ICWR2022
May 11-12, 2022; Tehran, Iran
8th International Conference on Web Research

Dcomg: Drug Combination Prediction by Applying Gnns on DDI Node2vec Features

Seyyed Sina Ziaee, Hossein Rahmani*, Mina Tabatabaei

School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran
Sinaziaee99@gmail.com, h_rahmani@iust.ac.ir, mina_tabatabaei@comp.iust.ac.ir

***Abstract*— Recent studies have been indicating that many clinical drug combinations surpass single-drug therapy efficacy. Machine learning, deep learning, network analysis, and search algorithms have been considered to facilitate the discovery of synergistic drug combinations, and two of the best state-of-the-art models in this area are under the deep learning category. In this paper, we present DComG, a Graph Auto Encoder method to predict synergistic drug combinations. Using the dataset provided in DCDB, our analysis shows tremendous improvement in the performance of predicting new drug combinations over previously introduced state-of-the-art models by an average of 4% in ROC_AUC scores. We highlight the importance of drug-drug interactions (DDI) in the form of node2vec features of DComG graph inputs for predicting new drug combinations. Finally, we address the results of our model in terms of biological interpretations of drug combinations based on recent medical drug combination papers in the literature.**

Keywords*— *Datamining; Graph Neural Network; Drug Combination; Drug-Drug Interaction; Node2vec; PCA; Graph.

I. INTRODUCTION

The synergistic effect of drug combination is a well-established concept in the treatment of various diseases [1]-[10]. Clinically experimenting all possible drug pairs is impossible due to the large number of possibilities and limited sources [33]-[34]. Thus, several computational methods have been introduced to prioritize combinations and decrease the amount of time and effort for clinical testing [35]-[37].

Among previous works in this field, the prediction methods depended extremely on customized feature engineering or statistical models built by domain experts [2]-[4]. This essentially reduces these methods' capability to detect hidden synergistic drug combinations, and the exercise tended to be very labor-intensive. Many machines learning approaches, such as SVM [9] and Naïve Bayesian classifiers [10] have also been applied to predict drug combinations. Deep Learning methods, on the other hand, capture high performance in an end-to-end fashion. The autoencoders in Zhang et al. were used to map new representations of features and eventually predict the synergy score of drug pairs [15]. Jiang et al. used a heterogeneous graph of drug-drug combination, drug-protein interaction, and protein-protein interactions as the input of their GCN model to predict drug combinations with high synergistic scores. However, scaling up these techniques is not an easy and efficient task, due to extremely high dimensional data in real-world graphs, which results in high computational overhead in execution time and storage.

In Deep graph neural networks, semi-supervised approaches such as selecting more vital information as node features of a graph can be highly useful alongside information gained from graph structures in their end-to-end approach.

The main challenges of previous methods in this area are the limitations in the size of datasets and the low number of drug indications (mainly anti-cancer drugs). Also, graph neural networks have been proved to be highly useful, especially on unstructured datasets, but have not been extensively used on drug combination predictions. In this paper, we propose DComG which uses a Graph Auto Encoder (GAE) model [22] using Graph Convolutional Network (GCN) [19] to find new synergistic drug pairs. The GCN method is a type of convolutional neural network model that is mainly used on graphs. Here we use GCN to do homogeneous graph embedding and subsequently solve a link prediction task. Inputs of DComG incorporate rich node2vec features extracted from drug-drug interactions from DCDB [17]. Dimension reduction techniques such as PCA are not only extremely helpful in improving model performance on validation and test but also result in reducing the training time of the DComG model. This approach makes DComG a fast and efficient model that outperforms previously introduced state-of-the-art models by 4% w.r.t AUC scores.

The remaining sections are as follows: In the second part, the related previous work is discussed. The third section lays out the DComG construction and the preparation of node2vec features. In the fourth section, we represent the performance of our model and the biological interpretation of top predicted drug combinations. Finally, the conclusion is stated following a brief discussion of the results.

II. BACKGROUND

To accelerate the discovery of synergistic drug combinations and to avoid the time-consuming and labor-intensive task of clinically experimenting with drug pairs, four primary categories of methods can be considered: search algorithms, network analysis, machine learning, and deep learning methods. The genetic algorithm as a stochastic search algorithm was used to deal with drug combinations prediction, starting with 18 random combinations of 19 drugs as the first generation [1]. In network analysis, the drug combinations predictions problem was converted to a link prediction problem and solved by Regularized Least Squares (KRLS) algorithm [2] or a different proximity metric was proposed. By investigating the relationship between drug targets and disease proteins in the protein-protein interaction networks, they concluded that drugs with close targets to disease can have a better effect [3] and others concluded that the proximity of drug-target modules in the protein-protein interaction network is correlated with the chemical and functional similarities of the two drugs [4].

