

CS224W Project Report: Drug Recommendation System to Minimize Polypharmacy Side Effects

Lea Jabbour, Marcos Torres, and Alex Wade

Abstract—Dealing with drug-drug interactions is time-consuming and costly. We synthesize work on drug-drug similarity network construction and process drug-drug interaction networks to construct a program that allows for automatic substitution of problematic drug pairs with combinations that have similar function but less severe interactions. First, we use a random walk algorithm on a protein-protein interaction network via drug-protein interactions to measure functional drug similarity. Then, we process existing drug-drug interaction networks to categorize interactions in terms of severity. We then combine these two to recommend pairs of drugs with high similarity to a problematic pair, but with less severe interactions.

I. INTRODUCTION

Patients with complex diseases or comorbidities often need to use combinations of drugs, which is known as polypharmacy. Polypharmacy is particularly common for the elderly; over 35% of adults aged 65 or older take between 5 and 10 prescription drugs [1]. Drugs used in combination can modulate protein activities, which can lead to improved or worsened outcomes. These drug-drug interactions (DDIs) that occur when using a drug combination lead to higher risk for side effects.

Since drugs are generally studied individually in clinical trials, it is very difficult to predict polypharmacy side effects. It would also be very costly and time consuming to investigate pairwise DDIs in a laboratory setting, let alone n -drug combinations. Treating polypharmacy side effects costs nearly \$180 billion a year in the U.S. [1], and the number of people taking drug combinations is increasing year to year. The impact on medical institutions is also serious. For example, 28% of admissions in a Dutch hospital encountered at least one DDI [2].

There is therefore a critical need to predict DDIs before prescribing drug combinations that may

have adverse effects on the patients and would be expensive to treat. Furthermore, there is a need to use DDI predictions to aid physicians in selecting better drug combinations to effectively treat patients, while minimizing the possibility of serious adverse side effects.

We propose a project that builds off of existing work on DDI prediction and on drug similarity networks. We aim to combine previous approaches in order to recommend drug pairs with high similarity to the originally requested drug pair, but with negligible side effects.

II. RELATED WORK

There has been much work on constructing drug-drug similarity metrics, and on predicting drug-drug interactions, but little work on suggesting viable alternatives to problematic drug combinations.

A. Chen et al. 2012, Assessing Drug Target Association Using Semantic Linked Data [3]

To give an example of work on drug-drug similarity, Chen et al. (2012) constructed a heterogeneous semantic network relating drugs, chemical compounds, protein targets, diseases, side effects, and pathways for use in drug-target interaction prediction [3]. They then used the results from their network to construct a drug-drug similarity network that hinted at drug interactions.

B. Park et al. 2015, Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks [4]

Another approach was taken by Park et al. (2015), who used a random walk with restart algorithm on a protein-protein interaction network

to model drug similarity. Their main insight was that if a drug targets a protein, they should let the influence of that protein diffuse through the protein-protein network to get a true sense of the drugs influence [4]. Subsequently, they compare the spread of each drugs influence through the protein-protein interaction network to determine the similarity between two drugs.

This method improved on earlier attempts to measure drug similarity through functional comparison. Chen et al. (2012) mention that functional comparison yields more meaningful drug similarity evaluations than structural comparison, Huang et al. (2013) [5] improve on the methods of Chen et al. (2012), and in this paper, Park et al. (2015) show that their method yields more impressive ROC curves than the technique of Huang et al. (2013) with known drug-drug interactions as a metric, with AUC of 0.842 on DrugBank, as compared to 0.786. Although ensemble models combining various techniques, some non-network, have been proposed by Zhang et al. (2017) [6], the method of Park et al. (2015) is the state of the art for a single network approach to drug similarity based on a protein interaction network.

C. Zitnik et al. (2018), *Modeling polypharmacy side effects with graph convolutional networks* [7]

Zitnik et al. (2018) proposed a model, Decagon, to predict specific polypharmacy side effects, rather than just the possibility of any side effect existing [7]. The authors formulated the problem as a multirelational link prediction problem in a two-layer multimodal graph G consisting of drug and protein node types. The network was constructed from protein-protein, drug-protein, and drug-drug interaction data. Unlike previous approaches which aimed to predict whether or not two drugs would interact (binary classification), Decagon aimed to determine whether a pair of drugs would interact with a given side effect type. Their method outperformed alternative approaches by up to 69%, and resulted in a 20% average gain in predictive performance. This approach also allowed the authors to find novel predictions.

