGAE is used for Link prediction tasks as well as designing new drugs or modifying already existing drugs. For this Later Application; A method purposed by researchers from IBM research, where they leveraged the protein-molecule binding affinity predictor that is pre-trained using SMILES VAE embeddings and protein sequence embeddings learned unsupervised from a large corpus. They applied this framework to three SARS-CoV-2 target proteins: main protease, receptor-binding domain of the spike protein, and non-structural protein 9 replicase[13]. Another effort of generative models was by Sumanta Ray et al. to use the GML on a combination of three different networks SARS-CoV-2-host PPI, human PPI, and drug-target network. It first created the integrated network by mapping the common interactions of three networks; then apply the Node2Vec to converts the network into fixed-size, low dimensional representations that preserve the properties of the nodes belonging to the three major components of the integrated network. Then used the learned node embeddings as well as the adjacency matrix from the integrated network to train a VGAE model, the decoding part of the VGAE, which re-raises the network based on the latent representation of the network provided by the encoder. This re-raising helps the network to find the edges between nodes that, although not having been explicit before, are imperative to exist relative to the encoded version of the network. Thereby, predicting links between drugs and SARS-CoV-2-associated human proteins, in particular[45].

6.3 Combination repurposing

Combination therapies provide an additional way to extend the indications and efficacy of available entities. The number of potential pairwise combinations of just two drugs makes a brute force empirical laboratory testing approach a lengthy and daunting prospect. To give a rough number, there exist around 4000 approved drugs which would require approx. 8 million experiments to test all possible combinations of two drugs at a single dose. Moreover, A Cocktails of drugs can be more effective than single drugs, because they can potentially work at lower doses and avoid resistance[59].

Some of the first work using GML to model combination therapy was DECAGON by Zitnik et al. [60] used to model polypharmacy side-effects via a multi-modal graph capturing drug–side effect–drug triplets in addition to PPI interactions. Deac et al. [15] forwent incorporation of a knowledge graph, instead modelling drug structures directly and using

Table 3: Text Mining Covid-19 Knowledge Graph

Covid KG	Data	Affiliation	Link	Comment
\CovidGraph[22]	CORD-19, Lens, Ensembl, NCBI Gene, Gene Ontology, experimental data	academic and industry org	https://covidgraph.org/	A knowledge graph of COVID-19 papers, case statistics, genes and functions, and molecular data
\BlenderLab COVID-KG[55]	CORD-19	UIUC	http://blender.cs.illinois.edu/	Knowledge graph with entity types genes, diseases, chemicals and organisms and subtypes derived from the text and figure/caption relations in literature
\COVID-KOP[36]	ROBOKOP, GO annotations, SciBite CORD-19 annotations	UNC Chapel Hill	https://covidkop.renci.org/	Combines ROBOKOP biomedical knowledge graph with information extracted from SciBite CORD-19 annotations

a coattention mechanism to achieve a similar level of accuracy. Wengong Jin et al. [31] proposed the idea of drug combinations for treating COVID-19. there proposed method ComboNet is composed of two networks a Drug target interaction [DTI] and a target disease association network. The antiviral effect of a single drug is predicted from its representation. The vector representation also characterizes the DTI features for a drug. The antiviral effect of a combination is predicted from its concatenated representation, which is computed from the molecular representations of each individual drug. ComboNet is trained on drug combination synergy, single-drug antiviral activity, and DTI data. two of the recommendation results from their paper were drug combinations of remdesivir + reserpine and remdesivir + IQ-1S. Discovering single agent therapies with activity against severe acute respiratory syndrome SARS-CoV-2 has been challenging. So that's why combination therapies play an important role in antiviral therapies, due to their improved efficacy and reduced toxicity.