Unsupervised learning, semi-supervised learning, and supervised learning as three major machine learning approaches, that can be considered in the task of drug combination prediction. In unsupervised methods, the purpose is to organize the data and to find out about the hidden structure of the data. Some assumed that a drug combination that targets the same or related pathways is more likely to have a synergistic effect [7]. Others believed that drug administration affects a subnetwork or pathway in the networked cellular system and they came up with an interaction score that showed how far apart the drug's effectiveness and side effects were [8]. In semi-supervised methods, by applying a manifold ranking algorithm they tried to achieve a drug pairs ranking based on the most similar to the labeled pairs [5]. Another semi-supervised approach concluded that drugs that have a synergistic effect with similar secondary drugs have a similar chemical structure and drug-target interactions, and they used the Laplacian Regularized Least Squares learning algorithm to predict promising synergistic drug combinations [6].

In the supervised category, Shi et al. used five SVM classifiers at the first level to learn five different features and one SVM at the next level to concatenate all the outputs [9], Bai et al.'s model used five features and created an upgraded Naïve Bayesian classifier [10], Li et al. have constructed five similarity matrix and used feature selection and ensemble model to predict [11], Liu et al. applied random walk with restart (RWR) on the heterogeneous network and got a probability distribution for each drug combination to use in gradient tree boosting (GTB) classifier [12].

In the deep learning category, Preuer et al, used a two-layer feedforward neural network with the input being a concatenated vector of chemical descriptors of both drugs and gene expression values of corresponding cancer cell lines [13]. Xia et al. passed every feature through its own deep fully connected neural network and passed the concatenated vector of features to a final fully connected neural network [14]. Zhang et al. used three autoencoders to obtain the encoded representations of features and a deep neural network to predict the synergy score [15]. Jiang et al. used GCN on an integrated drug-drug combination, drug-protein interaction, and protein-protein interaction networks as input to predict synergy scores [16].

In this paper, we consider the downfalls of previous methods like the limitations to only anti-cancer drugs, and propose DComG, which outperforms previous state-of-the-art models w.r.t AUC scores and can be performed on a variety of drugs with different indications.

III. METHOD

Datasets of drug combinations generate graphs, which are unstructured data. Thus, common machine and deep learning models that reach high performances on structured datasets are not necessarily effective on graphs. Graph auto encoders with simple machine learning approaches have been used on graphs but on the other hand, GNNs are not that frequent in this area. So, the idea of our model simply comes from this simple fact that GNNs are highly effective on graphs and have not been extensively used on drug combinations.

The main steps of our proposed DComG model are illustrated in Fig. 1. The empirical results section contains the description of the hyper-parameters used in DComG after tunning the model and the overall steps of DComG shown in Fig. 1 are as follows:

A. Feature Construction and Selection

The goal of this step is to extract appropriate node features for the input graph of the model.

- DComG accepts drug-drug interaction networks available in DCDB [17].
- DComG performs the node2vec method to extract features for each drug in the dataset. The result of this step is a unique vector of size 128 for each drug in the dataset. Therefore, a matrix of size N * 128 where N is the number of drugs available, is created as node features.
- DComG takes node features at the previous step and performs a dimension reduction method named PCA to reduce the number of features that need to be fed into the GCN model. The result of this method is a matrix of size N * 32, in which there exists a vector of size 32 for each drug in the dataset.

B. Construction of Drug Combination Graph

The goal of this step is to create a graph of previously approved drug combinations.

- DComG takes a set of approved drug combinations.
- DComG performs a graph construction from the available set of drug combinations, where each drug is a node and each drug combination is an edge between two drugs in the dataset.