D. Opportunity for novel work

For our project, we use drug-drug similarity and known drug-drug interactions to suggest al-

ternatives to problematic drug combinations. We therefore turn to the work done by Park et al.(2015) - which determines drug similarity more accurately than previous structural approaches - for the construction of our drug-drug similarity metric. While the authors mention that their results could be used to avoid bad drug combinations, they don't elaborate on this finding.

Zitnik et al. (2018) predicted specific drug-drug interactions very successfully in their Decagon model, but for our project we will instead use the Decagon drug-drug interaction (DDI) dataset as a reference to avoid interactions, rather than attempting to predict them ourselves.

III. APPROACH

A. Decagon Dataset

We used BioSNAP data, in particular the Decagon data (<http://snap.stanford.edu/decagon/>), to find drug-drug similarity. The relevant Decagon files for this part of the project are the drug-target interaction (DTI) network and the protein-protein interaction (PPI) network. There DTI network has data on 1774 drugs and 7795 proteins, while the PPI network has 19081 proteins. We did not condense the PPI network to just the 7795 proteins in the drug-target network, since our algorithm involves a random walk along the PPI network, and can therefore account for the influence of proteins in the PPI network even if they are not explicitly drug targets.

As our basis for determining side-effects between drugs, we use the Decagon drug-drug interaction (DDI) network, which models polypharmacy side-effects. In this network, nodes represent drugs, and edges represent the various types of side effects associated with the drug pair, which cannot be attributed to either individual drug in the pair. This dataset appears to be state of the art in determining DDIs, and was created from the third paper we reviewed in this report, Zitnik et al. (2018) [7]. The network contains 63,473 interactions across 645 drugs, and was generated based on national adverse event reporting systems.

B. TWOSIDES Dataset

We also utilized the TWOSIDES database, which the Decagon drug-drug interaction dataset

was built off of. This dataset was created by Tatonetti et al. (2012) [8], and contains information on polypharmacy side effects for pairs of drugs, as well as the likelihood of the interaction. In our approach, we used both the TWOSIDES and Decagon datasets so that we could compare results. This is important because the Decagon dataset is itself a model of the TWOSIDES data, and because we wished to use interaction statistics to refine our model. TWOSIDES contains 868,221 significant associations between 59,220 pairs of drugs and 1301 adverse events (side effects). The relationships that are included in this database are limited to those that cannot be clearly attributed to either drug alone. The database also includes 3,782,910 associations for which the drug pair has a higher proportional reporting ratio (PRR) than either individual drug alone. The PRR measures the extent to which a particular adverse event is reported for individuals taking a specific drug, compared to the frequency at which the same adverse event is reported for patients taking some other drug. This database was suitable for our algorithm because the reported adverse events are likely caused by the interaction of the drugs, rather than either drug alone - which is exactly what we are interested in. A very important feature of the TWOSIDES dataset, is that it contains a "confidence" level for each interaction it reports. The confidence level is assigned based on the p-value of the reported interaction, and ranges from 1 to 5 (5 corresponding to lowest p-value, and therefore highest confidence).

C. Methods

1) *Pre-processing Decagon Dataset:* We initially wanted to pre-process the Decagon model built by Zitnik et al. (2018) to include a heuristic for side effect severity. The Decagon model considers the 1317 types of polypharmacy side effects, and we had wished to assign each side effect a score indicating its severity (e.g. a minor side effect like dandruff would get a lower score, while a severe side effect like cardiac arrest would get a higher score). However, this heuristic would require significant medical and pharmacy expertise, so for this project we used a simpler heuristic. We will refer to the latter heuristic as the

