Visualization of drug interactions and drug interactions prediction task overview.

Moadine Fatima Zohra: National School of Computer Science and Systems Analysis (ENSIAS), Morocco.

Abstract:

In everyday life, we try to cure diseases by combining multiple drugs[1], and so the need for more analysis of drug interaction arises. For this purpose, the interaction between drugs was extracted from the dataset DDI_data[2], the latter is composed of five columns: drug1_id, drug2_id, drug1_name, drug2_name, and interaction type. Networkx and pandas libraries were the libraries most used for preprocessing the dataset from column extraction to drug preferences. After preprocessing, networkX generates a graph of 772 nodes and 997 edges, and a network density of 0.0033. The NetworkX greedy_modularity_communities() function detects three communities, non-overlapping and overlapping. The drug most frequently involved in drug interactions is Simvastatin, with a node degree of 402. Afterwards, we introduce some methods of drug interactions prediction task[3]. Studying drug-drug interactions helps identify and improve new drug combinations and their effects.

Index terms: Drug-drug interactions, NetworX, DDI, drug-drug interaction prediction, machine learning.

- - - - - - - - - - ◆ - - - - - - - - - -

1. Introduction:

Human beings suffer multiple diseases, to cure the latter we are in need of a combination of drugs([4-5]), a good combination with no side effects, or better the best combination. Research in this field is very expensive and time consuming without taking into consideration human limitation of analysis of big data compared to supercomputers. Therefore drug drug interaction network analysis is a much important study area to get more attention, in addition to that getting more insights from graphs and understanding more the workflow of drugs.

2. Materials and methods:

2.1. Data acquisition:

In order to analyze the interactions between drugs using network analysis, we searched for data from the website Mendeley data (https://www.drugbank.ca/; version 5.1.4, released 2019-07-02), which includes 2634 drugs The raw data is stored as a CSV file of 222696

rows and 5 columns, where each row represents five items: Drug1_id, Drug2_id, Drug1_name, Drug2_name and the interaction type between a particular pair of drugs. The Pandas Dataframe looks like this:

	drug1_id	drug2_id	drug1_name	drug2_name	interaction_type
0	DB00006	DB00346	Bivalirudin	Alfuzosin	serum concentration
1	DB00006	DB13783	Bivalirudin	Acemetacin	risk or severity of bleeding
2	DB00006	DB06605	Bivalirudin	Apixaban	anticoagulant activities
3	DB00006	DB06695	Bivalirudin	Dabigatran etexilate	anticoagulant activities
4	DB00006	DB09075	Bivalirudin	Edoxaban	anticoagulant activities

Fig 1: Sample rows of processed drug interactions data

2.2. Subset Data:

For more understandable network analysis and visualization, we narrow the dataset to a small one [6], and we add a column of weight with value 1 meaning equal severity for all interactions. Fig 2 shows a sample.

0	drug1_name	drug2_name	weight
0	Bivalirudin	Simvastatin	1
1	Goserelin	Metformin	1
2	Daptomycin	Simvastatin	1
3	Cyclosporine	Simvastatin	1
4	Cyclosporine	Lisinopril	1

Fig 2: sample rows of processed drug interactions data.

2.3. Network analysis:

Degree Centrality

By calculating degree centrality we try to find key nodes in the network, the degree of a node refers to the number of connections the node has. In other words, the degree centrality of a node is the number of edges it has. To calculate degree centrality, we use the networkX function Graph.degree().

Communities in the network

Many algorithms exist for community detection, networkx provides [7]. In this application we used the The NetworkX greedy_modularity_communities() function to detect communities based on ……..

2.4. Related work drug interaction prediction:

Drug drug interaction (DDI) prediction task is divided into 3 categories[3]**:** literature-based extraction, machine learning-based prediction and pharmacovigilance-based data mining methods. Deep learning is used in 2 categories, the first one is extracting DDI interaction from text using deep learning nlp techniques, or use it as a machine learning based method by generating strong features to extract novel DDI interactions.

3. Results and Discussion:

For more readability of the graph, we took the first 100 rows of the dataset, an undirected graph is generated (figure 3). The network density is very low meaning there is so fwe connections between drugs.

