Research paper

Deep variational graph autoencoders for novel host-directed therapy options against COVID-19

Sumanta Ray [a,b], Snehalika Lall [e], Anirban Mukhopadhyay [d], Sanghamitra Bandyopadhyay [e], Alexander Schönhuth [c,*]

[a] *Department of Computer Science and Engineering, Aliah University, New Town, Kolkata, India*
[b] *Health Analytics Network, PA, USA*
[c] *Genome Data Science, University of Bielefeld, Bielefeld, Germany*
[d] *Department of Computer Science and Engineering, University of Kalyani, Kalyani, India*
[e] *Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India*

ARTICLE INFO

ABSTRACT

The COVID-19 pandemic has been keeping asking urgent questions with respect to therapeutic options. Existing drugs that can be repurposed promise rapid implementation in practice because of their prior approval. Conceivably, there is still room for substantial improvement, because most advanced artificial intelligence techniques for screening drug repositories have not been exploited so far. We construct a comprehensive network by combining year-long curated drug–protein/protein–protein interaction data on the one hand, and most recent SARS-CoV-2 protein interaction data on the other hand. We learn the structure of the resulting encompassing molecular interaction network and predict missing links using variational graph autoencoders (VGAEs), as a most advanced deep learning technique that has not been explored so far. We focus on hitherto unknown links between drugs and human proteins that play key roles in the replication cycle of SARS-CoV-2. Thereby, we establish novel host-directed therapy (HDT) options whose utmost plausibility is confirmed by realistic simulations. As a consequence, many of the predicted links are likely to be crucial for the virus to thrive on the one hand, and can be targeted with existing drugs on the other hand.

1. Introduction

The ongoing pandemic of COVID-19 (Coronavirus Disease-2019), caused by SARS-CoV-2, an enveloped, single-stranded RNA virus [1], has led to more than a million deaths so far and keeps asking urgent questions. Accepting the challenge, researchers have been relentlessly searching for possible therapeutic strategies in the last few months. However, still, no truly reliable remedy has been showing on the horizon.

Repurposing drugs refers to screening databases for molecules whose risks have been found to be manageable in prior applications on the one hand, and that can be shown to target proteins that are crucial for SARS-CoV-2 to replicate and thrive on the other hand. If not even representing viable cures by themselves, repurposed drugs have the potential to mitigate the severity of the pandemic for the time being. The fact that the majority of 3D structures of the SARS-CoV-2 proteins has remained unknown so far, corroborates the need for artificial intelligence based screens of molecular interaction data that relate with COVID-19 further, because experimental, de novo drug

design crucially depends on the availability of such 3D structures. Note further that because the virus hijacks the host cell machinery for replication through interactions of viral with human proteins, a comprehensive understanding of the interactions between viral and human proteins is essential [2,3].

In that quest for therapy options, the most advanced artificial intelligence based approaches may mean a massive boost with respect to drug repurposing screens. However, despite their promising potential, many state-of-the-art AI – in particular deep neural network – based approaches have not yet been explored so far.

Here, we do exactly this. We combine existing, year-long curated and approved molecular (drug/human protein) interaction data with most recent experimental interaction screens (yielding new, and so far only insufficiently explored SARS-CoV-2–human protein interaction data). In this, we design an experimental setup that enables us to exploit most advanced (and hitherto unexplored) deep variational graph autoencoder techniques for generating novel therapy options.

---

* Corresponding author.
E-mail addresses: sumanta.ray@aliah.ac.in (S. Ray), alexander.schoenhuth@uni-bielefeld.de (A. Schönhuth).

In detail, we first learn the structure of the comprehensive drug–human–virus molecule interaction network while encoding the network. We then predict links between human proteins that are crucial for the virus to replicate on the one hand, and existing drugs as a result of decoding the autoencoder representation of our network, although these links were not part of the original network. As a result, we do not only predict drugs to act against COVID-19, but we also identify the human proteins that when blocked lead to disruption of the viral replication cycle, which fosters the biomolecular understanding of how viral replication can be controlled.

Note that we suggest to block human proteins such that the replication machinery of the virus falls apart. However, we do not suggest to target viral proteins themselves. The justification is that viral proteins, when being targeted, tend to elicit resistance-inducing mutations, such that the virus rapidly adapts to the (rather simpleminded) attack. In comparison to viral proteins, human proteins acquire mutations at rates that are smaller by orders of magnitudes. This renders human proteins substantially more sustainable therapy options when establishing actionable drug targets [4]. This explains why we focus on the corresponding host-directed therapy (HDT) options here. In summary, we suggest drugs that have the potential to be rapidly integrated in clinical practice (thanks to repurposing) and that the virus cannot easily escape (thanks to serving HDT based strategies).

Our combination of drug repurposing and HDT based on screening molecular interaction data is further supported by prior work that has been describing unprecedented opportunities lately [5]. Examples of pathogens that were treated earlier are Dengue [6], HIV [7], Ebola [8], next to various other, non-viral diseases.

As for related work, a handful of research groups have been trying to suggest drugs to be repurposed so as to counteract the spread of SARS-CoV-2 in the human body based on exploiting network resources since the outbreak of COVID-19. Zhou et al. made the first attempt through an integrative network analysis [9], followed by Li et al. who combined network data with a comparative analysis on the gene sequences of different viruses [10].

Only shortly thereafter, however, Gordon et al. generated a map that juxtaposes SARS-CoV-2 proteins with human proteins based on affinity-purification mass spectrometry (AP-MS) screens in pioneering work [11], closely followed by Dick et al. who, in independent work, identified high confidence interactions between human and SARS-CoV-2 proteins using sequence-based PPI predictors (a.k.a. PIPE4 & SPRINT) [12]. Recently, Sadegh et al. developed CoVex to visually explore the SARS-CoV-2 host interactome and repurposable drugs in an online interactive platform [13]. Giulia et al. [14], proposed a network similarity based approach to prioritize drug molecules associated with COVID-19. Gysi et al. [15] integrate artificial intelligence, network diffusion, and network proximity measure to rank the existing drugs for their expected efficacy against SARS-CoV-2.

Both of the studies [11,12] provide crucial data, because only because of the two studies we are able to link existing (long term curated and highly reliable) drug–protein and human protein–protein interaction data with the SARS-CoV-2 proteins, just as was possible for the above-mentioned diseases earlier [6,7,16,17].

Still, however, the exploitation of the novel data, in combination with year-long established, refined and curated interaction data using most advanced AI techniques needed to be brought into effect. As a brief summary of our contributions:

(1) We link existing high-quality, long-term curated and refined, large scale drug/protein–protein interaction data with

(2) molecular interaction data on SARS-CoV-2 itself, raised recently in literature,

(3) exploit the resulting overarching network using an advanced AI supported techniques (namely variational graph autoencoder based techniques)

(4) for repurposing drugs in the fight against SARS-CoV-2

(5) in the frame of HDT based strategies.

As for (3)–(5), we will demonstrate how to adapt most advanced deep learning based techniques to learn and exploit molecular interaction network data. By this, we are able to predict new links between drugs and proteins at utmost accuracy. The spectrum of drugs we reveal is fairly broad in terms of mechanism of action. We are therefore convinced that several drugs we suggest have solid potential to be amenable to developing successful HDTs against COVID-19.

2. Materials and methods

In the following, we will first describe the workflow of our analysis pipeline and the basic ideas that support it.

First, we raise a novel network by combining well-established and most recent resources into an overarching, comprehensive interaction network that puts drugs, human and SARS-CoV-2 proteins into encompassing context.

We then carry out a simulation study that proves that our AI supported pipeline predicts missing links in the encompassing drug–human protein–SARS-CoV-2–protein network at utmost accuracy. With this, we provide evidence for our predictions to reflect true interactions between molecular interfaces, at utmost likelihood.

Subsequently, in our real experiments, we predict links between drugs on the one hand, and SARS-Cov-2-associated human proteins on the other hand to be missing. Corroborated by our simulations, a large fraction (if not possibly even the vast majority) of predictions establish true molecular interactions, potentially actionable in HDT based strategies.