7. CHALLENGES

Drug repurposing does not always succeed as there are instances where a selected drug candidate fails. for example, Ceftriaxone[14] drug was originally purposed as a Antibiotic medicine but later purposed as a drug for Amyotrophic lateral sclerosis, but it failed to be repurposed because of efficacy issues during Phase 3 Trials. Other than efficacy failures, there are challenges in Patent, Regulatory and organizational. For instance, there are number of legal and intellectual right issues related to drug repurposing. As these are important with respect to future profits from the repurposed product[5]. It is possible to apply for the Patent for the repurposed drug, provided the new medical use is new and inventive. However, many of the potential uses are already known in literature but didn't get reused as it never went through clinical trials for that particular use case. Method-of-use (MOU) patents can be taken for the new repurposed generic drug. however, generic drug can be sold by multiple manufacturers and prescribed by clinicians for other non-patented usecases[44]. So in this case the enforcement of the patent would be difficult, and the monetary Return from finding the repurposed drug would be small for the companies.

In regulatory challenges, the FDA only offers a period of 3 years of data exclusivity for new use of an old drug But this 3 year time period is too short to recover the money a company has invested in repurposing a particular drug. These intellectual rights and regulatory challenges bolster by Organizational challenges, particularly if the repurposed indication is not within the organization's core disease area[44]. There are also lack of resources and funds to progress the potential drug repurposing project. From Alternative funding to government support and favourable property rights can help to tackle above issues and push the pharmaceutical companies toward drug repurposing.

8. CONCLUSIONS

We have discussed how Graph Machine Learning helps to solve the drug repurposing problem for Covid-19 disease. With its ability to do representations of unstructured multimodal datasets, one can expect to see tremendous advances being made within data integration. Although there are still problems like Oversmoothing problem for deeper Gnn models or issue with parallel computation for larger graphs. As GML is still in its infancy and lots of new research is coming out in this field. It is very exciting to see how GML will progress in coming years and become a de-facto tool for the Drug Repurposing Task.

9. ACKNOWLEDGMENTS

I would like to thanks *Prof. Erdem Sariyuce* for his continuous support while writing this survey Paper.

10. REFERENCES

[1] Cell division and anatomy. https://courses.lumenlearning.com/boundless-ap/chapter/cell-division.

[2] Revenue of pfizer from viagra sales. https://www.statista.com/statistics/264827/pfizers-worldwide-viagra-revenue-since-2003/.

[3] Who article on drug type. https://www.who.int/news-room/feature-stories/detail/

the-race-for-a-covid-19-vaccine-explained. Accessed: 2022-05-08.

[4] R. Apweiler, A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, and L.-S. L. Yeh. Uniprot: the universal protein knowledgebase. *Nucleic acids research*, 32 Database issue:D115–9, 2004.

[5] T. T. Ashburn and K. B. Thor. Drug repositioning: identifying and developing new uses for existing drugs. *Nature reviews Drug discovery*, 3(8):673–683, 2004.

[6] H. Ashoor, X. Chen, W. Rosikiewicz, J. Wang, A. Cheng, P. Wang, Y. Ruan, and S. Li. Graph embedding and unsupervised learning predict genomic sub-compartments from hic chromatin interaction data. *Nature communications*, 11(1):1–11, 2020.

[7] A.-L. Barabási, N. Gulbahce, and J. Loscalzo. Network medicine: a network-based approach to human disease. *Nature reviews genetics*, 12(1):56–68, 2011.

[8] Y. Bengio, A. C. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:1798–1828, 2013.

[9] S. Bonner, I. P. Barrett, C. Ye, R. Swiers, O. Engkvist, A. Bender, C. T. Hoyt, and W. L. Hamilton. A review of biomedical datasets relating to drug discovery: A knowledge graph perspective, 2021.

[10] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013.

[11] A. Breit, S. Ott, A. Agibetov, and M. Samwald. Openbiolink: a benchmarking framework for large-scale biomedical link prediction. *Bioinformatics*, 2020.