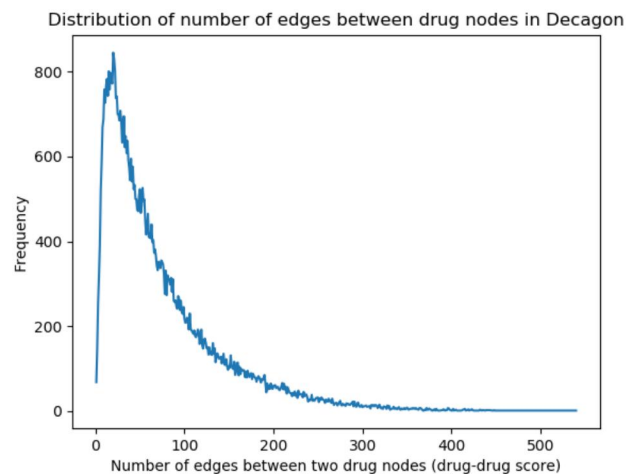


Fig. 1. Distribution of edge count between two drug nodes in Decagon drug-drug interaction network

drug-drug score. Instead of weighting each edge by the severity of the side effect, we used the total number of edges between two nodes as a heuristic for drug-drug interaction (DDI) severity. We chose this as an initial heuristic, because a larger number of potential side effects between two drugs would result in a larger drug-drug score, indicating a less favorable drug combination - as desired. Figure 1 shows the distribution of the drug-drug score in the Decagon network. We can see that the majority of drug pairs have about 50 edges between them, and that the frequency decreases as the drug-drug score increases.

2) *Pre-processing TWOSIDES Dataset:* After initial analysis of the Decagon dataset, we noticed that many drugs that are commonly taken together (e.g. aspirin and ibuprofen) had many listed side effects in Decagon. We therefore wanted to include some measure of the likelihood of side effects occurring, since many side effects were likely rare and would disproportionately affect our recommendations. Therefore, we decided to use the TWOSIDES dataset, because it contained data on the confidence of each interaction. Based on the p-value, 'confidence' ranged from 1 to 5, and a higher confidence level indicated that the interaction occurs with more certainty. We therefore pre-processed the TWOSIDES dataset to only include interactions with a confidence level of 5 (maximum), and plotted the drug-drug scores

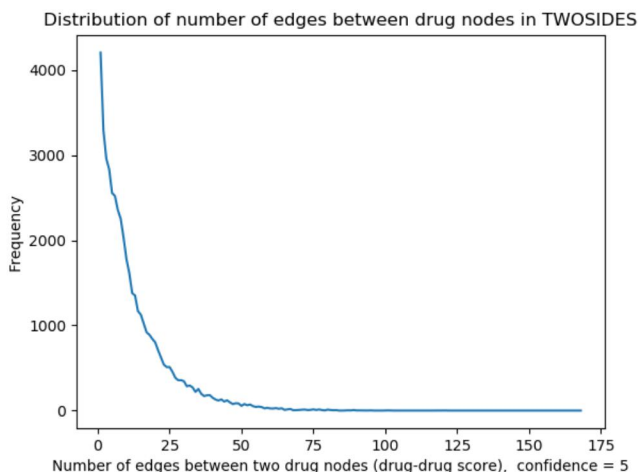


Fig. 2. Distribution of edge count between two drug nodes in TWOSIDES drug-drug interaction network, with interaction confidence equal to 5

using the same methods as before (Figure 2). We can see from Figure 2 that the distribution has the same overall shape as that of Decagon network, but that the maximum drug-drug similarity score is about three times smaller than it was in Decagon’s distribution. The overall graph also looks a bit smoother here, and may have less noise compared to the Decagon dataset.

The drug-drug score heuristic can therefore be used based on the Decagon DDI network, and/or the TWOSIDES network. We will report results from both for comparison.

3) *Recommendation Algorithm*: Next, given a query in the form of a pair of drugs, we reference a polypharmacy side-effect association network to get a heuristic of the drug-drug score (using either Decagon or TWOSIDES). We then search for the five most similar drugs to each drug in the input pair using a random walk with restart algorithm. Finally, we try all combinations of the resulting similar drugs and recommend the one with the lowest drug-drug score, which is a heuristic for lower DDI rate.

We have implemented the random walk with restart (RWR) algorithm described in Park et al. (2015) to calculate drug similarity. We first parsed the drug-target network file as well as the PPI network file and stored them for future uses.