Finally, we inspect the postulated mechanism of action of the suggested drugs in the frame of several diseases, including the closely related SARS-CoV ("SARS-classic") and MERS-CoV, documenting the plausibility of our predictions.

2.1. Workflow

See Fig. 1 for the workflow of our analysis pipeline. We will describe all important steps in the paragraphs of this subsection.

2.1.1. Raising a comprehensive interaction network

See A & B in Fig. 1. We have combined well-established drug–gene interaction and human interactome data (compiled from eight, year-long curated, much refined, well-established publicly accessible resources) with the SARS-CoV-2–human protein-interaction network published only a few weeks ago. The integrated network has four types of nodes:

(1) SARS-CoV-2 proteins, **(2)** SARS-CoV-2-associated host proteins (CoV-host), **(3)** human proteins other than (2) and **(4)** drugs. This means that we put drugs, human proteins and SARS-CoV-2 proteins into a context that is as comprehensive as currently possible. Still, it is highly likely, however, that links are missing. Because many such missing links reflect drugs that can be repurposed, it remains to set up an AI approach that can predict such links.

2.1.2. AI model first stage — Node2Vec

See C in Fig. 1. For the link prediction machinery to work, we operate in two stages. First, we employ Node2Vec [18], as a network embedding strategy that extracts node features from the integrated network. Formally, Node2Vec converts the adjacency matrix that represents the network into a fixed-size, low-dimensional latent feature space. As elements of this space, nodes correspond to feature vectors. Thereby, Node2Vec aims at preserving the properties of the nodes relative to their surroundings in the network. For efficiency reasons, Node2Vec makes use of a sampling strategy. The result of this step is a feature matrix (F) where rows refer to nodes and columns refer to the inferred network features (See supplementary text for detailed descriptions).
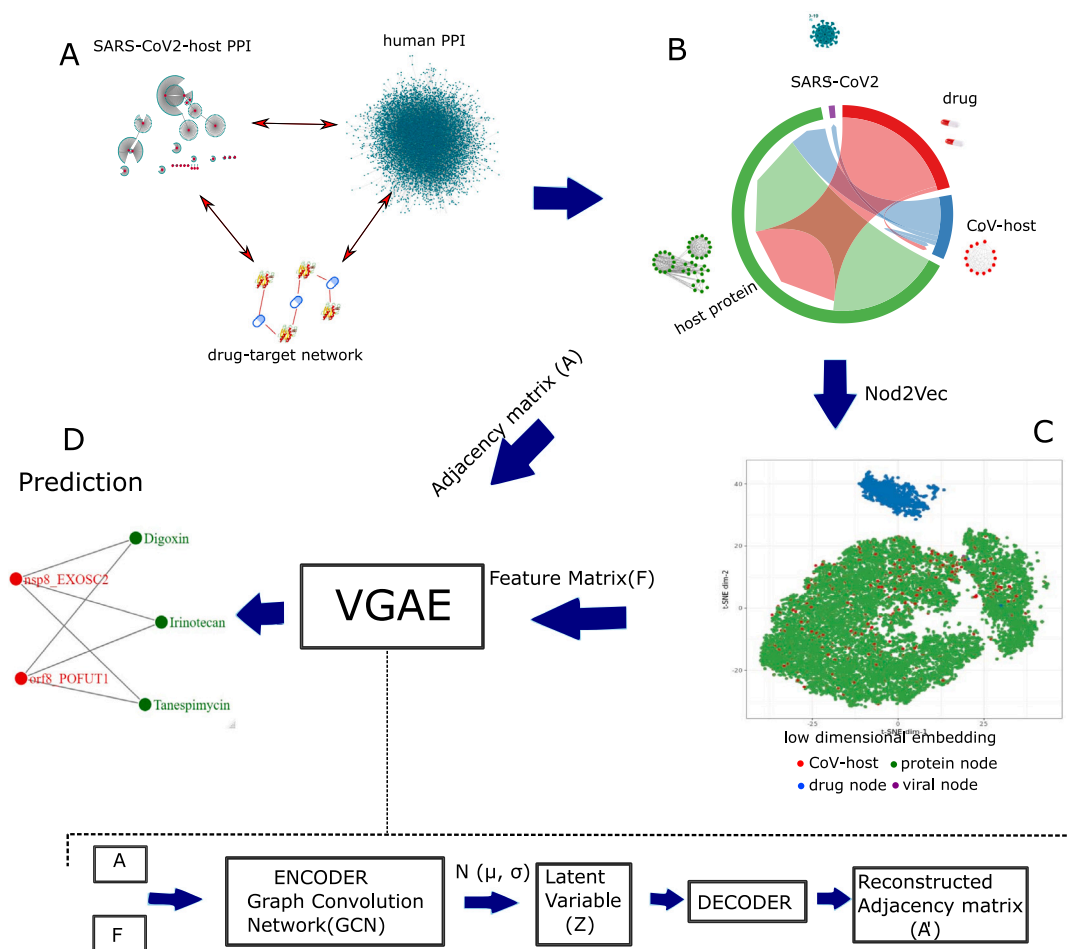
Fig. 1. Overall workflow of the proposed method: The three networks SARS-CoV-2–host PPI, human PPI, and drug–target network (Panel-A) are mapped by their common interactors to form an integrated representation (Panel-B). The neighborhood sampling strategy Node2Vec converts the network into fixed-size low dimensional representations that perverse the properties of the nodes belonging to the three major components of the integrated network (Panel-C). The resulting feature matrix (F) from the node embeddings and adjacency matrix (A) from the integrated network are used to train a VGAE model, which is then used for prediction (Panel-D).

2.1.3. AI model second stage — variational graph autoencoders (VGAE)

See B, C & D in Fig. 1. In the next step, we employ variational graph autoencoders (VGAE), as a most recent graph neural network based technique that was shown to predict missing links in networks at utmost accuracy [19]. VGAEs require the original graph (coded as its adjacency matrix A) and, optionally, a feature matrix F that annotates the nodes of the network with helpful additional information. Often, F does not necessarily refer to the topology of the network itself. Here, however, we do make use of the feature matrix F that was inferred from A by Node2Vec. We found that using F aided in raising prediction accuracy substantially, despite F only being an alternative representation of A. The explanation for this is that F consists of knowledge obtained using Node2Vec, which, as being complementary to VGAEs, indeed reveals additional information. Our pipeline thus unifies the virtues of both VGAE and Node2Vec. See Section 3-C and Fig. 2 for corresponding experiments.

2.1.4. Predicting missing links

See D in Fig. 1. After training the VGAE, we predict links in the encompassing drug–human–virus interaction network that had remained to be missing. For this, we make use of the decoding part of the VGAE, which re-raises the interaction network based on the latent representation the encoder had computed. Re-raising the network results in edges between nodes that although not having been explicit before, are imperative to exist relative to the encoded version of the network. Thereby, one predicts links between drugs and SARS-CoV-2-associated human proteins in particular. Although not having been

Table 1
Average AUC and AP across the last 10 training epochs of FastGAE. Validation AUC and AP for different numbers N_S of sampling nodes are reported.

N_s	Average performance on validation set		
	AUC (%)	AP (%)	Training time (in sec)
7000	89.21 ± 0.02	85.32 ± 0.02	1587
5000	89.17 ± 0.03	85.30 ± 0.04	1259
3000	88.91 ± 0.10	85.02 ± 0.04	1026
2500	88.27 ± 0.15	84.88 ± 0.13	998
1000	86.69 ± 0.17	83.58 ± 0.19	816

explicit before, the existence is implied by the topological constraints the comprehensive network imposes on such links to exist or not. Our model thus predicts both drugs and proteins: repurposing these drugs leads to targeting the matching proteins. See Fig. 1 for the total workflow we just described.