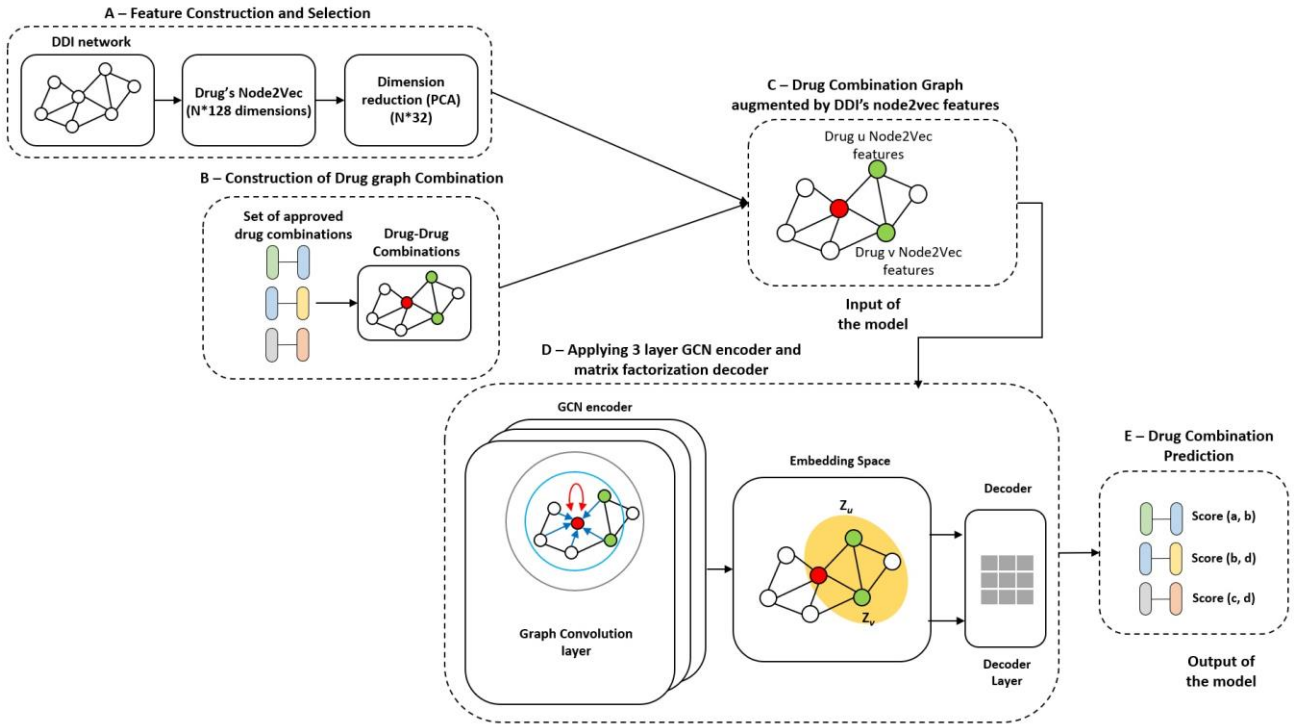

Fig. 1. The block diagram of DComG, which contains five steps: A. Feature Construction and selection, B. Construction of Drug Graph combination, C. Drug Combination Graph augmented by DDI's node2vec features, D. Applying 3-layer GCN encoder and matrix factorization decoder And E. Drug Combination synergistic score predictions

C. Drug Combination Graph augmented by DDI's node2vec features

The goal of this step is to incorporate the node2vec features extracted from step A in the graph created in step B as node features.

D. Applying a 3-layer GCN encoder and matrix factorization decoder

The model created here is a Graph Auto Encoder (GAE) which aims to create a new representation of the graph in another space to result in new synergistic drug pairs. The model includes two main components. The first is a GCN encoder and the second is a shallow decoder, which is a simple matrix factorization method. The idea behind incorporating GCNConv layers to perform drug combination tasks, comes from a paper by Zitnik et al. [40], where they used GCN layers to predict the side effects of drug combinations.

- DComG performs a GCN encoder to map each node of the graph in a new embedding space. Given a graph G = (V, E) where G is a set of N nodes as drugs, and E is a set of M edges specifically drug-drug combinations and drug-drug interactions as links, each of these N nodes have a unique feature vector x1, x2, x3, …, xN ∈ Rd, where d is the dimension of the feature vector. Also, edges are denoted as (vi, vj) between node vi and node vj.
- GCN model proposed by Kipf et al [19] in 2016, is a semi-supervised learning method that operates as a multi-layer propagation (MLP) model on graphs. For the previously mentioned graph G, D (Dii = ΣjAij) is a degree matrix. This model contains a layer-wise propagation (1):

$$H^{(l+1)} = \sigma(\bar{D}^{-\frac{1}{2}}\bar{A}\bar{D}^{\frac{1}{2}}H^{(l)}W^{(l)}) \quad (1)$$

Were, A ∈ RN*N is an adjacency matrix that is gained from undirected graph G and results in = A + I. Here, I is the identity matrix, = , W(l) is a weight matrix for each specific layer which is trainable, σ () denotes the activation function (i.e., sigmoid), H(l) ∈ RN*D is the matrix of activations of the lth layer and as an example, consider = X (X is the feature matrix consisting of x1, x2, x3, …, xN).

After k layers of feature vector xi, the final output is the embedding vector zi. GCN models are functions such as g (X, A) = Z where Z is the output of the model or the embedding feature matrix consisting of z1, z2, z3, …, zN.

