Identifying candidate drugs for repositioning by graph based modeling techniques based on drug side-effects

Ken McGarry, Nicole Slater and Angela Amanning
Department of Pharmacy, Health and wellbeing, Faculty of Applied Sciences, University of Sunderland, UK
Email:ken.mcgarry@sunderland.ac.uk

***Abstract*—Drug development is a lengthy and highly costly endeavor, often with limited success and high risk. The objective of drug repositioning is to apply existing drugs to different diseases or medical conditions than the original target, and thus alleviate to a certain extent the time and cost expended. The area of drug repositioning is a suitable application area for computational intelligence because numerous online databases containing technical information on drug targets, protein interactions, side-effects and biological knowledge are freely available. Thus *in-silico* analysis can be used as a useful first stage to screen potential candidate drugs for possible redeployment. This paper takes the position that drugs with side-effects are potential candidates for use elsewhere, it is a case of identifying potential diseases that may benefit from this re-deployment. The system uses graph based computational techniques to analyze drugs with known side-effects and compares the proteins involved in these side-effects with proteins known to be identified with other diseases. Our intention is to find potential candidates for treating Alzheimer's disease.**

***Keywords*—complex networks, side-effects, hubs, nearness, betweeness.**

I. Introduction

Drug repurposing involves using existing pharmaceutical products on novel targets, the advantages are many; off-the-shelf drugs have undergone extensive testing and their toxicological properties are well known, therefore the costs are greatly reduced and also time to product delivery [24]. Thus it is more economical to repurpose an existing drug than develop one from scratch [12].

Difficulties in drug development arise because diseases are often complex with multi-factorial components such as interactions between genes, proteins and the environment [5]. Furthermore, drugs that are highly selective in terms of their targets are very rare. The usual situation is that a drug will target the proteins involved in the defective biological process but also interact to a lessor or greater extent with off-target proteins[36], [6]. Hence, depending on the severity of the side-effects the therapy may be stopped and alternative drugs will have to be found [35], [7].

Interestingly, there are many examples where unanticipated side-effects have proven to be beneficial to patients suffering from unrelated problems to the original purpose of the drug thus allowing the drugs to be re-deployed [37]. The most well cited example is the drug developed by Pfizer (sildenafil) which was intended to treat angina by relaxing the coronary arteries and therefore allow greater blood flow when it was discovered to have an interesting side-effect on the male participants [1]. Later marketed as Viagra, the drug now has annual sales of $1.6 Billion.

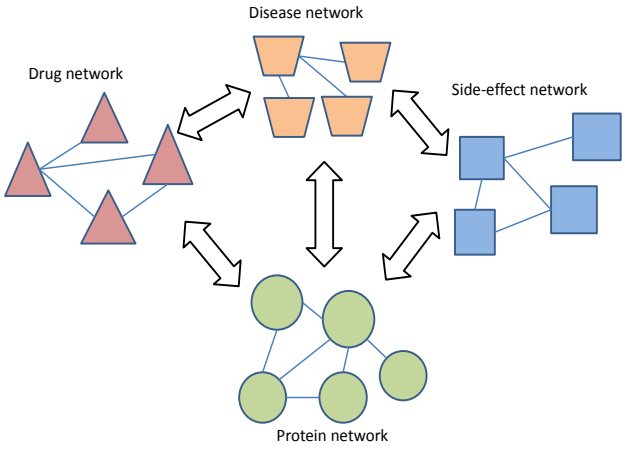


Fig. 1. The basic idea behind drug repositioning, it is the interplay between networks of interaction between drugs, proteins and side-effects, each having an effect on the other.

Alzheimers disease (AD) is a progressive neurodegenerative disease in which both genetic and non-genetic factors can contribute and is likely to affect 65 million people by 2030 [2]. There are several proteins that can affect the onset of the disease. Amyloid-β Precursor Protein (APP) is the major protein as it forms aggregates which will form the basis of the plaques, it is one of the leading causes of dementia and currently has no cure. Alzheimers is also a complex disease due to having many genes, proteins and many interactions between the proteins [32], [30]. The Amyloid-β hypothesis has driven research in this area for 20 years due to the understanding that the Amyloid-β plaques are the main cause for synaptic and neurodegeneration [34]. There are several competing arguments against the Amyloid-β plaques being the main cause, however most research provides evidence to supporting this hypothesis and now the majority scientists agree [29].