2.1.5. Addressing computation time

The biomolecular interaction networks one needs to consider for successful drug repurposing, i.e. standard protein–protein and drug–protein interaction networks, consist of hundreds of thousands of nodes, and are too large for standard implementations of VGAEs to deal with. This renders advanced, runtime friendly implementations of VGAEs crucial ingredients of our workflow. Most recent progress on that topic by Salha et al. (published Feb 5, 2020), running under the name FastGAE [20] provides the last key element for our approach to work in
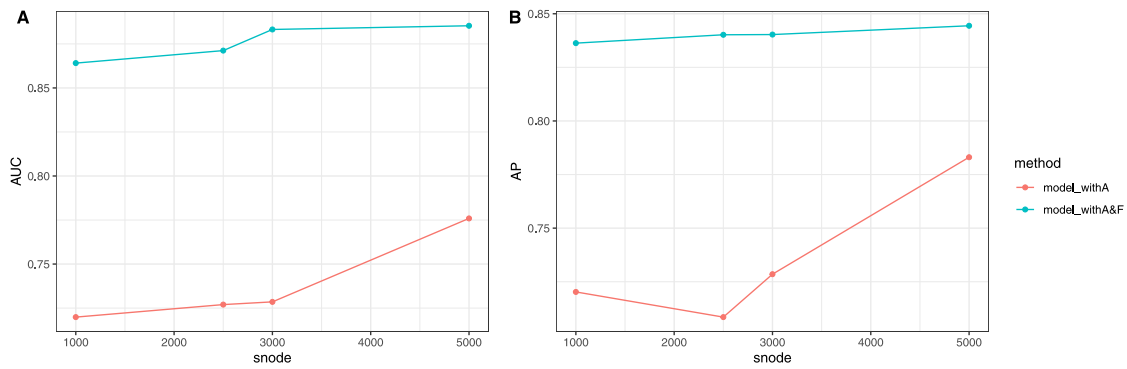
Fig. 2. Performance of the model (AUC on the validation set) with and without using feature matrix (F).

practice. FastGAE relies on a strategy by which to repeatedly subsample nodes from large graphs, and train VGAEs on the resulting subgraphs, and subsequently to join the resulting autoencoders in a consistent manner. In experiments (see Table 1), we determined 5000 nodes as an optimal size for a subsample.

2.2. Sampling strategy and feature matrix generation

We have utilized *Node2vec* [18], an algorithmic framework for learning continuous feature representations for nodes in networks. It maps the nodes to a low-dimensional feature space that maximizes the likelihood of preserving network neighborhoods.

The principle of feature learning framework in a graph can be described as follows: Let $G = (V, E)$ be a graph, where V represents a set of nodes, and E represents the set of edges. The feature representation of nodes ($|V|$) is given by a mapping function: $f : V \rightarrow R^d$, where d specifies the feature dimension. Alternatively, f may be considered as a node feature matrix of a dimension of $|V| \times d$. For each node, $v \in V$, a network neighborhood $NN_S(v) \subset V$ of node v is defined by employing a neighborhood sampling strategy S. The sampling strategy can be sketched as an interpolation between breadth-first search and depth-first search [18], with objective function

$$\max_f \left(\sum_{v \in V} \log P(NN_S(v) \mid f(v)) \right) \tag{1}$$

This maximizes the likelihood of observing a network neighborhood $NN_S(v)$ for a node v given its feature representation $f(v)$. The probability of observing a neighborhood node $n_i \in NN_S(v)$ given $f(v)$ is

$$P(NN_S(v) \mid f(v)) = \prod_{n_i \in NN_S(v)} P(n_i \mid f(v)). \tag{2}$$

where n_i refers to the ith neighbor of node v as part of $NN_S(v)$. Last, the conditional probability $P(n_i \mid f(v))$ of a neighborhood node $n_i \in NN_S(V)$ given the original node v is computed as the softmax the scalar product of their feature vectors $f(v)$ and $f(n_i)$

$$P(n_i \mid f(v)) = \frac{exp(f(v) \cdot f(n_i))}{\sum_{u \in V} exp(f(u) \cdot f(v))} \tag{3}$$

2.3. Drug–SARS-CoV-2 link prediction

2.3.1. Adjacency matrix preparation

In this work, we consider an undirected graph $G = (V, E)$ with $|V| = n$ nodes and $|E| = m$ edges. Let A be the binary adjacency matrix of G. Here V consists of SARS-Cov-2 proteins, CoV-host proteins, drug–target proteins and drugs. The matrix (A) contains a total of $n = 16444$ nodes given as:

$$n = |N_{Nc}| + |N_{DT}| + |N_{NT}| + |N_D|, \tag{4}$$

where, N_{Nc} is the number of SARS-CoV-2 proteins. N_{DT} is the number of drug targets, whereas N_{NT} and N_D represent the number of CoV-host and drugs nodes, respectively. The total number of edges is given by:

$$m = |E_1| + |E_2| + |E_3|, \tag{5}$$

where, E_1 represents interactions between SARS-CoV-2 and human host proteins, E_2 is the number of interactions among human proteins, and E_3 represents the number of interactions between drugs and human host proteins.

2.3.2. Feature matrix preparation

The neighborhood sampling strategy is used to compute feature representations for all nodes. A flexible biased random walk procedure is employed to explore the neighborhood of each node. A random walk in a graph G can be described as the probability

$$P(a_i = x \mid a_{i-1} = v) = \pi(v, x), \tag{6}$$

where, $\pi(v, x)$ is the transition probability between nodes v and x, where $(v, x) \in E$ and a_i is the ith node in the walk of length l. The transition probability is given by $\pi(v, x) = c_{pq}(t, x) \times w_{vx}$, where t is the previous node of v in the walk, w_{vx} is the (static) weight attached to the edge (v, x) and p, q are the two parameters that guide the walk. The coefficient $c_{pq}(t, x)$ is given by

$$c_{pq}(t, x) = \begin{cases} 1/p & distance(t, x) = 0 \\ 1 & distance(t, x) = 1 \\ 1/q & distance(t, x) = 2 \end{cases} \tag{7}$$

where $distance(t, x)$ represents the distance of the shortest path between nodes t and node x. The process of feature matrix $F_{n \times d}$ generation is governed by the Node2vec algorithm. It starts from every node, simulating r random walks of fixed length l. In every step of a walk the transition probabilities $\pi(v, x)$ govern the sampling. In each iteration, generated walks are added to a list of walks. Each random walk forms a sentence which is ultimately used by word2vec [21], a well-known algorithm that takes a set of sentences (walks), and outputs an embedding for each word. The log-likelihood in Eq. (1) is optimized in the Optimization step by using stochastic gradient descent algorithm on a two-layer Skip-gram neural network model used by word2vec.

2.3.3. Link prediction

We utilize scalable and fast variational graph autoencoder (FastV-GAE) [20] to reduce the computational time of VGAE in networks that are as large as ours. The adjacency matrix A and the feature matrix F are fed into the encoder of FastVGAE as input. The encoder uses a graph convolution neural network (GCN) on the entire graph to create the latent representation

$$Z = GCN(A, F) \tag{8}$$

The encoder works on the full adjacency matrix A. After encoding, one samples subgraphs, and decoding is performed on the sampled subgraphs.

The mechanism of the decoder of FastVGAE slightly differs from that of a traditional VGAE. For each subsample of graph nodes V_s, it regenerates an adjacency matrix $\hat{A}$. For subsampling graph nodes, it makes use of a technique that determines the nodes from which to reconstruct the adjacency matrix in each iteration. Therefore, each node is assigned with a probability

$$p(i) = \frac{d(i)^\alpha}{\sum_{j \in V} d(j)^\alpha} \tag{9}$$

where $d(i)$ is the degree of node i, and α is the sharpening parameter, where in our study $\alpha = 2$. Nodes are then selected during subsampling according to their probabilities p_i until the subsampled nodes amount to $|V_s| = n_s$, the prescribed number of sampling nodes.

The decoder reconstructs the smaller matrix, $\hat{A}_s$ of dimension $n_s \times n_s$ instead of decoding the main adjacency matrix A. The decoder function follows the following equation:

$$\hat{A}_s(i,j) = \text{Sigmoid}(z_i^T \cdot z_j), \ \forall (i,j) \in V_s \times V_s \tag{10}$$

where z_i, z_j reflect the representations of nodes i, j, as computed by the encoder, see (8). At each training iteration a different subgraph (G_s) is drawn using the sampling method.