Given a single drug, the RWR algorithm over the PPI network consists of the following: For each

timestep t , we calculate probability vector $p(t+1)$, whose values are the probability of the random walker being at a given protein at the timestep, as

$$p(t+1) = (1-r) \cdot M \cdot p(t) + r \cdot p(0) \quad (1)$$

where r is the restart probability (here 0.7 as in Park et al. (2015)), M is the adjacency matrix with rows normalized, $p(t)$ is the probability vector from the previous timestep, and $p(0)$ is the initial probability vector, consisting of a $\frac{1}{n}$ for each target of the drug in question and a 0 for each protein that is not a target, where n is the number of targets the drug has.

We terminate the walk when the sum of the absolute values of the differences between $p(t+1)$ and $p(t)$ is smaller than an epsilon 1.0×10^{-5} .

Then, we calculate the similarity between the two drugs by taking the element-wise square root of the product of the vector values of each protein for the two drugs A and B:

$$ProteinScore_i(A, B) = \sqrt{PR_i(A) \cdot PR_i(B)} \quad (2)$$

Then we simply sum up the protein scores over all proteins to get a metric of drug similarity:

$$DDIScore(A, B) = \sum_{i=1}^N ProteinScore_i(A, B) \quad (3)$$

Overall, the algorithm is quite similar to PageRank with teleportation, where the authors use protein influence as a proxy for drug influence and ultimately drug similarity.

In order to implement this algorithm, we chose to use the Python defaultdict data type to represent our sparse vectors and matrices in the PPI network, and implemented various operations such as normalization and dot product over this data type ourselves.

Our goal for this section was to input a pair of drugs (drugA, drugB), and output the five most similar drugs to drugA, and the five most similar drugs to drugB. We wanted our algorithm to be robust enough that we could test any pair of drugs in the evaluation portion, but the RWR algorithm was computationally intensive, and we did not have knowledge *a priori* of which drugs would be similar to one another (and therefore could

not prune any drug nodes). Therefore, we pre-computed the RWR vector for each of the 1774 drugs in the network, and saved all of these in a pickle file for future use (using Python’s pickle package).

Next, we wrote a function that would calculate the similarity score between drugA and all other drugs, using the saved RWR vectors, and return the five drugs with highest DDI score to drugA. Again, this was a computationally intensive step, so we saved results to a file. Furthermore, from literature searches, we noticed that a single drug is likely to interact with multiple other drugs, and thus we kept a cache to prevent from recomputing similarity rankings.

Finally, once we had the five most similar drugs to drugA and drugB, we used our heuristic to calculate the drug-drug score between all combinations between drugA and its five most similar drugs, and drugB and its five most similar drugs (36 combinations). For this section, we used the pre-processed drug-drug scores. Since the drug-drug score was used as a heuristic for severity of drug-drug interactions, we then recommended the drug pair with the lowest drug-drug score because it would theoretically minimize the adverse events from drug-drug interactions.

IV. RESULTS AND DISCUSSION

A. Software

Our deliverables include 1GB of computed similarity scores between all 1774 drugs for which we have information in the drug-target network in Decagon. We also produced software in Python for processing the Decagon and TWOSIDES database files into internal representations, caching our computed RWR vectors and similarity scores as Python pickle files, and creating databases of our heuristics for drug-drug interaction severity.

We also produced text files of the most similar drugs to the drugs in our 25 sample pairs with their scores, and all the possible pairs of those alternatives with their scores as per our heuristics. Finally, and most importantly, we created Python software that computes drug-drug similarity using the RWR algorithm on the PPI network, and calculates alternatives given a pair of drugs, the

similarity algorithm, and DDI files containing the heuristic scores.

Our repository may be found at <https://bitbucket.org/drugdrug224w/drugdrug/src> and a backup is at <https://github.com/alexjwade/drugdrugrecommender>.

B. Evaluation

As our test set of drug pairs, we first did a literature review and selected 15 pairs of drugs that are known to be problematic. Next, we used our data and our drug-drug score heuristic to select 10 pairs of drugs with the highest number of interactions in the TWOSIDES database. Note that there was no overlap between the 15 pairs from the literature review and the TWOSIDES database.