The key objective then is how to identify drug repurposing candidates, the approach taken in this paper is to view the

978-1-4799-5538-1/14/$31.00

problem as one of identifying and assessing side-effects in known drugs [27], [21].

A deeper understanding of the causes of disease is necessary, in particular knowledge of the genetic differences between individuals will eventually lead to improved treatments [31], [19]. This has only recently been made possible by the development of advanced genomic and proteomic techniques which are able to provide detailed and accurate data on individual cellular processes [16], [9]. We are now able to determine which genes (and proteins) interact together and form related functional groups that modify the behavior and ultimately the health of the cell [25].

Recent studies based on modeling cooperating modular groups of genes have suggested that diseases themselves are in fact network like [15]. The concept of the human disease network or *diseasome* is relatively new and is now starting to be explored as means of developing new drug products to tackle and combat diseases [11], [18], [26].

A. Related work

The PREDICT system developed by Gottleib et al tackles the issue from the viewpoint of large scale drug indications for personalised medicine [17]. The system analyses both approved and novel drugs, basing its recommendations on the observation that similar drugs affect similar diseases and they devise appropriate distance measures to implement this. PREDICT is able to build a model of disease specific signatures that can be used assist the drug targeting process.

The connectivity map (CMAP) of Cheng et al was developed by compiling the data available from Affymetrix genechips taken from 13,000 human samples (9,000 diseased and 3,400 healthy individuals) [9]. These were mapped against 152 drug profiles and 145 disease gene signatures, this was achieved by using the *eXtreme Sum* (XSum) scoring algorithm. The authors concluded that although they could identify drug-interaction pairs they admitted the CMAP method required better validation data as some diseases had worse than random performance.

The system developed by Chiang and Butte uses a guilt by association (GBA) method which involves mapping the diseases to one logical set and mapping the drugs to another set, the degree of overlap between the two sets was determined systematically [10]. The authors claim to have produced 57,000 novel drug uses of which only a small number could be verified and the false positive rate appears rather high.

The remainder of this paper is structured as follows; section two describes the methods, consisting of data used and the processes along with assumptions that have been made, it also provides a brief discussion of system architecture in terms of flow of information and the overall structures modeled; further information including the statistical measures used to calculate the protein networks; section three outlines the results of modeling a known group of gene modules for the Alzheimer disease state and the protein targets that are shared by the candidate drugs; finally section five presents the conclusions and future work.

II. Methods

The most appropriate databases containing the necessary side-effects information, protein connectivity patterns were idenified and are described in the following sections. The R language was used along and the RStudio programming environment on an Intel Xenon CPU, 64-bit with dual processors (3.2GHz) and 128 GB of RAM. The R code was not compiled or optimized.

A. Data sources

Figure 2 gives a basic idea of the flow of information and data sources. A number of publicly available databases are used in this work which are listed in table I, the key resources are briefly described.

TABLE I. Ontologies and databases utilized for mapping, conversion and integration

Data source	Type	Function
STITCH V9	database	chemical-to-protein interaction
STRING	database	protein-to-protein interactions
GO.db	ontology	hierarchical structure of gene products
org.Hs.eg.db	database	mapping of gene symbol-to-gene ID's
KEGG	ontology	pathways of interactions
DrugBank	database	drug types and their targets, chemical information
SIDER2	database	drug side-effects
DO	ontology	hierarchy disease relationships

The raw data of known protein-to-protein interactions is held in the STITCH and STRING databases [28]. The STRING database also makes use of nearly 6 million known associations made by text mining and annotation by experts, through co-occurrence and and natural language processing but with confidence levels. Furthermore, about 30,000 associations are predicted based on similar protein characteristics. The STITCH database contains protein to chemical mappings, again with confidence factors. Annotations by human curators have higher values than through text mining.

The DrugBank database was utilized because it lists the majority of drugs that are prescribed or have reached the clinical trial stage and is widely used by those developing drugs, chemists, pharmacologists and those involved in pharmaceutics research [22]. Each drug is listed with its main targets, off known off-targets and the chemical structure and other properties.