[12] I. Chami, S. Abu-El-Haija, B. Perozzi, C. Ré, and K. Murphy. Machine learning on graphs: A model and comprehensive taxonomy. *arXiv preprint arXiv:2005.03675*, 2020.

[13] V. Chenthamarakshan, P. Das, S. Hoffman, H. Strobelt, I. Padhi, K. W. Lim, B. Hoover, M. Manica, J. Born, T. Laino, and A. Mojsilovic. Cogmol: Target-specific and selective drug design for covid-19 using deep generative models. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 4320–4332. Curran Associates, Inc., 2020.

[14] M. E. Cudkowicz, S. Titus, M. Kearney, H. Yu, A. Sherman, D. Schoenfeld, D. Hayden, A. Shui, B. Brooks, R. Conwit, et al. Safety and efficacy of ceftriaxone for amyotrophic lateral sclerosis: a multi-stage, randomised, double-blind, placebo-controlled trial. *The Lancet Neurology*, 13(11):1083–1091, 2014.

[15] A. Deac, Y.-H. Huang, P. Veličković, P. Liò, and J. Tang. Drug-drug adverse effect prediction with graph co-attention. *arXiv preprint arXiv:1905.00534*, 2019.

[16] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen. Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics*, 47:20–33, 2016.

[17] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. E. Gillespie, M. R. Kamdar, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C. Duenas, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. M. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio. The reactome pathway knowledgebase. *Nucleic Acids Research*, 42:D472 – D477, 2018.

[18] H. Fang, H. De Wolf, B. Knezevic, K. L. Burnham, J. Osgood, A. Sanniti, A. Lledó Lara, S. Kasela, S. De Cesco, J. K. Wegner, et al. A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nature genetics*, 51(7):1082–1091, 2019.

[19] E. N. Feinberg, E. Joshi, V. S. Pande, and A. C. Cheng. Improvement in admet prediction with multitask deep featurization. *Journal of medicinal chemistry*, 63(16):8835–8848, 2020.

[20] M. Fey and J. E. Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.

[21] T. Gaudelet, B. Day, A. R. Jamasb, J. Soman, C. Regep, G. Liu, J. B. Hayter, R. Vickers, C. Roberts, J. Tang, et al. Utilizing graph machine learning within drug discovery and development. *Briefings in bioinformatics*, 22(6):bbab159, 2021.

[22] L. Gütebier, R. Henkel, A. Jarasch, T. Bleimehl, S. Müller, J. Munro, M. Preusse, D. Walthemath, and T. HealthEcco. Covidgraph: Connecting biomedical covid-19 resources and computational biology models. In *2nd Workshop on Search, Exploration, and Analysis in Heterogeneous Datastores, SEA-Data 2021*, pages 34–37, 2021.

[23] J.-D. J. Han. Understanding biological functions through molecular networks. *Cell research*, 18(2):224–237, 2008.

[24] M. A. Harris, J. I. Deegan, A. Ireland, J. Lomax, M. Ashburner, S. Tweedie, S. Carbon, S. E. Lewis, C. J. Mungall, J. Richter, K. Eilbeck, J. A. Blake, C. J. Bult, A. D. Diehl, M. E. Dolan, H. J. Drabkin, J. T. Eppig, D. P. Hill, N. Li, M. Ringwald, R. Balakrishnan, G. Binkley, J. M. Cherry, K. R. Christie, M. C. Costanzo, Q. Dong, S. R. Engel, D. G. Fisk, J. E. Hirschman, B. C. Hitz, E. L. Hong, C. J. Krieger, S. R. Miyasato, R. S. Nash, J. Park, M. S. Skrzypek, S. Weng, E. D. Wong, K. K. Zhu, D. Botstein, K. Dolinski, M. S. Livstone, R. Oughtred, T. Z. Berardini, L. Donghui, S. Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, R. P. Huntley, N. J. Mulder, V. K. Khodiyar, R. C. Lovering, S. Povey, R. L. Chisholm, P. Fey, P. Gaudet, W. Kibbe, R. Kishore, E. M. Schwarz, P. W. Sternberg, K. V. Auken, M. G. Giglio, L. I. Hannick, J. R. Wortman, M. Aslett, M. Berriman, V. Wood, H. J. Jacob, S. J. F. Laulederkind, V. Petri, M. Shimoyama, J. L. Smith, S. N. Twigger, P. Jaiswal, T. E. Seigfried, D. G. Howe, M. Westerfield, C. W. Collmer, T. T. Alalibo, E. Feltrin, G. Valle, S. Bromberg, S. C. Burgess, and F. M. McCarthy. The gene ontology project in 2008. *Nucleic Acids Research*, 36:D440 – D444, 2008.