From the literature review, we chose 15 pairs of drugs which have problematic interactions between them, based on Kheshti et al. (2016)[9], which asked trained pharmacists for their knowledge about the pairs. We believed that this would represent our typical test case, the problem we wished to solve. However, it turned out that the Decagon drug-target networks we use for our RWR algorithm lacked information on many of the drugs from these pairs. When we only included pairs that were present in the data we used to construct drug-drug similarity, the number of pairs dropped to 11. We therefore added in another 4 drug pairs from another article, Roe et al. (2016), on harmful drug-drug interactions.

Next, we wanted to use our data to find potentially problematic pairs. For this method, we used our pre-processed TWOSIDES dataset, which had the count of side effects (with confidence level 5) between drug pairs. Using this again as a heuristic for the severity of drug-drug interactions, we found the top 10 pairs with the highest drug-drug scores, and used those as input to our algorithm. Again, some of these drug were not present in the drug-target network we use in the RWR algorithm. Therefore, we used the top 10 drug pairs which were present in the drug-target database. This meant that we had to search the the top 27 pairs in our database according to our metric, because 17 of those pairs did not exist in our drug-target data.

Our algorithm first found the 5 most similar drugs to each of the drugs in the pair, then calcu-

lated the severity of interactions between the drugs most similar to each, and finally proposed a pair of similar drugs with low interactions to recommend in lieu of the problematic pair.

Following the steps of our algorithm, we first evaluate our results in terms of similarity, followed by our drug-drug interaction metrics, and finally the quality of the recommended pairs.

1) *Drug similarity evaluation*: We hoped to compare our results directly to the implementation of Park et al. (2015)[4]. They have a website (<http://biosoft.kaist.ac.kr/targetrw/index.html>) which allows users to find similarity scores between drugs. Unfortunately, their website did not have a similarity score between acetaminophen and erythromycin.

For a sanity check to ensure that our results are approximately correct, we compared acetaminophen and aspirin, two painkillers with similar chemical structures and effects, hoping to see a high similarity score. The similarity is 0.93578 in our implementation.

We also compared aspirin to glyburide, which is used in the treatment of Type II diabetes and has a very dissimilar chemical structure to aspirin. The similarity is 0.19818 in our implementation, which contrasts markedly with the score for acetaminophen and aspirin, thus confirming our sanity check.

Besides comparing our results directly to Park et al. (2015), we compare our results to those of Brown et al. (2016)[10], who used a literature-based method to find similarities between FDA approved drugs. Importantly, they have an implementation of their algorithm online at <http://apps.chiragjgroup.org/MeSHDD/>. We chose to compare acetaminophen and its most similar drugs according to MeSHDD using our algorithm. Unfortunately their online implementation does not include dissimilar drugs, only similar ones, but we tried finding random drugs (glyburide and erythromycin), computing the similarity score, and seeing whether they were listed as being similar to acetaminophen on MeSHDD (they weren't). (Note that MeSHDD scores correspond to distance, whereas our scores correspond to similarity.) Table I summarizes the scores for the sample drugs, and targetRW corresponds to Park et al. (2015)'s

scores.

TABLE I

SAMPLE DRUG SIMILARITY SCORES WITH ACETAMINOPHEN

Drug name	1 - MeSHDD score	targetRW	Our score
Ibuprofen	0.222	2.112	0.764
Naproxen	0.165	2.0	0.738
Codeine	0.144	0.178	0.226
Ketoprofen	0.162	2.026	0.247
Erythromycin	N/A	N/A	0.205
Glyburide	N/A	0.168	0.197

Additionally, MeSHDD did not have aspirin listed as similar to acetaminophen. The scores for codeine and ketoprofen are rather low on our end but still higher than the similarity scores for erythromycin and glyburide.

Our results follow those of targetRW (from Park et al. (2015)) fairly closely apart from our low score for ketoprofen, which is much lower than the targetRW score. Since we followed the same algorithm as targetRW, this discrepancy in results must be due to the differences in the datasets we ran our algorithms on. We hypothesize that the Decagon data we used had fewer links (direct and indirect) between acetaminophen and ketoprofen. Ultimately, though, our results were roughly similar and our dissimilar drugs did have lower scores.