The SIDER2 database is an important repository of hundreds of drugs and thousands of their known side-effects on humans [21], [20]. The latest release contains 4,192 side-effects and 996 drugs, this information is extracted from public documents and package inserts. The database is structured into drug classes and types of side-effect and has been very useful in drug-target prediction [8]. Further information is collected from clinical trials where they report side-effects from the placebo group, technically this group should not be reporting side-effects. However, note is taken of the frequency of reported side-effects and this is related to the population in general versus those on the actual drugs.

The Gene Ontology (GO) and Disease Ontology (DO) are used to annotate the proteins and drug targets with additional information useful for a deeper interpretation of the biological processes and structures [23], [33]. The DO database contains knowledge on 8,043 inherited, developmental and acquired

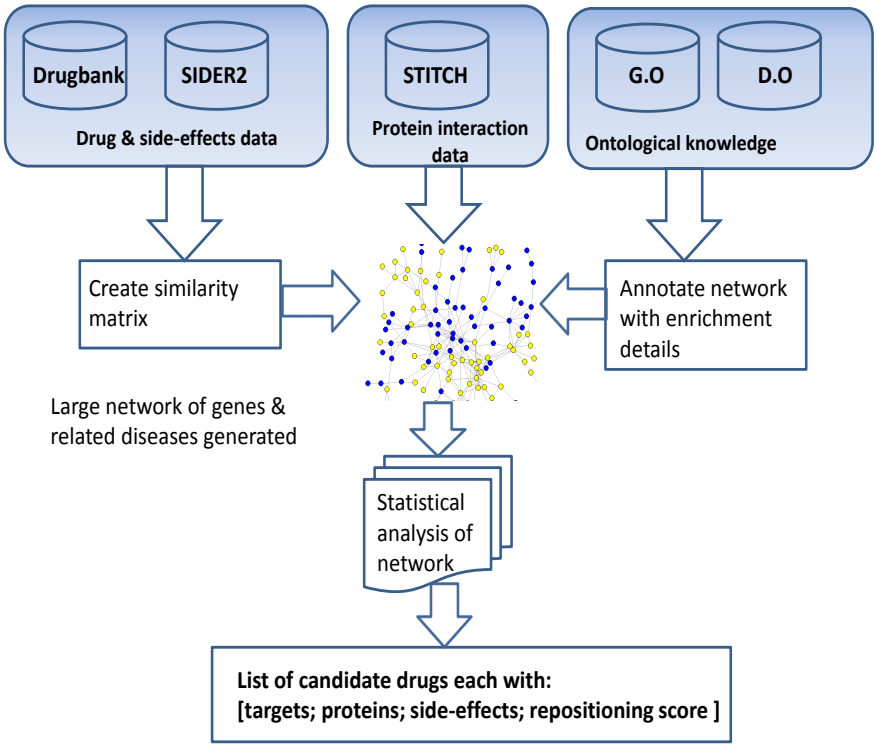


Fig. 2. Overview of system operation, showing database sources, data flow and statistical analysis

human diseases. Through enrichment analysis, the R package DOSim is able to explore the biological meaning of related genes in terms of structure, function and hierarchy. The concepts in DOSim are organized into a directed acyclic graph (DAG) similar to a tree structure, the concepts are linked by' is-a' relationships. The lower the term or concept is positioned in the hierarchy then the more specific the term is, higher-up terms describe higher level or more abstract concepts.

However, a number of challenges are presented by the nature of the data used in the construction of networks, most notably it breaks a number of statistical assumptions, for example there is a high degree of correlation within the cell of the metabolites and various activity patterns. This invalidates the important statistical assumption of independent variables. Also, the most powerful statistical techniques are parametric but the majority of proteomic data are highly skewed. Considering these limitations our aim is to develop a computational model to identify, assess and rank candidate drugs based on targets/side-effect proteins in common with the candidate drug and the specific disease (Alzheimer's) we are interested in.