After training the model, the drug–CoV-host links are predicted using the equation

$$p(A_{ij} = 1 \mid z_i, z_j) = \text{Sigmoid}(z_i^T z_j), \tag{11}$$

where $A_{ij} = 1$ reflects a link between nodes i and j to exist, where i and j further reflect human proteins that interact with SARS-CoV-2 on the one hand and drugs on the other hand (recalling that we would like to predict links between human proteins that when targeted lead to the replication machinery of the virus falling apart). For each such combination of nodes the model computes the probability based on the logistic sigmoid function.

3. Formal details and background

3.1. Variational graph autoencoder

The Variational Graph Autoencoder (VGAE) is a framework for unsupervised learning on graph-structured data [19]. This model uses latent variables and is effective in learning interpretable latent representations for undirected graphs. The VGAE consists of two subnetworks that are stacked onto another: (1) Encoder and (2) Decoder. First, a graph convolution networks (GCN) based encoder [19] maps the nodes into a low-dimensional embedding space. Subsequently, a decoder attempts to reconstruct the original graph structure from the encoder representations. Both models are jointly trained to optimize the quality of the reconstruction from the embedding space, in an unsupervised way. The encoder and the decoder are described in the following.

Encoder. The encoder consists of a Graph Convolution Network (GCN) that takes the adjacency matrix A and the feature representation matrix F as input. The encoder generates a d'-dimensional latent variable z_i for each node $i \in V$, with $|V| = n$, where $d' \leq n$. Let $Z = (z_i)$ represent all such latent variables. The probability to generate a particular choice of Z is given by the formula

$$q(Z \mid A, F) = \prod_{i=1}^{|v|} q(z_i \mid A, F), \tag{12}$$

which assumes conditional independence between the z_i given the adjacency matrix A (that is, the graph) and the Node2Vec based feature representation F of the graph. The probability $r(z_i|A, F)$ follows a normal distribution, $\mathcal{N}(z_i|\mu_i, \text{diag}(\sigma_i^2))$ where μ_i and $\text{diag}(\sigma_i^2)$ parameterize

the d'-dimensional distribution $\mathcal{N}$. Both μ_i and σ_i reflect the output of two graph convolutional networks (GCN) that share parameters in the first layer. For further details, see [19]. In that sense, z_i reflects to be sampled from the Gaussian distributions that are learned by two GCNs that partially share parameters.

Decoder. The decoder is a generative model that seeks to reconstruct the graph, as represented by its adjacency matrix A from the latent variables z_i. The result is an estimate $\hat{A}$ of the adjacency matrix that is supposed to match the original A as well as possible. The probability that A_{ij} is one (that is there is an edge between node i and j), given the embedding vector Z, evaluates as

$$p(A_{i,j} = 1 \mid z_i, z_j) = \text{Sigmoid}(z_i^T z_j), \tag{13}$$

that the application of a sigmoid function to the scalar product of z_i and z_j. The objective function of the variational graph autoencoder (VGAE) reads as

$$C_{VGAE}$$
$$= E_{q(Z|A,F)}[\log p(A \mid Z)] - D_{KL}(q(Z \mid A, F) \parallel p(Z)) \tag{14}$$

where $D_{KL}(. \parallel .)$ reflects Kullback–Leibler divergence and $p(Z)$ is the prior distribution that governs the latent variables Z. C_{VGAE} is maximized using stochastic gradient descent; for details see again [19].

3.2. Practical implementation: FastGAE

We utilize FastGAE, a fast version of VGAE, for the implementation of variational graph autoencoding in practice. Note that while encoding is feasible in practice also for large networks, decoding is not. Therefore, FastGAE is identical to the original VGAE during the encoding phase. Decoding, however, is computationally too expensive if the underlying graph, and hence its adjacency matrix is too large. Note that here the number of nodes corresponds to the number of drugs, human and SARS-CoV-2 proteins together, which clearly exceeds the limits of the original VGAE.

To resolve the issue, FastGAE randomly samples subgraphs G_S, referring to smaller sets of nodes $S \subset V$ of size N_S, and reconstructs the corresponding submatrices A_S. FastGAE proceeds in several iterations in each of which a different subset of nodes S of size N_S is sampled. The decoding step then estimates an adjacency matrix A_S whose entries refer only nodes from S. The submatrices A_S resulting from single iterations are eventually combined into an overarching matrix $\tilde{A}$, as an approximation of the matrix $\hat{A}$ that gets reconstructed as a whole in the decoding phase of the original VGAE. The justification is that depending on the number of iterations and the size N_S of a sample, $\tilde{A}$ was shown to be a highly accurate estimate of the original A [20].

For evaluating the effects of including a feature matrix F that reflects an alternative representation of the original adjacency matrix A, we evaluated the performance of the model with and without including F. Fig. 2 shows the average performance of the model on validation sets with and without F as input for different numbers of sampling nodes. The average AUC, and AP scores are reported for 50 complete runs. From Fig. 2, it is evident that including F as a feature matrix enhances the model's performance markedly. As mentioned in Results, the explanation for this – at first glance surprising – effect is the complementarity of the methods Node2Vec (which generates F independently of VGAE) and VGAE. The integration of F evidently leads to synergetic effects between Node2Vec and VGAE.

4. Result

4.1. Dataset preparation

We have utilized three categories of interaction datasets: human protein–protein interactome data, SARS-CoV-2–host protein interaction data, and drug–host interaction data.

Table 2
Description of data sets.

Index	Dataset Category	Dataset	#Edges	#Nodes	
1	Human PPI	CCSB [29]	13 944		4303
		HPRD [27]	39 240		9617
2	SARS-CoV-2–Host PPI	Gordon et al [11]	332	27 (#SARS-CoV-2)	332 (#Host)
		Dick et al [12]	261	6 (#SARS-CoV-2)	202 (#Host)
3	Drug–target interaction	DrugBank (v4.3) [22] ChEMBL [23] Therapeutic Target Database (TTD) [24] PharmGKB database	1 788 407	1307 (# Drug)	12 134 (# Host-target)

SARS-CoV-2-host interaction data. We have taken SARS-CoV-2–host interaction information from two recent studies by Gordon et al. and Dick et al. [11,12]. In [11], 332 high confidence interactions between SARS-CoV-2 and human proteins are predicted using affinity purification mass spectrometry (AP-MS). In [12], 261 high-confidence interactions are identified using sequence-based PPI predictors (PIPE4 & SPRINT).

Drug–host interactome data. The drug–target interaction information has been collected from five databases: DrugBank database (v4.3) [22], ChEMBL [23] database, Therapeutic Target Database (TTD) [24], PharmGKB database, and IUPHAR/BPS Guide to PHARMACOLOGY [25]. The total number of drugs and drug–host interactions used in this study are 1309 and 1 788 407, respectively.

The human protein–protein interactome. We have built a comprehensive list of human PPIs from two datasets:

(1) CCSB human Interactome database consisting of 7000 genes, and 13 944 high-quality binary interactions [26]

(2) The Human Protein Reference Database [27] which consists of 8920 proteins and 53 184 PPIs.

The summary of all the datasets is provided in Table 2. The CMAP database [28] is used to annotate the drugs according to their usage with respect to different diseases.

4.2. Advantages of including a feature matrix F

For evaluating the effects of including a feature matrix F that reflects an alternative representation of the original adjacency matrix A, we evaluated the performance of the model with and without including F. Fig. 2 shows the average performance of the model on validation sets with and without F as input for different numbers of sampling nodes. The average AUC, and AP scores are reported for 50 complete runs. From Fig. 2, it is evident that including F as a feature matrix enhances the model's performance markedly. As mentioned in Results, the explanation for this—at first glance surprising—effect is the complementarity of the methods Node2Vec (which generates F independently of VGAE) and VGAE. The integration of F evidently leads to synergetic effects between Node2Vec and VGAE.

4.3. Predicting missing links: Validation

Let $G = (V, E)$ be the entire drug–human–virus interaction network in the following, where nodes $v \in V$ represent molecules (drugs, proteins). Edges $(u, v) \in E$ between molecules u, v reflect known interactions where we are interested in the case of u being a drug and v being a human protein that was found to interact with SARS-CoV-2 proteins recently [11,12]. The goal of this study is to predict such edges (u, v) to exist with great probability, despite not making explicit part of the network G.

