Polypharmacy Side-Effect Prediction with Graph Neural Networks

# Overview/Abstract:

Significant number of diseases directly or indirectly require combination of multiple drugs for treatment. Usage of multiple drugs can significantly alleviate the chances of adverse drug reaction event (ADE), hindering the treatment of primary underlying condition. Hence, it is utmost import to study the ADE, develop methodologies to predict the ADE and take preventative actions to avoid ADE. In this study we present a graph neural network (GNN) model that processes three types of interactions namely, drug to protein (in humans) interactions, protein to protein interactions, and drug to drug interactions, to detect an ADE corresponding to a pair of drugs. We used publicly available dataset XXX to train our model and achieved XXX% of accuracy in detection of an ADE corresponding to a given pair of drugs. The GNN model proposed in this study improves the accuracy by XXX% compared to a basic neural network classifier. With our results we conclude the proposed GNN model efficiently detects the critical patterns in the inter and intra-connections between drug and proteins, to accurately predict and ADE.

# Introduction:

# Problem:

In this study we specifically study the problem of detecting if an ADE occurs for a given pair of chemicals. Moreover, we also extend our problem to detect the type of ADE if an ADE occurs. In other words, given following information,

|  |  |
| --- | --- |
|  | A graph of protein-to-protein interactions in humans where,is set of unique proteins[[1]](#footnote-1) in humans and is the set of connections between proteins. This graph is unweighted and undirected graph. |
|  | A graph of chemical-to-chemical interactions (reported events of ADEs) where, is the set of unique chemicals[[2]](#footnote-2) and is the set ADE corresponding to two chemicals. Let . This is an undirected graph with links having multiple features. |
|  | A graph of chemical-to-protein interactions (reported in previously published studies). This is a directed and unweighted graph. |

We develop a model such that,

and

We express the objective problem of interest as follows,

Where, is the set of model parameters and is the objective function to be minimized. We express as follows,

Equation 1: Objective function

Where,

is a negative sample of a chemical, sampled from set with a distribution

# Data:

We use data published by XXX, which has three main components. These three components, as previously mentioned, are protein-to-protein interactions, drug-to-protein interactions, and drug-to-drug interactions. We will discuss each one in detail in the following sections.

## Protein-to-protein interactions:

The in the Problem section is the protein-to-protein interactions data. The data is read as a graph with nodes () representing unique proteins and edges representing the experimentally validated physical reactions between a pair of proteins. Authors of Zitnik et.al. have compiled this dataset from multiple previously published and publicly available datasets (menche, aryamountri, string database, proteome scale map). This graph is undirected and unweighted. There are over 19,000 proteins and over 700,000 physical interactions documented in this graph. The node-degree distribution of this graph follows the power law with a mean value of 75 and median of 33. Eight components exist in the graph, the largest component subsume 99.9% of nodes while the others are small, disconnected sets of nodes.

## Drug-to-drug interactions:

The in the Problem section is the chemical-to-chemical (or drug-to-drug) interactions data. Authors of Zitnik et.al. have combined information from SIDER, OFFSIDES and TWOSIDES datasets. SIDERS dataset is created by mining ADEs from drug label text, while OFFSIDES and TWOSIDES datasets are created from reports from patients, doctors, and drug-companies on ADEs. Nodes () in the final combined data represents unique chemical/drug and the edges (or ) between drug represents the category of the side effect occurring for the pair of drugs. There are 645 unique chemicals and over 63,000 side-effects documented in the final dataset. All side-effects present in the dataset belong to 561 unique categories. The number of chemicals (number of nodes) in the graph exponentially reduce with increase in the number of side-effects associated with the chemical i.e., there are very few chemicals with more than 300 side-effects associated it.

## Drug-to-protein interactions:

# Methodology:

# Results:

# Discussion/Observations:

# Conclusion:

1. : Note that the set is not exhaustive [↑](#footnote-ref-1)
2. : Note that the set is not exhaustive [↑](#footnote-ref-2)