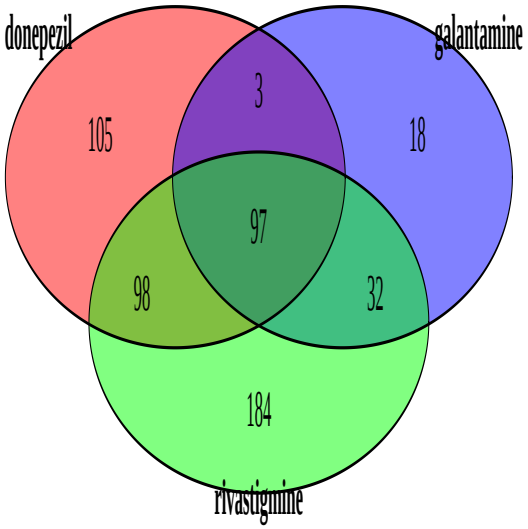


Fig. 3. Venn diagram showing common side-effects between the three main Alzheimers drugs

B. Process

The side-effects pertaining to the three drugs commonly used to treat Alzheimer's (Donepezil, Galantamine and Rivastigmine) were obtained from the SIDER2 database. Figure 3 shows a Venn diagram illustrating the common side-effects shared by the three drugs and those unique to each drug. The decision was made to tackle the problem by using only the side-effects common to all three drugs which meant that 97 side-effects were available for study.

The SIDER2 database was searched for drugs that produced the same side-effects, those drugs that had a similarity score of 10% and above we retained, thus 42 drugs were available for scrutiny. As an example in table II the side-effects for one drug (levobunolol) are shown.

A further analysis of the SIDER2 database revealed the numbers of side-effects per drug and drugs per side-effect. Figure 4, this analysis is important and gives an indication

TABLE II. EXAMPLE OF SIDE-EFFECTS FOR ONE DRUG

levobunolol	ptosis
levobunolol	rash
levobunolol	cerebrovascular accident
levobunolol	respiratory failure
levobunolol	bronchospasm

of what to expect in terms of frequency of side-effects for a typical drug. Some drugs may have several hundred side-effects, not all are life threating and not all may be directly related to the drug. The drug companies must report each patents/reported side-effect on the box insert, it is plausible to expect in many cases that some side-effects may have their causes elsewhere and thus a source of bias.

III. GRAPH THEORY AND NETWORKS OF INTERACTION

Graph theoretic methods can be applied to any discipline where the entities of interest are linked together through various associations or relationships. Quite diverse application areas such as social network analysis and biological networks are particularly suited to the mathematics of graph construction, traversal and inferencing. A graph G = (V, E) consists of a set of nodes often called vertices V and a set of links called edges E. The links in this case are undirected, that is to say there is no implied direction to the relationship in the sense that A causes B. The criteria we use to determine the relevance of disease connectivity is based upon well known measures from the graph theoretic literature [13], [3], [4].

- CLOSENESS. The simplest of all measures is degree centrality (DC). DC(i) is the number of edges present upon node i, i.e. the number of other proteins that protein interacts with. Closeness centrality: This measure is the closeness centrality (CC). The closeness centrality of protein i is the sum of graph-theoretic distances from all other proteins in the PPI network, where the distance $d(v_i, v_j)$ from one protein i to another j is defined as the number of links in the shortest path from one to the other. The closeness centrality of protein i in a PPI network is given by the following expression:

$$CC(v_i) = \frac{N-1}{\sum_j d(v_i, v_j)} \tag{1}$$

- BETWEENNESS. Betweenness centrality: Is a measure of the degree of influence a protein has in facilitating communication between other protein pairs and is defined as the fraction of shortest paths going through a given node. If $p(v_i, v_j)$ is the number of shortest paths from protein i to protein j, and $p(v_i, v_k, v_j)$ is the number of these shortest paths that pass through protein k in the PPI network, then the BC of node k is given by:

$$BC(v_k) = \sum_i \sum_j \frac{p(v_i, v_j, v_k)}{p(v_i, v_j)}, i \neq j \neq k \tag{2}$$

The z-score is defined by the following equation where k_i defines the number of links of node *i* to the nodes in nearby modules si, and ks_i is the average of k over all nodes in si and $\sigma_{k_{s_i}}$ is the standard deviation of k in s_i :

$$Z_i = \frac{k_i - k_{s_i}}{\sigma_{k_{s_i}}} \tag{3}$$

The z-score allows us to measure how well-connected protein i is to the other proteins within the module.