For approving and corroborating the quality of the predictions of such virtual, non-explicit edges in the following, we designed the following canonical experiment. We first removed *all existing edges* $(u, v) \in E$ between drugs and SARS-CoV-2-associated human proteins ("CoV-host proteins") from G, resulting in an interaction network $\tilde{G} =$

$(V, \tilde{E})$ where V is the same set of molecules as before, while the edges $\tilde{E}$ lack any interaction of the type we are looking for, as just described. Note that removing *all* such edges creates a particularly challenging scenario (in comparison to, for example, only removing selected subsets of such edges).

We then ran our pipeline on $\tilde{G}$, yielding an adjacency matrix $\tilde{A}$ and a feature matrix $\tilde{F}$ (resulting from running Node2Vec on $\tilde{G}$) for training the VGAE. We used the resulting decoder for predicting missing edges. In the evaluation of this experiment, we focused on the edges that we had removed before, because these are known to be true.

Evaluating the resulting predictions confirmed that predicting missing edges by means of our pipeline operates at utmost performance (ROC–AUC: 93.56 ± 0.01 AP: 90.88 ± 0.02 averaged across 100 runs); we recall that we considered the most challenging scenario conceivable (it is conceivable that performance rates increase when re-integrating existing edges, because the model profits from the additional structure provided during training). The purpose of these simulations is to point out that the following results are trustworthy; note that in the following we do make use of all the edges that we removed in the simulations described here, which provides the additional structure, as just explained.

To choose the correct size of sampling node (N_S) in the decoding stage of FastGAE, we tested the model performance for different numbers of N_S and kept track of the corresponding performance (area under the ROC curve (AUC), average precision (AP) score) and model training time) in the frame of a train-validation-test split at proportions 8:1:1. Table 1 shows the performance of the model for sampled subgraph sizes $N_S = 7000, 5000, 3000, 2500$ and 1000. For 5000 sampled nodes, the model's performance is sufficiently good concerning its training time and validation-AUC and -AP score. The average test ROC–AUC and AP score of the model for $N_s = 5000$ are 88.53 ± 0.03 and 84.44 ± 0.04.

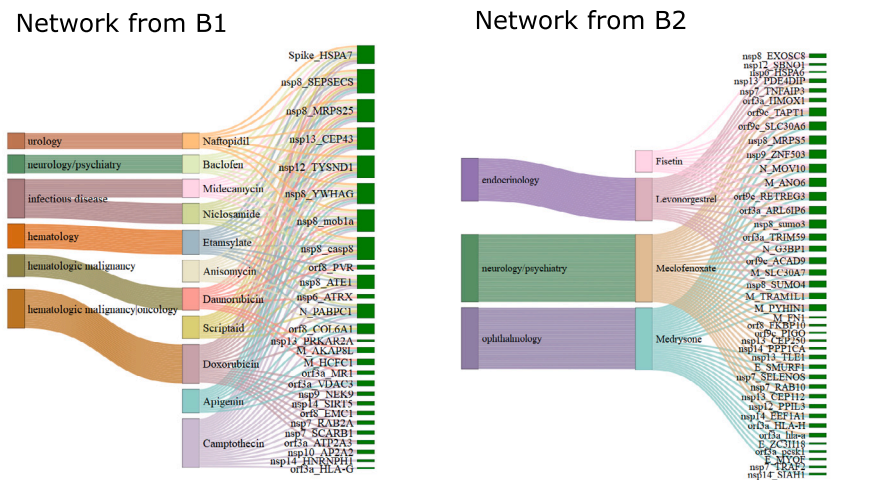### 4.4. Drug–CoV-host interaction prediction

The overall number of possible links between drugs and CoV-host proteins amounts to 332×1302 (CoV-host $\times$ drugs). While many such links make part of the network, the majority of such possible links does not make part of the network. We refer to all such links that do not make part of network as "non-edges". Any such "non-edge" that is predicted to exist at sufficiently high probability is a prediction for an interaction of a drug with a CoV-host protein. We train the model on the whole network reflected by adjacency matrix A and feature matrix F (the latter computed by Node2Vec). The trained model is then applied to the "non-edges" to discover the most probable missing drug–Cov–host interactions. Fig. 3(a), Panel-A shows the heatmap of probability scores between predicted drugs and CoV-host proteins. We identified 692 links, connecting 92 drugs with 78 CoV-host proteins whose probability to exist exceeded a threshold of 0.8. As we will illustrate further in the following, the predicted CoV-host proteins are involved in different pathways that are crucial for viral infections (Supplementary Table 3). We further used a Weighted bipartite clustering algorithm [30] for analyzing the bipartite graph whose partitions consist of drugs on the one hand, and CoV-host proteins on the other hand further. Application of
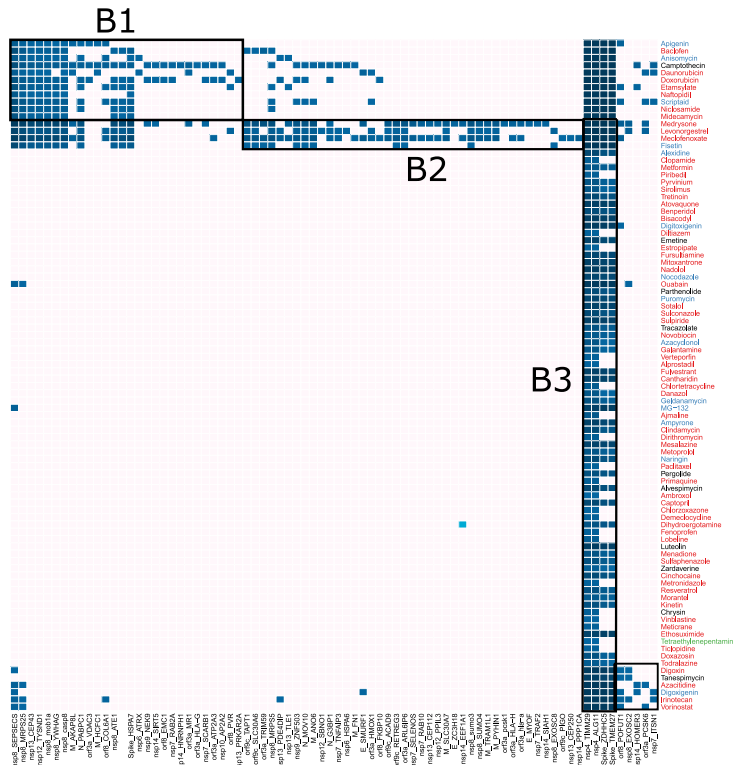
(a) Drug–CoV-host predicted interaction: panel-A shows heatmap of probability scores between 92 drugs and 78 CoV-host proteins. The four predicted bipartite modules are annotated as B1, B2, B3 and B4 within the heatmap. The drugs are colored based on their clinical phase (red–launched, preclinical–blue, phase2/phase3–green and phase-1/ phase-2–black).

(b) Panel-B represents networks corresponding to B1 and B2. The drugs are annotated using the disease area found in CMAP database [25]

Fig. 3. Drug–CoV-host predicted interaction.

the algorithm results in 4 bipartite modules (Panel-A Fig. 3(a)): B1 (11 drugs, 28 CoV-host), B2 (4 drugs, 41 CoV-host), B3 (71 rugs and 4 CoV-host), and B4 (6 drugs and 5 CoV-host). Panels B–D in Figs. 3(b), 3(c), and 3(d) show the network diagram of four bipartite modules. Of note,

several antibiotics (Anisomycin and Midecamycin in B1; Puromycin, Demeclocycline, Dirithromycin, Geldanamycin, and Chlortetracycline in B3), anti-cancer drugs (Doxorubicin, Camptothecin) and other drugs (Lobeline and Ambroxol in B3) have a variety of therapeutic uses,
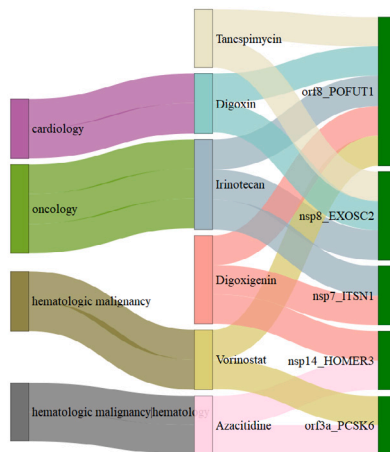
Panel-C