When we put the similarity algorithm into practice, we noticed a strange feature in the most similar drugs that came up. Many of them had the same ID, but beginning with a 1 instead of a 0. Technically, these are different molecules, with vastly different structures, and of the ones we examined, not even licensed medications. It seems that the same drug is listed twice in the drug-target database, but with some of its mentions preceded by a '1' for some reason. In our similarity algorithm, we ended up having to throw these strange molecules away as it seemed that they were clones of the actual drugs we wanted to examine, a glitch in the database.

We then analyzed the top 5 similar drugs returned by our algorithm to see if our algorithm truly returned drugs with similar functions. We analyzed the distinct drugs in the 30 drug pairs we had previously tested. For each drug, we extracted the drug in its list of 5 most similar drugs with the highest similarity score (ignoring the drugs with the problem mentioned above). Next, we

researched the drug and its most similar drug, and noted the results. Finally, we reported a representative sample in Table II of the types of relations we found.

TABLE II
DRUG SIMILARITY ANALYSIS

Drug	Most Similar Drug	Sim Score	Observations
Colchicine	Vincristine	1.16	Colchicine is an anti-inflammatory drug that can treat and prevent gout attacks, and vincristine is a chemotherapy drug. Both bind to tubulin, and inhibit cell growth.
Ergotamine	Bromocriptine Mesylate	1.08	Ergotamine and Bromocriptine Mesylate are closely related to ergoline and dopamine agonists.
Epinephrine	Ciprofloxacin	1.08	Epinephrine can treat severe asthma attacks and allergic reactions, and ciprofloxacin treats infections. While they act on different systems, they both seem to be involved in immune responses.
Carvedilol	Fentanyl	1.07	Carvedilol can treat high blood pressure and heart failure, and fentanyl can treat severe pain. While they act on different systems, they might have overlapping mechanisms.
Fluoxetine	Citalopram	1.05	Fluoxetine is a Selective Serotonin Reuptake Inhibitor (SSRI). It can treat depression, obsessive-compulsive disorder (OCD), bulimia nervosa, and panic disorder. Citalopram is also an SSRI used to treat depression.
Haloperidol	Asenapine	1.05	Haloperidol is an antipsychotic used to treat certain types of mental disorders. Asenapine is also an antipsychotic, often used to treat schizophrenia and acute mania associated.
Trimethoprim	Triamterene	1.03	Triamterene is an antibiotic for bladder infections. Trimethoprim is a diuretic. Interestingly, both seem to affect renal system.
Simvastatin	Etravirine	0.83	Simvastatin is used to treat cholesterol. Etravirine is an HIV drug (NNRTI). These drugs seem fairly different.

We see that drugs with high similarity (higher sim score) tend to either affect the same organ system or have similar targets, effects, or molecular structure. Generally, higher similarity score indicates a closer match, though there do seem to be nonsensical recommendations. We did note that many of these nonsense pairings included an antibiotic. It is possible that, since antibiotics tend to target bacterial proteins, the Decagon data is not well suited because it primarily contains

human protein interactions. One example of a good similarity matching is that of fluoxetine's and citalopram. Both drugs are Selective Serotonin Reuptake Inhibitors (SSRIs), which are used to treat depression. One surprising and interesting match is that of trimethoprim and triamterene. While trimethoprim is used to treat bladder infections, triamterene is a diuretic, which increases production of urine. At first glance, these drugs seem to act differently, but they both act on the renal system. Finally, one match that seems irrelevant is that of simvastatin and etravirine. Simvastatin is used to treat cholesterol, while etravirine is an HIV drug. While there may be some similarity in molecular mechanism that we are unaware of, it seems unlikely that these drugs are actually similar.

2) *Drug-drug interaction severity evaluation:* Subsequently, we evaluate the quality of our drug-drug interaction metrics. This is the component of our algorithm that has the potential to most severely impact the patient. It was of great importance that we avoid deadly interactions while not ruling out a drug combination based on minor side effects.

Our first, coarsest metric, was based off of the Decagon DDI database directly. We had assumed that the more interactions recorded between a pair of drugs, the more harmful the drug combination would be. However, we had noticed that acetaminophen and aspirin, which are actually sold together as one medication (together with caffeine) over the counter as Excedrin, had 337 different interactions, while atazanavir and simvastatin, which are contraindicated and one of our test pairs, had only 11 interactions in the database. This was an indicator that this first heuristic had limitations.