Where k_{s_i} defines the number of links from protein i to proteins in modules s, k_i is the degree of protein i. We can interpret the participation coefficient for each protein by noting that the value is constrained between 1.0 and 0, that is to say 1.0 implies that all of protein i links are evenly distributed between all of the modules and 0 if its links are only within one module.

IV. RESULTS

The STITCH database was manually searched with keywords and entered each drug (Donepezil, Galantamine and Rivastigmine). The screenshots are from the STITCH interface, drugs and chemicals are lozenge shaped while proteins are circles, the three networks were saved as CSV files and loaded into R from which the statistics were calculated. Three protein networks were created for each of the three main drugs and the statistics were calculated for each one and shown in table III. The protein networks shown in figure 5(a) to 5(c) and are the drug-to-protein interactions.

In table III we have the statistical breakdown of the network connectivity patterns. The networks in this case are not complex with few nodes and smaller numbers of connections. It appears that these drugs significantly target proteins with low degrees, based on the number of nodes and connections.

TABLE III. NETWORK STATISTICS FOR THE ALZHEIMER'S DRUGS: DATA FROM STITCH DATABASE

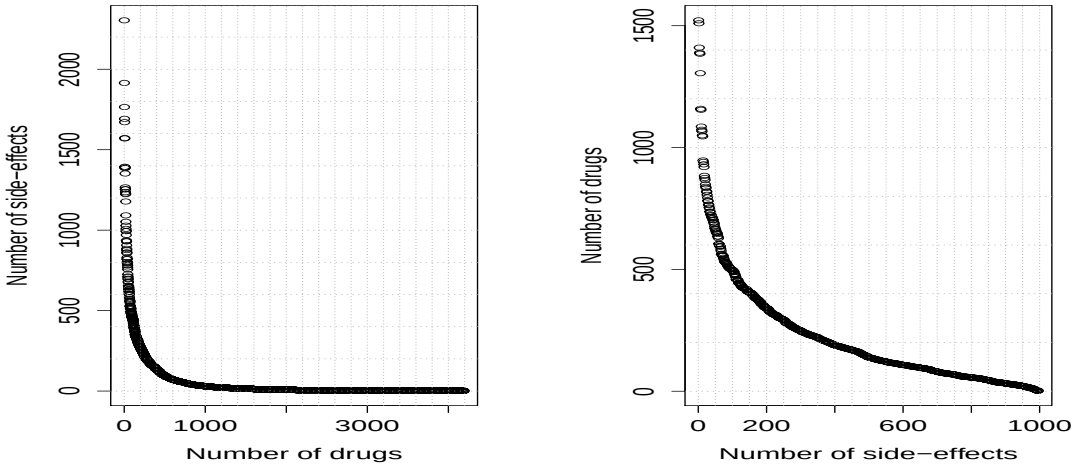
Drug	nodes	connect	diameter	transitivity	reciprocity	path
Donepezil	17	34	3	0.35	0.0	1.25
Galantamine	3	6	4	0.0	0.0	1.0
Rivastigmine	4	8	1	0.0	0.0	1.0

Analysis of the key proteins and their PPI network connectivity statistics was performed, this data is presented in table IV. The APOE and APP proteins have the greatest number of interacting proteins with considerable connectivity between them. BACE1 appears to have the least number of interacting partners with few connections between them. It is therefore likely that APOE and APP will have a better chance of interacting with potential drugs, and of course causing more side-effects.