Network from B3
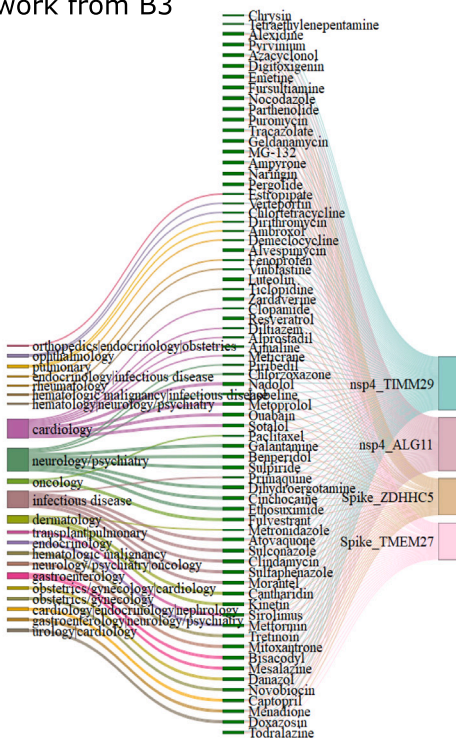
(c) Panel-C represents networks corresponding to B3. The drugs are annotated using the disease area found in CMAP database [25]

Panel-D

Network from B4

(d) Panel-D represents networks corresponding to B4. The drugs are annotated using the disease area found in CMAP database [25]

Fig. 3. (*continued*).

including bronchitis, pneumonia, and respiratory tract infections [31] which provides further evidence of the reasonability of our results.

High-confidence interactions (exceeding a probability threshold of 0.9) are further shown in Fig. 4(a), Panel A. To highlight some re-purposable drug combination and their predicted CoV-host target, we perform a weighted clustering (using clusterONE [32]) on this network, resulting in several quasi-bicliques (shown in Panels B–E of Fig. 4(b))

We matched our predicted drugs with the drug list recently published by Zhou et al. [9] and found six common drugs: Mesalazine,
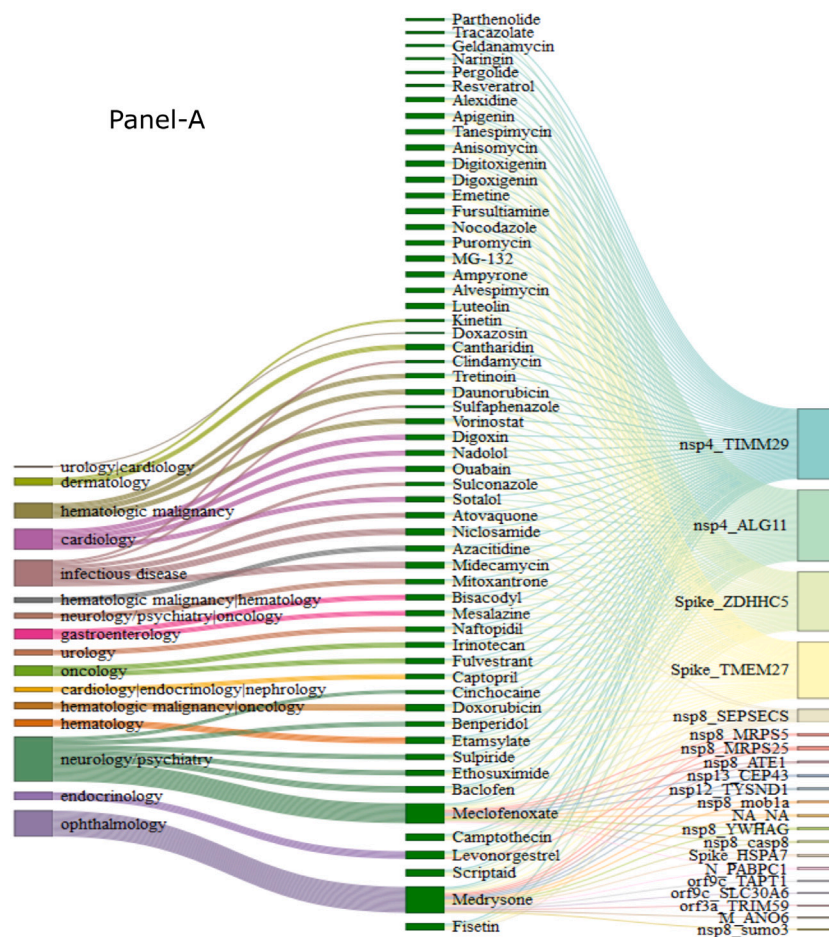
(a) Panel-A shows the interaction graph between drugs and CoV-host. Drugs are annotated with their usage.
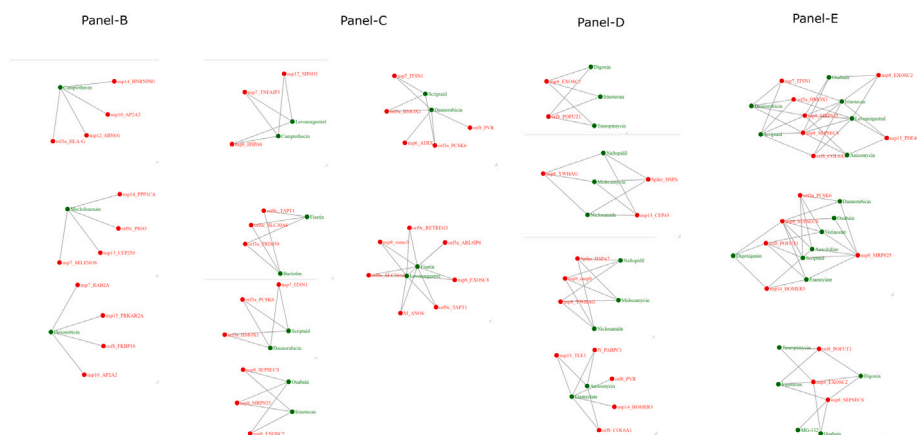
(b) Panel-B, C, D and E represents quasi-bicliques for one, two, three and more than three drugs molecules respectively.

Fig. 4. Predicted interactions for probability threshold: 0.9. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Vinblastine, Menadione, Medrysone, Fulvestrant, and Apigenin. Among them, Apigenin has a known effect on the antiviral activity together with quercetin, rutin, and other flavonoids [33]. Mesalazine is also proven to be extremely effective in the treatment of other viral diseases like influenza A/H5N1 virus [34].

4.5. Repurposable drugs for SARS-CoV-2

Here we showcased some repurposable drugs that have prominent literature-reported antiviral evidence, especially for two other coronaviruses SARS-CoV and MERS-CoV. Some drugs are directly associated with the treatment of SARS-CoV-2 as well. The details of the

predicted drugs and their uses are given in supplementary text and Supplementary Table-2.

Topoisomerase inhibitors. Topoisomerase Inhibitors such as Camptothecin, Daunorubicin, Doxorubicin, Irinotecan and Mitoxantrone are in the list of predicted drugs. The anticancer drug camptothecin (CPT) and its derivative Irinotecan have a potential role in antiviral activity [35]. Daunorubicin (DNR) is demonstrated as an inhibitor of HIV-1 virus replication in human host cells [36]. The anticancer antibiotic Doxorubicin was previously identified as a selective inhibitor of *in-vitro* Dengue and Yellow Fever virus [37]. Mitoxantrone shows antiviral activity against the human herpes simplex virus (HSV1) by reducing the transcription of viral genes in many human cells that are essential for DNA synthesis [38].

Histone deacetylases inhibitors (HDACi). Our predicted drug list (supplementary table-2) contains two HDACi: Scriptaid and Vorinostat. Both drugs can be used to achieve latency reversal in the HIV-1 virus safely and repeatedly [39]. Asymptomatic patients infected with SARS-CoV-2 are of significant concern as they are more vulnerable to infect large population than symptomatic patients. Moreover, in most cases (99-percentile), patients develop symptoms after an average of 5–14 days, which is longer than the incubation period of SARS and MERS. To this end, HDACi may serve as good candidates for recognizing and clearing the cells in which SARS-CoV-2 latency has been reversed.