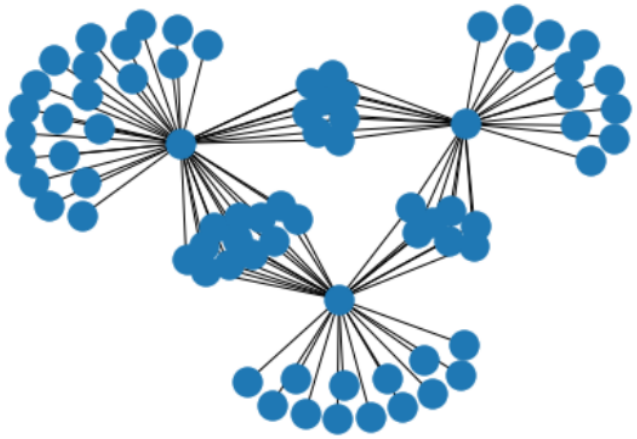

Fig 3: graph network of drug drug interactions with 76 nodes and 100 edges and a Network density of 0.035.

We obtained the top 20 most important nodes as shown in figure 4, based on the number of edges of each node. Simvastatin, Lisinopril and Metformin are most connected in the network.

```
Top 20 drugs by degree:
('Simvastatin', 41)
('Lisinopril', 32)
('Metformin', 27)
('Cyclosporine', 2)
('Fluvoxamine', 2)
('Ramipril', 2)
('Amphetamine', 2)
```

Fig 4: degree centrality of drugs.

Three communities were detected using the function greedy_modularity_communities() (figure 5 and 6), in figure we noticed the importance of drugs Simvastatin, Lisinopril and Metformin as they connect communities and keep them intact. We also see overlapping between some communities.

Fig 5: detection of 3 communities of drugs, each community is represented in a different color.

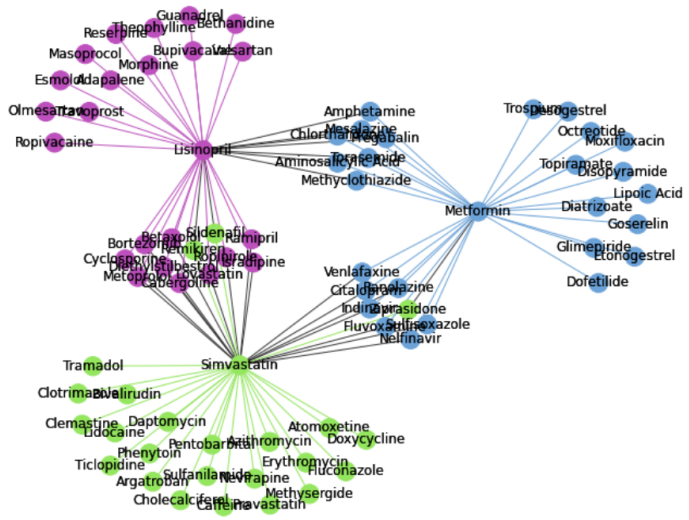

Fig 6: detection of 3 communities of drugs, each community is represented in a different color.

4. Conclusion:

In this work , we tried to give a starting point of how to create a graph network from a specific dataset, here we had drug drug interactions. We applied statistics on DDI as degree centrality and visualized the graphs for more human intuition to be involved visually. For future work we can add severity between pairs of drugs to explore more potentials from the graphs. As we have seen, machine learning based methods and nlp techniques can enhance the domain of drug drug interaction prediction.

References:

[1] “Drug-Drug Interaction,” *TDC*. http://localhost:4000/multi_pred_tasks/ddi/ (accessed May 25, 2022).

[2] H. Yu, “data of multiple-type drug-drug interactions,” vol. 1, Jul. 2020, doi: 10.17632/md5czfsfnd.1.

[3] Y. Qiu, Y. Zhang, Y. Deng, S. Liu, and W. Zhang, “A Comprehensive Review of Computational Methods for Drug-drug Interaction Detection,” *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. PP, May 2021, doi: 10.1109/TCBB.2021.3081268.

[4] Y.-H. Feng, S.-W. Zhang, and J.-Y. Shi, “DPDDI: a deep predictor for drug-drug interactions,” *BMC Bioinformatics*, vol. 21, no. 1, p. 419, Sep. 2020, doi: 10.1186/s12859-020-03724-x.

[5] G. Lee, C. Park, and J. Ahn, “Novel deep learning model for more accurate prediction of drug-drug interaction effects,” *BMC Bioinformatics*, vol. 20, no. 1, p. 415, Aug. 2019, doi: 10.1186/s12859-019-3013-0.

[6] K. Leung, *Drug Drug Interactions - Network Analysis and Visualization*. 2022. Accessed: May 25, 2022. [Online]. Available: https://github.com/kennethleungty/Drug-Interactions-Network-Analysis-and-Visualization/blob/1f0ddc7e868e03b75e6e16b3601ad78b8b9934c8/02_Network_Analysis_and_Visualization.ipynb

[7] “Communities — NetworkX 2.8.2 documentation.” https://networkx.org/documentation/stable/reference/algorithms/community.html (accessed May 25, 2022).