TABLE IV. NETWORK STATISTICS OF KEY PROTEINS LINKED TO ALZHEIMER'S: BASED ON STITCH DATABASE

Protein	nodes	connect	diameter	transitivity	reciprocity	path
PSEN1	176	1261	6	0.25	0.0	1.9
PSEN2	89	377	4	0.25	0.0	1.87
APOE	301	7990	6	0.46	0.0	1.8
APP	300	13000	6	0.76	0.0	1.75
BACE1	51	87	4	0.11	0.0	1.78

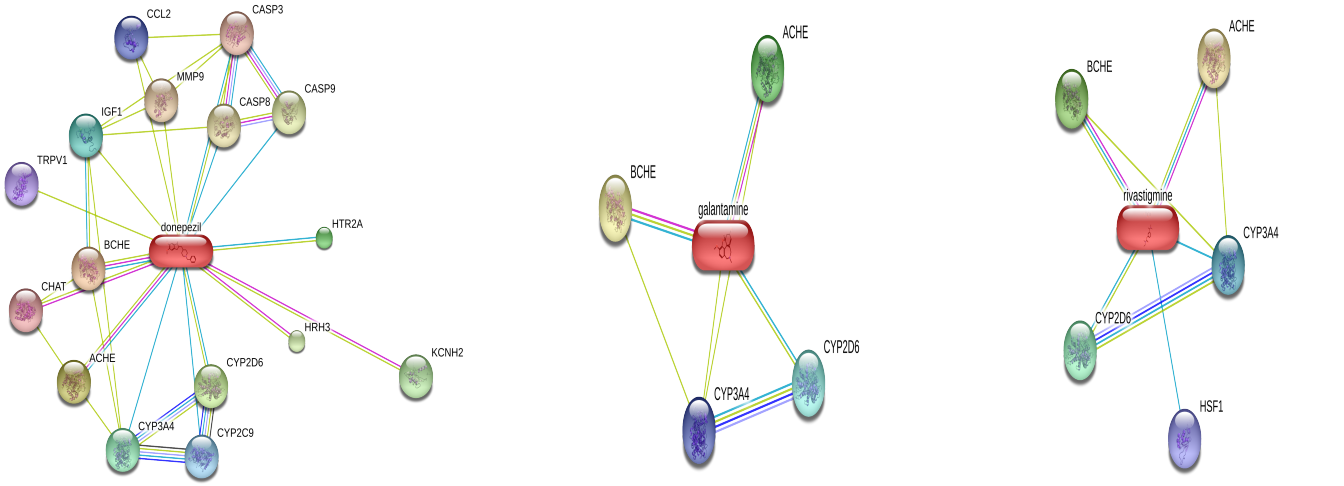
Referring to table V, these are the top 10 candidate drugs identified by their similarity to the current the drugs (Galantiamine, Donepezil and Rivastigmine) in terms of similar side-effects experienced. The most similar drug is Ropinirole with



(a) Number of side-effects per drug

(b) Number of drugs per side-effect

Fig. 4. The relationship between the number of drugs in the SIDER2 database and their corresponding side-effects



(a) Proteins associated with Donepezil

(b) Proteins associated with Galantamine

(c) Proteins associated with Rivastigmine

Fig. 5. Interacting proteins affiliated with the three conventional Alzheimer's drugs

a score of 17.5% similarity, the next is Tramadol with a score of 16.4%, the number of side-effects experienced by patients on these drugs is 659 and 506 respectively.

TABLE V. TOP TEN OF 42 DRUGS WITH SIMILAR SIDE-EFFECTS

id	drugnames	NoSideEffects	coverage (%)
740	ropinirole	659.00	17.53
467	tramadol	506.00	16.49
498	pregabalin	724.00	16.49
848	citalopram	474.00	16.49
299	naltrexone	396.00	15.46
330	paroxetine	501.00	15.46
234	ofloxacin	560.00	14.43
141	fluoxetine	516.00	13.40
154	aripiprazole	662.00	13.40
866	olanzapine	385.00	13.40

The top ten candidate drugs were originally targeted at central nervous system diseases; Ropinirole is targeted at Parkinson's disease; Tramadol is a analgesic but several of the pathways involved are common targets of Parkinson's and Alzheimer's; Pregabalin is an anti-convulsant; Citalopram and Paroxetine are antidepressants; Aripiprazole and Olanzapine are for schizophrenia and anti psychotic use. The drug to target map in figure 6 shows several of the shared pathways of interaction, the blue circles indicate the candidate drug and the sizes are related to the number of connections. The targets (proteins) are indicated by the pink circles.