HSP inhibitor. Heat shock protein 90 (HSP) is described as a crucial host factor in the life cycle of several viruses that includes an entry in the cell, nuclear import, transcription, and replication [40,41]. HSP90 is also shown to be an essential factor for SARS-CoV-2 envelop (E) protein [42]. In [43], HSP90 is described as a promising target for antiviral drugs. The predicted drug list contains three HSP inhibitors: Tanespimycin, Geldanamycin, and its derivative Alvespimycin. The first two have a substantial effect in inhibiting the replication of Herpes Simplex Virus and Human enterovirus 71 (EV71), respectively. Recently in [44], Geldanamycin and its derivatives are proposed to be an effective drug in the treatment of COVID-19.

Antimalarial agent, DNA-inhibitor, DNA methyltransferase/synthesis inhibitor. Inhibiting DNA synthesis during viral replication is one of the critical steps in disrupting the viral infection. The list of predicted drugs contains six such small molecules/drugs, viz., Niclosamide, Azacitidine, Anisomycin, Novobiocin, Primaquine, Menadione, and Metronidazole (see supplementary text). Recently Hydroxychloroquine (HCQ), a derivative of CQ, has been evaluated to efficiently inhibit SARS-CoV-2 infection *in vitro* [45]. Therefore, another anti-malarial aminoquinolin drug Primaquine may also contribute to the attenuation of the inflammatory response of COVID-19 patients. Primaquine is already established to be effective in the treatment of Pneumocystis-pneumonia (PCP) [46].

Cardiac glycosides ATPase inhibitor. The predicted list of drugs contains three cardiac glycosides ATPase inhibitors: Digoxin, Digitoxigenin, and Ouabain. These drugs have been reported to be effective against different viruses such as herpes simplex, influenza, chikungunya, coronavirus, and respiratory syncytial virus [47].

Mg132, resveratrol and captopril. MG132, a proteasomal inhibitor, is a strong inhibitor of SARS-CoV replication in early stage [48]. Resveratrol has also been demonstrated to be a significant inhibitor MERS-CoV infection [49]. Another drug Captopril is known as Angiotensin II receptor blockers (ARB), which directly inhibits the production of angiotensin II. In [50], Angiotensin-converting enzyme 2 (ACE2) is demonstrated as the binding site for SARS-CoV-2. So Angiotensin II receptor blockers (ARB) may be good candidates to use in the tentative treatment for SARS-CoV-2 infections

5. Discussion

In this work, we have successfully generated a list of candidate drugs that can be repurposed to counteract SARS-CoV-2 infections. As novelties, we have integrated the most recently published SARS-CoV-2 interaction data into well-established network resources to raise an encompassing network putting drugs, viral and human proteins into a comprehensive context. Further, to exploit this novel network, we have made use of the most recent and advanced deep learning methodology that addresses learning and exploiting network data, establishing another novelty. Experiments validate that our predictions are of utmost accuracy, which confirms the quality of the novel interactions between drugs and virus related proteins that we suggest.

The recent publication of two novel SARS-CoV-2–human protein interaction resources [11,12] has unlocked enormous possibilities in studying the mechanisms that drive virulence and pathogenicity of SARS-CoV-2. Only now sufficiently systematical and accurate, AI supported drug repurposing strategies for fighting COVID-19 have become conceivable.

To the best of our knowledge, we have raised such a systematic approach of utmost accuracy with an advanced AI boosted model for the first time. We have integrated the new SARS-CoV-2 protein interaction data into well established, carefully curated resources, capturing hundreds of thousands of approved interfaces between molecules that reflect drugs or human proteins. As a result, we have been able to raise a comprehensive drug–human–SARS-CoV-2 network that reflects the latest state of the art with respect to the interactions that it displays.

This new network already establishes a novel resource in its own right. For exploiting it, we have opted for using variational graph autoencoders (VGAE), which have been most recently presented as the state of the art in analyzing large network datasets, and which allow to predict links that are missing in the network whose structure and the rules that underlie the interplay of links it has "learned" at utmost accuracy. Note that FastGAE, the practical implementation of VGAEs that enables us to analyze networks of sufficiently large sizes, was presented only very recently [20] as well, pointing out the timeliness of our study yet again.

Simulation experiments, reflecting scenarios where links known to exist are predicted upon their artificial removal, have pointed out that our approach operates with utmost accuracy.

Encouraged by these simulations, we predicted links to be missing without prior removal of links. Our predictions have revealed 692 high confidence interactions between human proteins that are essential for the virus on the one hand and 92 drugs on the other hand; note that we had been emphasizing host-directed therapy (HDT) strategies, which explains why we have focused on the type of interaction just described. We recall that the combination of HDT and drug repurposing promises to yield drugs that not only enable accelerated usage, but also guarantee a sufficiently high degree of sustainability.

We further systematically categorized the 92 repurposable drugs into 70 categories based on their domains of application and molecular mechanism. According to this, we identified and highlighted several drugs that target host proteins that the virus needs to enter and subsequently hijack human cells. One such example is Captopril, which directly inhibits the production of Angiotensin-Converting Enzyme-2 (ACE-2), in turn already known to be a crucial host factor for SARS-CoV-2. Further, we identified Primaquine, as an antimalaria drug used to prevent the Malaria and also Pneumocystis pneumonia (PCP) relapses, because it interacts with the TIM complex TIMM29 and ALG11. Moreover, we have highlighted drugs that act as DNA replication inhibitor (Niclosamide, Anisomycin), glucocorticoid receptor agonists (Medrysone), ATPase inhibitors (Digitoxigenin, Digoxin), topoisomerase inhibitors (Camptothecin, Irinotecan), and proteasomal inhibitors (MG-132). Note that some drugs are known to have rather severe side effects from their original use (Doxorubicin, Vinblastine),

but the disrupting effects of their short-term usage in severe COVID-19 infections may mean sufficient compensation.

In summary, we have compiled a list of drugs, when repurposed, are of great potential in the fight against the COVID-19 pandemic, where therapy options are still urgently needed. Our list of predicted drugs suggests both options that had been identified and thoroughly discussed before, as well as new opportunities that had not been pointed out earlier. The latter class of drugs may offer valuable chances for pursuing new therapy strategies against COVID-19.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability

All codes and datasets are given in the github link: https://github.com/sumantaray/Covid19.

Acknowledgment

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.artmed.2022.102418.

References

[1] Wu F, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020;579(7798):265–9.

[2] Forst CV. Host–pathogen systems biology. In: Infectious disease informatics. Springer; 2010, p. 123–47.

[3] Ray S, Lall S, Bandyopadhyay S. A deep integrated framework for predicting SARS-CoV2–human protein-protein interaction. IEEE Transactions on Emerging Topics in Computational Intelligence 2022.

[4] Kaufmann SH, Dorhoi A, Hotchkiss RS, Bartenschlager R. Host-directed therapies for bacterial and viral infections. Nat Rev Drug Discov 2018;17(1):35.

[5] de Chassey B, Meyniel-Schicklin L, Aublin-Gex A, André P, Lotteau V. New horizons for antiviral drug discovery from virus–host protein interaction networks. Curr Opin Virol 2012;2(5):606–13.

[6] Doolittle JM, Gomez SM. Mapping protein interactions between dengue virus and its human and insect hosts. PLoS Negl Trop Dis 2011;5(2).

[7] Bandyopadhyay S, Ray S, Mukhopadhyay A, Maulik U. A review of in silico approaches for analysis and prediction of HIV-1-human protein–protein interactions. Brief Bioinform 2015;16(5):830–51.

[8] Cao H, Zhang Y, Zhao J, Zhu L, Wang Y, Li J, Feng Y-M, Zhang N. Prediction of the Ebola virus infection related human genes using protein-protein interaction network. Combin Chem High Throughput Screen 2017;20(7):638–46.

[9] Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov 2020;6(1):1–18.