After mapping each node into a new embedding space, the primary task is to represent the graph in the new space, which is known as decoding. From the GCN encoder, each node (drug) is first encoded into a new embedding space with a unique vector representation, then we have a set of vectors for all the nodes (Fig. 1). After the optimization is

performed, the synergy scores of all possible drug combinations between all pairs of drugs are calculated. Therefore, predicting the existence of an edge between drug u and v in the new embedding space with the nodes Zu and Zv method is improved and the mentioned score is computed with the following matrix factorization (2):

$$\text{score}(u, v) = ZTu \,.\, Zv \qquad (2)$$

Here Zu and Zv are the new embedding vectors for node u and v respectively and score (u, v) is the probability of node u and v being adjacent to each other or in other words the probability in which the combination of u and v is synergistic in the new embedding space. A higher value score represents a higher potential for synergy.

E. Drug Combination Prediction

The result of this step is a matrix of size N * N and each cell of this matrix is a score. Higher scores represent a higher probability of an edge between two nodes and eventually a higher combination probability.

IV. EMPIRICAL RESULTS

A. Data Collection and Preparation

In DComG, both the drug-drug interaction dataset and drug combination dataset are taken from DCDB (Version 2.0) [17]. DCDB is collected from about 140,000 clinical studies and the Food and Drug Administration (FDA) Orange Electronic Book, which includes 866 new drug combinations (1363 in total) compared to the previous version. It covers 904 individual drugs and 805 targets. For combinations with more than two drugs, we considered all the pairwise drug combinations as synergistic combinations. In the end we reached 760 drugs (drugs that exist in both drug-drug interaction dataset and drug combination dataset) and 1866 drug combinations to use as the input of DComG.

DCDB includes drug interactions in drug combinations which can be divided into pharmacokinetic and pharmacodynamic interactions. When one drug's effects are altered by the presence of another at the drug's site of action, this is referred to as a pharmacodynamic interaction, whereas pharmacokinetic interactions occur when the presence of another drug alters the absorption, distribution, metabolism, and excretion processes of one drug.

All of the drug interactions used in this article are extracted from DCDB. To create a network of drug-drug interactions in which nodes and edges represent drugs and interactions, respectively. Using the node2vec algorithm [18], we have extracted a representation for each drug using the drug-drug interaction network that will be used in the next steps.

The step following graph construction is splitting the data into train, validation, and test sets shown in Fig. 2. Here 10% of the edges are used for the validation set and another 10% of the edges are used for the test set. The remaining data is used for the train set. Here we need to introduce a notion called positive and negative edges (edge_index in code [38]). Positive edges are the true edges between drugs in the original drug-drug combination graph and negative edges are edges that do not exist in the original graph. A portion of positive edges are used for message passing in the graph neural network (GNN) model and the other edges in the positive portion combined with negative edges are used for training the link prediction task. One of the main steps in training a GNN model is message passing [21] in which neighboring nodes and edges exchange information and influence each other's updated embeddings. DComG uses negative sampling to select negative edges. The number of negative samples in each set is proportional to the number of positive edges in that set, therefore all train, validation, and test sets are balanced. When machine learning models have many times more negative samples than positive samples, negative sampling is one of the most vital procedures used, which can estimate the model's performance accurately by generating a certain proportion of negative samples from the sampling distribution.

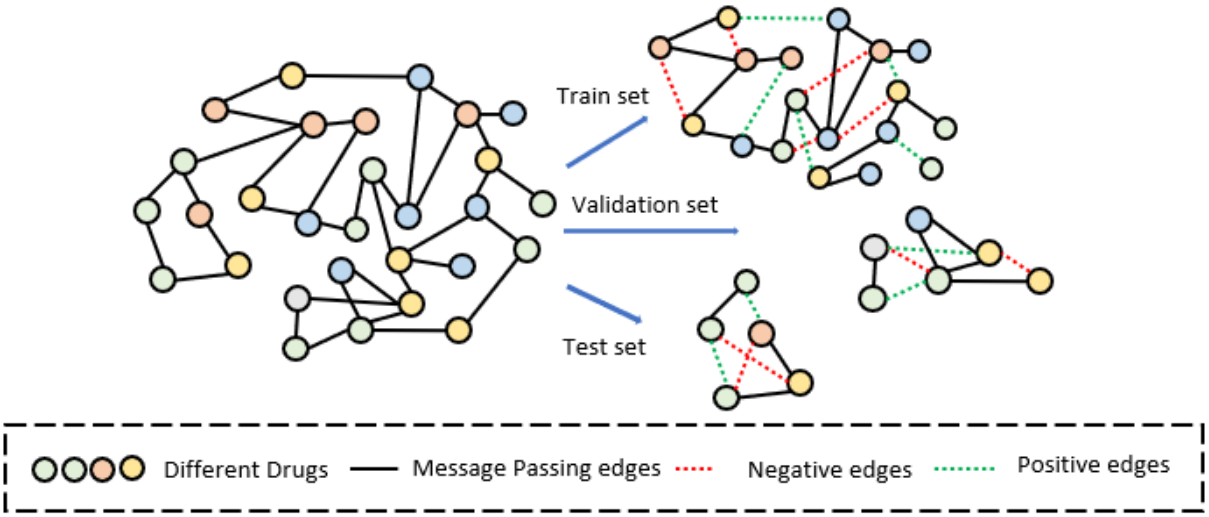

Fig. 2. Dataset split and negative sampling on edges.