A. *Enrichment from Disease Ontology*

The next part of the analysis was to augment the statistical information extracted from the networks by adding biological knowledge from two popular and comprehensive ontologies. This supporting knowledge enables a more realistic *in-silico* evaluation of the candidate drugs. Disease ontology

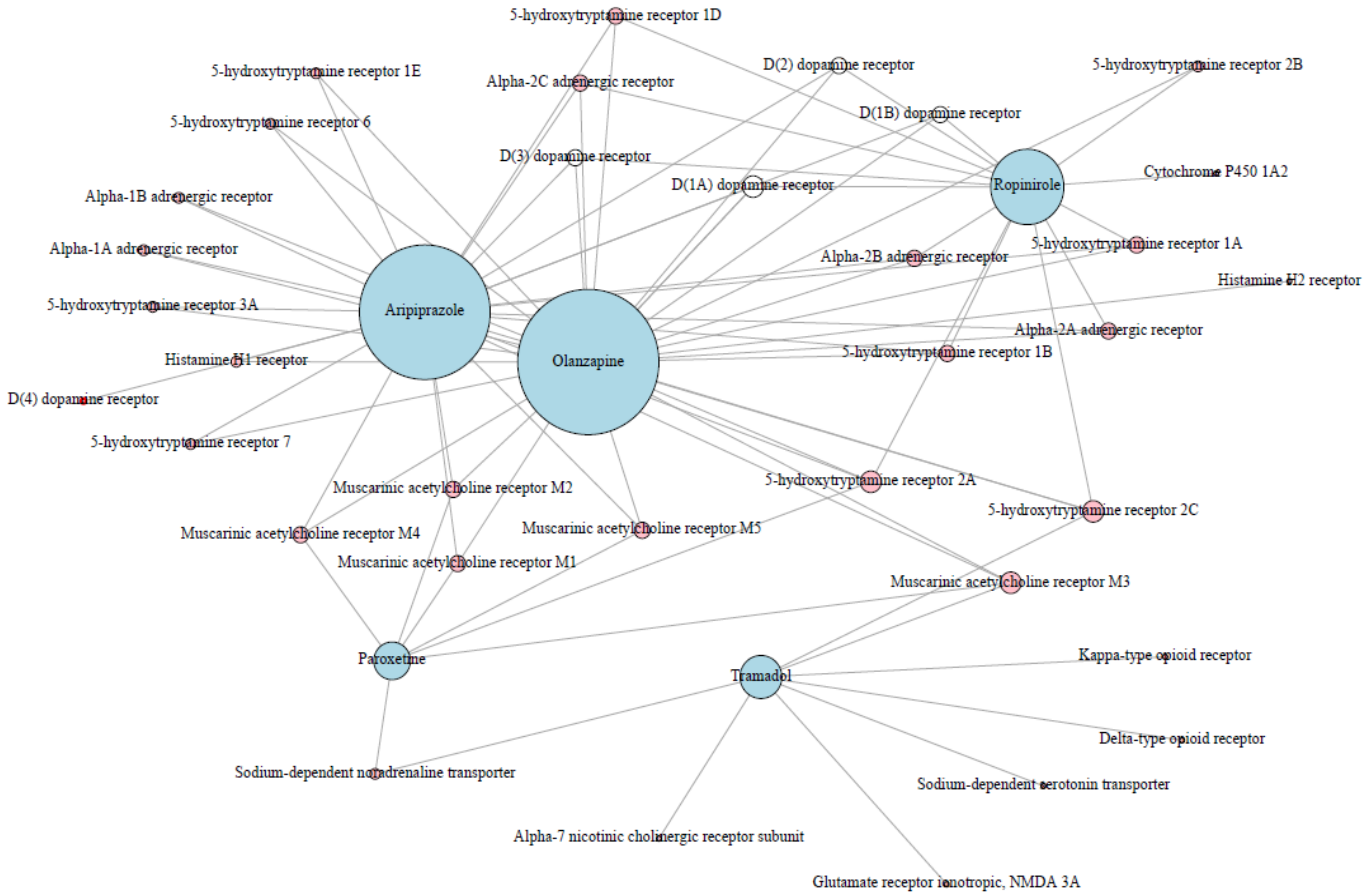


Fig. 6. Key candidate drugs and their target receptor connectivities (Aripiprazole, Olanzapine, Ropinirole, Tramadol and Paroxetine) and their common targets. The graph shows the interlinked relationship of drugs and shared targets