[10] Li X, Yu J, Zhang Z, Ren J, Peluffo AE, Zhang W, Zhao Y, Yan K, Cohen D, Wang W. Network bioinformatics analysis provides insight into drug repurposing for COVID-2019. 2020, Preprints.

[11] Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O'Meara MJ, Rezelj VV, Guo JZ, Swaney DL, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 2020;1–13.

[12] Dick K, Biggar KKG, R. J. Comprehensive prediction of the SARS-CoV-2 vs. Human interactome using PIPE4, SPRINT, and PIPE-sites. 2020, http://dx.doi.org/10.5683/SP2/JZ77XA.

[13] Sadegh S, Matschinske J, Blumenthal DB, Galindez G, Kacprowski T, List M, Nasirigerdeh R, Oubounyt M, Pichlmair A, Rose TD, et al. Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. Nature Commun 2020;11(1):1–9.

[14] Fiscon G, Conte F, Farina L, Paci P. SAveRUNNER: a network-based algorithm for drug repurposing and its application to COVID-19. PLoS Comput Biol 2021;17(2):e1008686.

[15] Gysi DM, Do Valle Í, Zitnik M, Ameli A, Gan X, Varol O, Ghiassian SD, Patten J, Davey RA, Loscalzo J, et al. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. Proc Natl Acad Sci 2021;118(19).

[16] Ray S, Alberuni S, Maulik U. Computational prediction of HCV-human protein-protein interaction via topological analysis of HCV infected PPI modules. IEEE Transactions on NanoBioscience 2018;17(1):55–61.

[17] Ray S, Bandyopadhyay S. A NMF based approach for integrating multiple data sources to predict HIV-1–human PPIs. BMC bioinformatics 2016;17(1):1–13.

[18] Grover A, Leskovec J. Node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016, p. 855–64.

[19] Kipf TN, Welling M. Variational graph auto-encoders. 2016, arXiv preprint arXiv:1611.07308.

[20] Salha G, Hennequin R, Remy J-B, Moussallam M, Vazirgiannis M. FastGAE: Fast, scalable and effective graph autoencoders with stochastic subgraph decoding. 2020, ArXiv Preprint.

[21] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. 2013, p. 3111–9.

[22] Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res 2014;42(D1):1091–7.

[23] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 2012;40(D1):D1100–7.

[24] Yang H, Qin C, Li YH, Tao L, Zhou J, Yu CY, Xu F, Chen Z, Zhu F, Chen YZ. Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. Nucleic Acids Res 2016;44(D1):D1069–74.

[25] Pawson AJ, Sharman JL, Benson HE, Faccenda E, Alexander SP, Buneman OP, Davenport AP, McGrath JC, Peters JA, Southan C, et al. The IUPHAR/BPS guide to PHARMACOLOGY: an expert-driven knowledgebase of drug targets and their ligands. Nucleic Acids Res 2014;42(D1):D1098–106.

[26] Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, et al. Towards a proteome-scale map of the human protein–protein interaction network. Nature 2005;437(7062):1173–8.

[27] Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjan V, Muthusamy B, Gandhi T, Gronborg M, et al. Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res 2003;13(10):2363–71.

[28] Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Davis JF, Tubelli AA, Asiedu JK, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell 2017;171(6):1437–52.

[29] Yu H, Tardivo L, Tam S, Weiner E, Gebreab F, Fan C, Svrzikapa N, Hirozane-Kishikawa T, Rietman E, Yang X, et al. Next-generation sequencing to generate interactome datasets. Nature Methods 2011;8(6):478.

[30] Beckett SJ. Improved community detection in weighted bipartite networks. R Soc Open Sci 2016;3(1):140536.

[31] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006;34(suppl_1):D668–72.

[32] Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. Nature Methods 2012;9(5):471.

[33] Salehi B, Venditti A, Sharifi-Rad M, Kregiel D, Sharifi-Rad J, Durazzo A, Lucarini M, Santini A, Souto EB, Novellino E, et al. The therapeutic potential of apigenin. Int J Mol Sci 2019;20(6):1305.

[34] Zheng B-J, Chan K-W, Lin Y-P, Zhao G-Y, Chan C, Zhang H-J, Chen H-L, Wong SS, Lau SK, Woo PC, et al. Delayed antiviral plus immunomodulator treatment still reduces mortality in mice infected by high inoculum of influenza A/H5N1 virus. Proc Natl Acad Sci 2008;105(23):8091–6.

[35] Horwitz SB, Chang C-K, Grollman AP. Antiviral action of camptothecin. Antimicrob Agents Chemother 1972;2(5):395–401.

[36] Filion L, Logan D, Gaudreault R, Izaguirre C. Inhibition of HIV-1 replication by daunorubicin. Clin Investig Med. Med Clin Exp 1993;16(5):339–47.

[37] Kaptein SJ, De Burghgraeve T, Froeyen M, Pastorino B, Alen MM, Mondotte JA, Herdewijn P, Jacobs M, De Lamballerie X, Schols D, et al. A derivate of the antibiotic doxorubicin is a selective inhibitor of dengue and yellow fever virus replication in vitro. Antimicrob Agents Chemother 2010;54(12):5269–80.

[38] Huang Q, Hou J, Yang P, Yan J, Yu X, Zhuo Y, He S, Xu F. Antiviral activity of mitoxantrone dihydrochloride against human herpes simplex virus mediated by suppression of the viral immediate early genes. BMC Microbiol 2019;19(1):274.

[39] Archin NM, Kirchherr JL, Sung JA, Clutton G, Sholtis K, Xu Y, Allard B, Stuelke E, Kashuba AD, Kuruc JD, et al. Interval dosing with the HDAC inhibitor vorinostat effectively reverses HIV latency. J Clin Investig 2017;127(8):3126–35.

[40] Ju H-Q, Xiang Y-F, Xin B-J, Pei Y, Lu J-X, Wang Q-L, Xia M, Qian C-W, Ren Z, Wang S-Y, et al. Synthesis and in vitro anti-HSV-1 activity of a novel Hsp90 inhibitor BJ-B11. Bioorg Med Chem Lett 2011;21(6):1675–7.

[41] Shim HY, Quan X, Yi Y-S, Jung G. Heat shock protein 90 facilitates formation of the HBV capsid via interacting with the HBV core protein dimers. Virology 2011;410(1):161–9.

[42] DeDiego ML, Nieto-Torres JL, Jiménez-Guardeño JM, Regla-Nava JA, Alvarez E, Oliveros JC, Zhao J, Fett C, Perlman S, Enjuanes L. Severe acute respiratory syndrome coronavirus envelope protein regulates cell stress response and apoptosis. PLoS Pathogens 2011;7(10).

[43] Wang Y, Jin F, Wang R, Li F, Wu Y, Kitazato K, Wang Y. HSP90: a promising broad-spectrum antiviral drug target. Arch Virol 2017;162(11):3269–82.

[44] Sultan I, Howard S, Tbakhi A. Drug repositioning suggests a role for the heat shock protein 90 inhibitor geldanamycin in treating COVID-19 infection. 2020, ArXiv.

[45] Liu J, Cao R, Xu M, Wang X, Zhang H, Hu H, Li Y, Hu Z, Zhong W, Wang M. Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro. Cell Discov 2020;6(1):1–4.

[46] Vöhringer H-F, Arastéh K. Pharmacokinetic optimisation in the treatment of Pneumocystis carinii pneumonia. Clin Pharmacokinet 1993;24(5):388–412.

[47] Amarelle L, Lecuona E. The antiviral effects of na, K-ATPase inhibition: A minireview. Int J Mol Sci 2018;19(8):2154.

[48] Schneider M, Ackermann K, Stuart M, Wex C, Protzer U, Schätzl HM, Gilch S. Severe acute respiratory syndrome coronavirus replication is severely impaired by MG132 due to proteasome-independent inhibition of M-calpain. J Virol 2012;86(18):10112–22.

[49] Lin S-C, Ho C-T, Chuo W-H, Li S, Wang TT, Lin C-C. Effective inhibition of MERS-CoV infection by resveratrol. BMC Infect Dis 2017;17(1):144.

[50] Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F. Structural basis of receptor recognition by SARS-CoV-2. Nature 2020;1–4.