[25] D. S. Himmelstein, A. Lizee, C. Hessler, L. Brueggeman, S. L. Chen, D. Hadley, A. Green, P. Khankhanian, and S. E. Baranzini. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *Elife*, 6:e26726, 2017.

[26] K. Hsieh, Y. Wang, L. Chen, Z. Zhao, S. Savitz, X. Jiang, J. Tang, and Y. Kim. Drug repurposing for covid-19 using graph neural network and harmonizing multiple evidence. *Scientific reports*, 11(1):1–13, 2021.

[27] P. D. Hsu, E. S. Lander, and F. Zhang. Development and applications of crispr-cas9 for genome engineering. *Cell*, 157:1262–1278, 2014.

[28] V. N. Ioannidis, X. Song, S. Manchanda, M. Li, X. Pan, D. Zheng, X. Ning, X. Zeng, and G. Karypis. Drkg-drug repurposing knowledge graph for covid-19. *arXiv preprint arXiv: 2010.09600*, 2020.

[29] V. N. Ioannidis, D. Zheng, and G. Karypis. Few-shot link prediction via graph neural networks for covid-19 drug-repurposing. *arXiv preprint arXiv:2007.10261*, 2020.

[30] M. Jiang, Z. Li, S. Zhang, S. Wang, X. Wang, Q. Yuan, and Z. Wei. Drug–target affinity prediction using graph neural network and contact maps. *RSC Advances*, 10(35):20701–20712, 2020.

[31] W. Jin, J. M. Stokes, R. T. Eastman, Z. Itkin, A. V. Zakharov, J. J. Collins, T. S. Jaakkola, and R. Barzilay. Deep learning identifies synergistic drug combinations for treating covid-19. *Proceedings of the National Academy of Sciences*, 118(39), 2021.

[32] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

[33] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[34] T. N. Kipf and M. Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.

[35] J. Klicpera, J. Groß, and S. Günnemann. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020.

[36] D. Korn, T. Bobrowski, M. Li, Y. Kebede, P. Wang, P. Owen, G. Vaidya, E. Muratov, R. Chirkova, C. Bizon, et al. Integrating emerging covid-19 data with the robokop database. chemrxiv 2020.

[37] C. le Sage, S. Lawo, P. Panicker, T. M. Scales, S. A. Rahman, A. S. Little, N. J. McCarthy, J. D. Moore, and B. Cross. Dual direction crispr transcriptional regulation screening uncovers gene networks driving drug resistance. *Scientific reports*, 7(1):1–10, 2017.

[38] L. Mabonga and A. P. Kappo. Protein-protein interaction modulators: advances, successes and remaining challenges. *Biophysical Reviews*, 11:559 – 581, 2019.

[39] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[40] E. N. Muratov, R. Amaro, C. H. Andrade, N. Brown, S. Ekins, D. Fourches, O. Isayev, D. Kozakov, J. L. Medina-Franco, K. M. Merz, et al. A critical overview of computational approaches employed for covid-19 drug discovery. *Chemical Society Reviews*, 2021.