Our second heuristic was designed based on an observation from the data's original source in the TWOSIDES database, which includes statistical significance of the side effects. While acetaminophen and aspirin had no interaction with confidence over 4 (the confidence scores in the database range from 1 to 5 and are based on the logarithm of the p-score of the interaction), atazanavir and simvastatin have two interactions with confidence of 5. Thus, we decided to count only the interactions with confidence over 5.

Due to the fact that most pairs lack any in-

formation in TWOSIDES or Decagon, we limit a quantitative evaluation of the interaction metric to the pairs with data which were found in our search for alternatives to epinephrine and metoprolol, which was one of the problematic pairs we used as a test. For each of these candidate pairs with data, we compare the score that the two heuristics assigned them as well as the status of the pair on RxList, a website that records drug interactions.

TABLE III
SAMPLE DRUG PAIR INTERACTION SCORES

Drug 1	Drug 2	Score 1	Score 2	Status
epinephrine	metoprolol	94	8	Significant
epinephrine	ciprofloxacin	98	28	None
metoprolol	ciprofloxacin	249	0	Minor
betaxolol	metoprolol	58	0	Serious
betaxolol	ciprofloxacin	77	0	Significant

As a qualitative assessment of the first two pairs, the first pair, which we were trying to avoid in the first place, is associated with kidney failure with highest confidence. The second pair, however, is also associated with terrible effects such as heart attack with highest confidence.

The problem is that both heuristics not only somewhat disagree with each other, but that when they do seem to agree, as with betaxolol and metoprolol receiving a relatively low score, RxList actually reported the severity of interactions as the opposite- highly problematic.

On the bright side, we did manage to write software that calculated and used these metrics for recommendation of less problematic drug pairs, and if given the right network data, the software would give correct results. The problem is that there does not seem to be a good way to measure severity of drug interactions.

While our heuristics are certainly not a replacement for the advice of licensed medical professionals, we hoped to obtain a metric that was as valid as possible for use in our algorithm, without having to ask medical professionals to spend hours annotating drug-drug interactions. We had assumed that TWOSIDES and its downstream incarnation in Decagon carried the information that we needed, but due to the amount of information they have gathered, it became difficult to identify the truly problematic drug combinations amidst the noise.

We also faced a problem that was a bit of the reverse- many drug combinations that we wanted to find information about as part of the candidate-ranking stage of our algorithm simply did not have data listed in the database. Without this information, we could hazard a guess as to whether they would interact or not, and these pairs simply had to be thrown out. For example, while we came up with 22 distinct candidate alternative pairs for dopamine and phenylephrine, none of the pairs had any information in Decagon or TWOSIDES.

Another issue is that both databases contain information on deadly effects as well as less severe side effects, or phenomena which are not side effects at all. Some examples of these are ‘road traffic accident’, ‘herpes simplex’, and ‘dandruff’. While we wanted to narrow the document down to the side effects we deem most significant, given that there were 1301 side effects in TWOSIDES, we deemed it unfeasible for us to properly weight all of them, and we defer this work to a team of medical professionals in the future.

In the future, we need a gold standard with actual pharmacists and side effects weighted by not just significance but how severe the actual side effect is. This would be crucial not just for future computational work but for the well-being of all patients. When tested, pharmacists only correctly identified 66% of drug-drug interactions[11], and even current computerized systems only scored 250 out of a possible 400 points for accuracy in the study of Kheshti et al. (2016)[9]. Clearly there is a need for better data.

V. CONCLUSIONS AND FUTURE WORK

Drug-drug interactions (DDIs) can cause numerous adverse side effects for patients, and cost billions to the healthcare system each year. However, they are a seemingly necessary evil for patients that require a cocktail of drugs for multiple afflictions or a particularly difficult disease. Our goal was to provide a tool that can automatically alleviate the burden of DDIs caused by a pair of drugs.

Our system accepts a pair of drugs and recommends a new drug pair that is similar in function, yet reduces the number and severity of polypharmacy side effects, for a set of roughly 1774 drugs whose targets are known (though some of these don’t have data in TWOSIDES).