B. Method Configuration

As discussed in the "Methods" section, DComG consists of 2 major components. The encoder with 3 GCN layers and dropout layers between each layer. And also, the decoder, which is a matrix factorization method and can be considered as the 4th layer of the model. After hyper-parameter tunning of DComG, the results indicate that three GCNConv layers result in the highest AUC scores. Each of the three GCNConv layers has an input and an output channel, therefore, in and out channels for all layers after hyperparameter tunning are as follows: (32, 128), (128, 128), (128, 64). In the DComG model, between each of the two layers of the GCN model, there exists a ReLu activation function that sends its output as the input of a dropout layer with a 0.5 dropout rate and the dropout layer sends its output as the input of the next GCNConv layer. The activation function available in the last layer is sigmoid (3):

$$f(x) = \frac{1}{1+e^{-x}} \quad (3)$$

Where x is the previous layer's output and f(x) is the embedding vector. The encoder's output is the new embedding space, which includes the embedding vectors for each drug. The matrix decoder then performs the mathematical operations necessary to decode all of the given embedding vectors in the mentioned embedding space. Matrix decoding is the fourth layer. Finally, the decoder produces an N * N matrix, where N is the number of drugs, which is 749. Each cell (i, j) in the matrix introduces a probability score between nodes i and j, where i is the row number and j is the column number. Higher scores are interpreted as a higher probability which exists an edge between every 2 drugs (The highest is 1 and the lowest is 0 after sigmoid operation). Binary-cross-entropy with (4) formula is the loss function used to implement the optimization problem with the help of the Adam optimizer [39].

$$L = -\frac{1}{N}\sum [y_i * log\,\bar{y}_i + (1 - y_i) * \log(1 - \bar{y}_i)] \quad (4)$$

Here, yi is in {0, 1}, which 1 denotes the existence and 0 non-existence of an edge between two nodes. i is the predicted value between 0 and 1 and L is the total loss based on the presence of an edge between all pairs of nodes. The training phase of the model is performed with backpropagation in which the loss computed from the decoded matrix is sent back to both the deep GCN model and the dot product decoder components of the model (Fig. 1). The performance of the model is significantly improved when all the trainable parameters are trained in this end-to-end manner as all parameters gain their gradients jointly from the loss function in the backpropagation phase [31].

C. Comparison with related work

We compared our DComG model with two of the best previously introduced state-of-the-art models performed on DCDB w.r.t AUC scores. The performance of DComG is assessed on 5-fold cross validation with negative sampling. DComG outperformed the previous state-of-the-art method (the AuDNNsynergy method [15]) in terms of AUC scores, with an average improvement of 4%. The comparison is available in Fig. 3.

D. Biological Interpretation

The ten drug pairs with the highest predicted probabilities are listed in Table I. These drug pairs do not exist in the original dataset but are predicted with a high probability to treat different diseases. The original drug combinations were gathered in DCDB and proposed in their paper in 2014 [17], Therefore some of the drug pairs shown here are predicted to be synergistic in the literature in recent years. J. Zhou et al. showed the antifungal activity of Artemisinin and its derivatives (Dihydroartemisinin, Artesunate, and Artemether) [23]. Q. Li et al. expressed that a combination of Lamivudine and Indinavir is used as a therapy against human immunodeficiency virus (HIV) [24]. R. A. Jonas et al. showed that Fotemustine or Cisplatin in combination with Sunitinib or Tamoxifen may be used in the future to treat uveal melanoma (UM) [25]. K. Kosilov et al. found that the efficacy and safety of taking Tadalafil, Solifenacin, and Dutasteride at the same time were beneficial in treating benign prostatic hyperplasia (BPH) with overactive bladder symptoms, as well as lowering urinary tract obstruction in previously unsuccessfully treated men [26]. Another study by D. L. Richardson et al., showed the improvement of women with recurrent ovarian cancer taking a combination of Paclitaxel and Pazopanib over the improvement of patients taking Paclitaxel and Placebo weekly [27].