[41] A. C. Nascimento, R. B. Prudêncio, and I. G. Costa. A multiple kernel learning algorithm for drug-target interaction prediction. *BMC bioinformatics*, 17(1):1–16, 2016.

[42] S. Pittala, W. Koehler, J. Deans, D. Salinas, M. Bringmann, K. S. Volz, and B. Kapicioglu. Relation-weighted link prediction for disease gene identification. *arXiv preprint arXiv:2011.05138*, 2020.

[43] N. Pržulj, D. G. Corneil, and I. Jurisica. Modeling interactome: scale-free or geometric? *Bioinformatics*, 20(18):3508–3515, 2004.

[44] S. Pushpakom, F. Iorio, P. A. Eyers, K. J. Escott, S. Hopper, A. Wells, A. Doig, T. Guilliams, J. Latimer, C. McNamee, et al. Drug repurposing: progress, challenges and recommendations. *Nature reviews Drug discovery*, 18(1):41–58, 2019.

[45] S. Ray, S. Lall, A. Mukhopadhyay, S. Bandyopadhyay, and A. Schönhuth. Predicting potential drug targets and repurposable drugs for covid-19 via a deep generative model for graphs. *arXiv preprint arXiv:2007.02338*, 2020.

[46] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. v. d. Berg, I. Titov, and M. Welling. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer, 2018.

[47] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018.

[48] J. Stebbing, A. Phelan, I. Griffin, C. Tucker, O. Oechsle, D. Smith, and P. Richardson. Covid-19: combining antiviral and anti-inflammatory treatments. *The Lancet Infectious Diseases*, 20(4):400–402, 2020.

[49] U. Storz. Rituximab: how approval history is reflected by a corresponding patent filing strategy. In *MAbs*, volume 6, pages 820–837. Taylor & Francis, 2014.

[50] Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*, 2019.

[51] W. Torng and R. B. Altman. Graph convolutional neural networks for predicting drug-target interactions. *Journal of chemical information and modeling*, 59(10):4131–4149, 2019.

[52] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[53] B. Walsh, S. K. Mohamed, and V. Nováček. Biokg: A knowledge graph for relational learning on biological data. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 3173–3180, 2020.

[54] L. L. Wang, K. Lo, Y. Chandrasekhar, R. Reas, J. Yang, D. Eide, K. Funk, R. Kinney, Z. Liu, W. Merrill, et al. Cord-19: The covid-19 open research dataset. *ArXiv*, 2020.

[55] Q. Wang, M. Li, X. Wang, N. Parulian, G. Han, J. Ma, J. Tu, Y. Lin, H. Zhang, W. Liu, et al. Covid-19 literature knowledge graph construction and drug repurposing report generation. *arXiv preprint*

arXiv:2007.00576, 2020.

[56] X. Zeng, X. Song, T. Ma, X. Pan, Y. Zhou, Y. Hou, Z. Zhang, K. Li, G. Karypis, and F. Cheng. Repurpose open data to discover therapeutics for covid-19 using deep learning. *Journal of proteome research*, 19(11):4624–4636, 2020.

[57] R. Zhang, D. Hristovski, D. Schutte, A. Kastrin, M. Fiszman, and H. Kilicoglu. Drug repurposing for covid-19 via knowledge graph completion. *Journal of biomedical informatics*, 115:103696, 2021.

[58] S. Zheng, J. Rao, Y. Song, J. Zhang, X. Xiao, E. F. Fang, Y. Yang, and Z. Niu. Pharmkg: a dedicated knowledge graph benchmark for bomedical data mining. *Briefings in bioinformatics*, 22(4):bbaa344, 2021.

[59] A. Zimmer, A. Tendler, I. Katzir, A. Mayo, and U. Alon. Prediction of drug cocktail effects when the number of measurements is limited. *PLoS biology*, 15(10):e2002518, 2017.

[60] M. Zitnik, M. Agrawal, and J. Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466, 2018.