Our proposed method combines prior work done by Zitnik et al. (2018) on predicting polypharmacy side effects, and Park et al. (2015) on predicting drug similarity using a protein-protein interaction network. We first processed the Decagon model built by Zitnik et al. (2018) to distinguish between side effects by their severity. The Decagon model considers the 964 most commonly occurring types of polypharmacy side effects, but we wanted to focus on the most severe side effects, and therefore created a heuristic scoring system to assign weights to the different side effects based on confidence. This allowed us to determine if the side effects between two drugs could be categorized as severe or not. In the case of a severe interaction, we use a random walk algorithm to find similar drugs that can be substituted to minimize DDI severity.

While we successfully built a software framework which allowed us to fully implement our vision for an alternative recommender to problematic pairs of drugs (with the exception of name-to-CID and CID-to-DrugID lookup for drugs, which would be simple to achieve through a dictionary), our final product was hampered by data quality, in particular as it related to our metrics. It turned out that our similarity metric didn't always accurately predict the most similar cousins to a drug, and we had trouble with reliably capturing drug-drug interaction severity, even given a large dataset which Zitnik et al. (2018) used successfully for side effect prediction.

In the future, we hope to create a better interaction severity metric, which could offer a better sense of the combined toll potential side effects would have on a patient. We imagine this could use unsupervised learning methods to assign weights to the various side effects using existing medical data. We would also like to incorporate the severity and likelihood of each individual side effect in this model, and weight each side effect's score accordingly. We also hope to validate our drug recommendation system with physicians and pharmacists to determine which aspects need to be further refined.

To turn an eye even farther towards the future, we also believe our system, with different data sources, could be used in veterinary medicine. This would be very challenging due to the lack of

data on protein-protein interactions in animals, but could produce many new insights because drug-drug interactions in animals are not well studied.

VI. ACKNOWLEDGEMENTS

We would like to thank Jure Leskovec and the entire CS224W teaching staff for their guidance throughout this project, and to the BioSNAP group for providing access to their datasets.

Please note that we are happy with an equal grade split, and therefore did not outline individual contributions.

VII. CODE LOCATION

Our repository may be found at <https://bitbucket.org/drugdrug224w/drugdrug/src> and a backup is at <https://github.com/alexjwade/drugdrugrecommender>.

REFERENCES

- [1] E. F.R. and G. A.J., "Drug-related morbidity and mortality: updating the cost-of-illness model," *J. Am. Pharm. Assoc.*, vol. 41, no. 192199, 2001.
- [2] J. Z. Rijkom, E. Uijtendaal, M. T. Berg, W. V. Solinge, and A. Egberts, "Frequency and nature of drug-drug interactions in a dutch university hospital," *British Journal of Clinical Pharmacology*, vol. 70, no. 4, pp. 618–618, 2010.
- [3] B. Chen, Y. Ding, and D. J. Wild, "Assessing drug target association using semantic linked data," *PLoS Computational Biology*, vol. 8, 2012.
- [4] K. Park, D. Kim, S. Ha, and D. Lee, "Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks," *PLoS One*, vol. 10, no. 10, 2015.
- [5] J. Huang, C. Niu, C. D. Green, L. Yang, H. Mei, and J. J. Han, "Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network," *PLoS Computational Biology*, vol. 9, no. 3, 2013.
- [6] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, and X. Li, "Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data," *BMC Bioinformatics*, vol. 18, no. 1, 2017.
- [7] A. M. . L. J. Zitnik, M., "Modeling polypharmacy side effects with graph convolutional networks," *Bioinformatics*, 2018.
- [8] e. a. Tatonetti, Nicholas P., "Data-driven prediction of drug effects and interactions," *Science Translational Medicine*, Mar 2012.
- [9] R. Kheshti, M. Aalipour, and S. Namazi, "A comparison of five common drug-drug interaction software programs regarding accuracy and comprehensiveness," *Journal of research in pharmacy practice*, vol. 5, no. 4, pp. 257–263, 2016.
- [10] A. S. Brown and C. J. Patel, "Meshdd: Literature-based drug-drug similarity for drug repositioning," *Journal of the American Medical Informatics Association*, 2016.
- [11] M. . Science Translational Medicine, American Association for the Advancement of Science, "Pharmacist recognition of potential drug interactions," *Am J Health Syst Pharm*, vol. 56, no. 15, 1999.