V. Satyanarayana et al. demonstrated that combining Lisinopril, Simvastatin, Aspirin, and Hydrochlorothiazide resulted in lower cholesterol and blood pressure levels. It also increased adherence in patients who had at least one

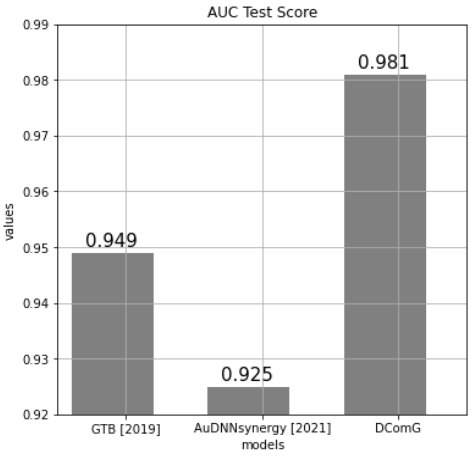

Fig. 3. Comparison of DComG with two of the best previously introduced models with highest AUC scores in the literature.

TABLE I. TOP PREDICTED SYNERGISTIC DRUG COMBINATIONS WITH DCOMG

Drug 1	Drug 2	Cited as Synergistic
Artesunate	Artemether	23
Lamivudine	Indinavir	24
Cisplatin	Fotemustine	25
Hydrochlorothiazide	Azellnidipine	-
Metformin	Vildagliptin	-
Solifenacin	Dutasteride	26
Paclitaxel	Pazopanib	27
Hydrochlorothiazide	Simvastatin	28
Docetaxel	Gemcitabine	29
Docetaxel	Axitinib	30

cardiovascular risk factor, such as coronary artery disease, hypertension, or dyslipidemia [28]. Several studies showed the antitumor activity of the combination of Gemcitabine and Docetaxel (GD) against various subtypes of sarcoma [29]. L. P. Martin et al. represented the effect of Docetaxel and Axitinib. In their study on the effect of Docetaxel, Axitinib, Paclitaxel, and Capecitabine pharmacokinetics, although there was no significant difference between the presence or absence of axitinib, Docetaxel exposure was increased in the presence of Axitinib [30].

V. DISCUSSION

Computational methods have been used to predict high synergistic drug pairs due to the enormous possibilities of drug combinations to avoid clinically experimenting with all of them, which is impractical in terms of time and cost, considering the enormous number of drug combinations. In this paper, we introduce DComG which is a graph auto encoder (GAE) model, that can predict synergy scores of drug combinations based on the PCA method performed on the external node2vec features extracted from drug-drug interactions. Although GCN models have been recently used in the field of Knowledge Graphs prediction and social network analysis problems, GCNs and GAE are fairly new in the field of predicting side effects caused by drug-drug interactions and drug-drug combination predictions.

Amongst the top predicted drug combinations in the new embedding space after decoding layer of the GAE model, many of them have been reported to be synergistic in the literature in recent years and some of them are open for future clinical research.

Although DComG outperforms other state-of-the-art models, there are few limitations in this study. It is limited to only the drugs that already have drug-drug interactions, which may cause some hidden associations in undiscussed drugs. Also, the size of the dataset is considered to be small. The study is performed on only 749 drugs and 1866 drug combinations available in the dataset.

VI. CONCLUSION

Computational drug combination prediction methods in recent papers are limited to the low number of drugs. They are mostly limited to specific drug indications like anti-cancer drugs. Also, the usage of GNN models as a new evolutionary approach in unstructured datasets is vacant in biological papers. In this paper, we propose a novel method for predicting synergistic drug combinations by developing a graph auto encoder (GAE) model. The encoder model is a deep GCN model and the decoder is a simple matrix factorization (dot product). This study showcases the important factor of DDI node features in drug combination prediction. Also, a dimension reduction technique on node features like PCA enhances the model performance and increases AUC Score between 0.5 to 1 percent. For our test set from 10% of the reported drug-drug combination dataset, we reached the mean AUC score over of 0.98, which outperforms the best of previous models by 4%. Remarkably, these numbers prove the great improvement of performance in predicting synergistic drug combinations compared to previously state-of-the-art machine learning and deep learning methods.

Regarding future research driven by our work, we find some particular research areas for refinement and extension of our research. First, to perform our method on more datasets and more drug features. Second, consider not only drug pairs but also a combination of more than two drug pairs as the input of DComG model. Third, to expand our research on drug-drug interaction (DDI) prediction (i.e. side effects prediction). We are currently focusing on performing DComG on more datasets.

Overall, considering drug combinations as a graph and a network of nodes and edges, and performing GCN operations on these graphs, can open new and effective opportunities for drug combination predictions.

Regarding the performance of DComG, the GCN models could be a useful resource for predicting new drug combinations with high potential of synergy and thus guide rational undiscovered medical treatments in the future.

REFERENCES

[1] R. G. Zinner *et al.*, "Algorithmic guided screening of drug combinations of arbitrary size for activity against cancer cells," *Mol. Cancer Ther.*, vol. 8, no. 3, pp. 521–532, Mar. 2009, doi: 10.1158/1535-7163.MCT-08-0937.

[2] H. Chen, S. K. Iyengar, and J. Li, "Large-scale Analysis of Drug Combinations by Integrating Multiple Heterogeneous Information Networks," in *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, Sep. 2019, pp. 67–76, doi: 10.1145/3307339.3342142.

[3] E. Guney, J. Menche, M. Vidal, and A.-L. Barábasi, "Network-based in silico drug efficacy screening," *Nat. Commun.*, vol. 7, no. 1, p. 10331, Apr. 2016, doi: 10.1038/ncomms10331.

[4] F. Cheng, I. A. Kovács, and A.-L. Barabási, "Network-based prediction of drug combinations," *Nat. Commun.*, vol. 10, no. 1, p. 1197, Dec. 2019, doi: 10.1038/s41467-019-09186-x.

[5] Y. Sun *et al.*, "Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer," *Nat. Commun.*, vol. 6, no. 1, p. 8481, Dec. 2015, doi: 10.1038/ncomms9481.

[6] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, and G. Yan, "NLLSS: Predicting Synergistic Drug Combinations Based on Semi-supervised Learning," *PLOS Comput. Biol.*, vol. 12, no. 7, p. e1004975, Jul. 2016, doi: 10.1371/journal.pcbi.1004975

[7] D. Chen, H. Zhang, P. Lu, X. Liu, and H. Cao, "Synergy evaluation by a pathway–pathway interaction network: a new way to predict drug combination," *Mol. Biosyst.*, vol. 12, no. 2, pp. 614–623, 2016, doi: 10.1039/C5MB00599J.

[8] Z. Wu, X.-M. Zhao, and L. Chen, "A systems biology approach to identify effective cocktail drugs," *BMC Syst. Biol.*, vol. 4, no. S2, p. S7, Sep. 2010, doi: 10.1186/1752-0509-4-S2-S7.

[9] J.-Y. Shi *et al.*, "Predicting combinative drug pairs via multiple classifier system with positive samples only," *Comput. Methods Programs Biomed.*, vol. 168, pp. 1–10, Jan. 2019, doi: 10.1016/j.cmpb.2018.11.002.

[10] L.-Y. Bai *et al.*, "Prediction of Effective Drug Combinations by an Improved Naïve Bayesian Algorithm," *Int. J. Mol. Sci.*, vol. 19, no. 2, p. 467, Feb. 2018, doi: 10.3390/ijms19020467.

[11] J. Li, X.-Y. Tong, L.-D. Zhu, and H.-Y. Zhang, "A Machine Learning Method for Drug Combination Prediction," *Front. Genet.*, vol. 11, Aug. 2020, doi: 10.3389/fgene.2020.01000.

[12] H. Liu, W. Zhang, L. Nie, X. Ding, J. Luo, and L. Zou, "Predicting effective drug combinations using gradient tree boosting based on features extracted from drug-protein heterogeneous network," *BMC Bioinformatics*, vol. 20, no. 1, p. 645, Dec. 2019, doi: 10.1186/s12859-019-3288-1.

[13] K. Preuer, R. P. I. Lewis, S. Hochreiter, A. Bender, K. C. Bulusu, and G. Klambauer, "DeepSynergy: predicting anti-cancer drug synergy with Deep Learning," *Bioinformatics*, vol. 34, no. 9, pp. 1538–1546, May 2018, doi: 10.1093/bioinformatics/btx806.

[14] F. Xia *et al.*, "Predicting tumor cell line response to drug pairs with deep learning," *BMC Bioinformatics*, vol. 19, no. S18, p. 486, Dec. 2018, doi: 10.1186/s12859-018-2509-3.

[15] T. Zhang, L. Zhang, P. R. O. Payne, and F. Li, "Synergistic Drug Combination Prediction by Integrating Multiomics Data in Deep Learning Models," 2021, pp. 223–238.

[16] P. Jiang, S. Huang, Z. Fu, Z. Sun, T. M. Lakowski, and P. Hu, "Deep graph embedding for prioritizing synergistic anticancer drug combinations," *Comput. Struct. Biotechnol. J.*, vol. 18, pp. 427–438, 2020, doi: 10.1016/j.csbj.2020.02.006.

[17] Y. Liu, Q. Wei, G. Yu, W. Gai, Y. Li, and X. Chen, "DCDB 2.0: a major update of the drug combination database," *Database*, vol. 2014, pp. bau124–bau124, Dec. 2014, doi: 10.1093/database/bau124.

[18] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 855–864, doi: 10.1145/2939672.2939754

[19] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," Sep. 2016, [Online]. Available: http://arxiv.org/abs/1609.02907.

[20] X. Li and Y. Cheng, "Understanding the Message Passing in Graph Neural Networks via Power Iteration Clustering," May 2020, doi: 10.1016/j.neunet.2021.02.025.

[21] T. N. Kipf and M. Welling, "Variational Graph Auto-Encoders," Nov. 2016, [Online]. Available: http://arxiv.org/abs/1611.07308.

[22] J. Zhou, J. Li, I. Cheong, N.-N. Liu, and H. Wang, "Evaluation of artemisinin derivative artemether as a fluconazole potentiator through inhibition of Pdr5," *Bioorg. Med. Chem.*, vol. 44, p. 116293, Aug. 2021, doi: 10.1016/j.bmc.2021.116293.

[23] Q. Li *et al.*, "Indinavir Alters the Pharmacokinetics of Lamivudine Partially via Inhibition of Multidrug and Toxin Extrusion Protein 1 (MATE1)," *Pharm. Res.*, vol. 35, no. 1, p. 14, Jan. 2018, doi: 10.1007/s11095-017-2290-4.

[24] R. A. Jonas, A. C. Rokohl, and L. M. Heindl, "Targeted therapy for malignant ocular melanomas," *Ann. Eye Sci.*, vol. 6, no. March, pp. 1–7, 2021, doi: 10.21037/aes-20-101.

[25] K. Kosilov, I. Kuzina, V. Kuznetsov, O. Barabash, and E. Fedorishcheva, "Efficacy of a combination of dutasteride, tadalafil, and solifenacin in the treatment of previously unsuccessful patients," *Asian J. Urol.*, vol. 9, no. 1, pp. 42–50, Jan. 2022, doi: 10.1016/j.ajur.2021.04.002.

[26] D. L. Richardson *et al.*, "Paclitaxel With and Without Pazopanib for Persistent or Recurrent Ovarian Cancer," *JAMA Oncol.*, vol. 4, no. 2, p. 196, Feb. 2018, doi: 10.1001/jamaoncol.2017.4218.

[27] V. Satyanarayana, G. Rajani, Kiranmai, K. Rupesh Reddy, and A. B. Raju, "Effect of fixed dose combination of Simvastatin, Aspirin, Hydrochlorothiazide and Lisinopril on blood pressure and cholesterol levels," *J. Pharm. Res.*, vol. 6, no. 4, pp. 452–455, Apr. 2013, doi: 10.1016/j.jopr.2013.02.031.

[28] H. Hara *et al.*, "Gemcitabine and docetaxel combination chemotherapy for advanced bone and soft tissue sarcomas: protocol for an open-label, non-randomised, Phase 2 study," *BMC Cancer*, vol. 19, no. 1, p. 725, Dec. 2019, doi: 10.1186/s12885-019-5923-7.

[29] L. P. Martin *et al.*, "Phase i study of axitinib combined with paclitaxel, docetaxel or capecitabine in patients with advanced solid tumours," *Br. J. Cancer*, vol. 107, no. 8, pp. 1268–1276, 2012, doi: 10.1038/bjc.2012.407.

[30] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," Jun. 2016, [Online]. Available: http://arxiv.org/abs/1606.09375.

[31] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," Jun. 2016, [Online]. Available: http://arxiv.org/abs/1606.09375.

[32] [15] T. Zhang, L. Zhang, P. R. O. Payne, and F. Li, "Synergistic Drug Combination Prediction by Integrating Multiomics Data in Deep Learning Models," 2021, pp. 223–238.

[33] [16] L. Wu *et al.*, "Machine learning methods, databases and tools for drug combination prediction," *Brief. Bioinform.*, vol. 23, no. 1, Jan. 2022, doi: 10.1093/bib/bbab355.

[34] B. Güvenç Paltun, S. Kaski, and H. Mamitsuka, "Machine learning approaches for drug combination therapies," *Brief. Bioinform.*, vol. 22, no. 6, Nov. 2021, doi: 10.1093/bib/bbab293.

[35] A. Ling and R. S. Huang, "Computationally predicting clinical drug combination efficacy with cancer cell line screens and independent drug action," *Nat. Commun.*, vol. 11, no. 1, p. 5848, Dec. 2020, doi: 10.1038/s41467-020-19563-6.

[36] J. Li, X.-Y. Tong, L.-D. Zhu, and H.-Y. Zhang, "A Machine Learning Method for Drug Combination Prediction," *Front. Genet.*, vol. 11, Aug. 2020, doi: 10.3389/fgene.2020.01000.

[37] S.S. Ziaee, https://github.com/sinaziaee/gnn_drugs, 2022.

[38] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6980.

[39] M. Zitnik et , M. Agrawal, and J. Leskovek, "Modeling polypharmacy side effects with graph convolutional networks," Bioinformatics., vol. 34, July. 2018, doi: 10.1093/bioinformatics/bty294.