was used to compare the semantic similarity between the diseases in the list of drugs, the greater the degree of disease similarity implied a greater chance of drug target similarity and hence better chance of drug usefulness. The Wang measure was used from the DOSE R-package developed by Li [14]. When presented with the appropriate DOID terms (e.g. Alzheimers=DOID:10652) the DOSE functions will search the related diseases (stored as a directed acyclic graph or DAG) and return a list of genes associated with them. The DAG is represented by (A, T_A,E_A) where T_A is a set of terms for A and its ancestors in the graph and E_A is the set of edges connecting DO terms. The semantic value (S) for a given term (t) is calculated by:

$$S_A(t) = \begin{cases} S_A(A) = 1 & : \\ S_A(t) = max\{w_e \times S_A(\acute{t}|\acute{t}) \in \text{children of}(t)\} & : \\ \text{if } t \neq A & : \end{cases}$$

Where w_e links term t with the semantic value (S) for edge $e \in E_A$. Term A semantic value is calculated by:

$$SV(A) = \sum_{t \in T_A} S_A(t)$$

$$Sim_{Wang}(A,B) = \frac{\sum_{t \in T_A \cap T_B} S_A(t) + S_B(t)}{SV(A) + SV(B)} \quad (4)$$

Where $S_A(t)$ is the semantic value of DO term t which is related to term A and $S_B(t)$ is the semantic value of DO term t related to to term B. The following DOID's are the top level items for the main associated diseases, similarity between them is calculated from the shared sub-level items.

- alzheimers DOID:10652
- parkinsons DOID:14330
- schizophrenia DOID:5419
- psychosis DOID:8646
- depression DOID:1596
- pain disorder DOID:0060164

In table VI the main types of diseases treated by the 42 drugs are shown. The similarity measure is constrained to lie between 0 and 1, with 1 being the highest level of similarity. Ignoring the diagonal, the highest measure is between psychosis and schizophrenia (0.64) and between Parkinson's and Alzheimer's (0.44), depression and psychosis (0.39), followed by epilepsy and Alzheimer's (0.35).

TABLE VI. DISEASE ONTOLOGY SEMANTIC SIMILARITY SCORES BETWEEN DISEASE TERMS

	alzheimers	parkinsons	schizophrenia	psychosis	depression	pain_disorder	epilepsy	ulcer	candidiasis	amenorrhea
alzheimers	1.00	0.44	0.06	0.06	0.06	0.08	0.35	0.20	0.04	0.15
parkinsons	0.44	1.00	0.06	0.06	0.06	0.08	0.35	0.20	0.04	0.15
schizophrenia	0.06	0.06	1.00	0.64	0.39	0.27	0.07	0.11	0.06	0.09
psychosis	0.06	0.06	0.64	1.00	0.39	0.27	0.07	0.11	0.06	0.09
depression	0.06	0.06	0.39	0.39	1.00	0.27	0.07	0.11	0.06	0.09
pain_disorder	0.08	0.08	0.27	0.27	0.27	1.00	0.09	0.14	0.08	0.11
epilepsy	0.35	0.35	0.07	0.07	0.07	0.09	1.00	0.23	0.05	0.17
ulcer	0.20	0.20	0.11	0.11	0.11	0.14	0.23	1.00	0.08	0.27
candidiasis	0.04	0.04	0.06	0.06	0.06	0.08	0.05	0.08	1.00	0.06
amenorrhea	0.15	0.15	0.09	0.09	0.09	0.11	0.17	0.27	0.06	1.00

B. *Enrichment from Gene Ontology*

For the proteins associated with the three standard drugs enrichment was performed using gene ontology (GO), the enrichment is based on similarity measures using information content techniques. The content of which is calculated by taking the negative log probability of the terms t appearing in the database.

$$p(t) = \frac{nt}{N} \mid t' \in \{t, \text{children of } t\} \tag{5}$$

Where nt number of terms and N is total of all GO terms in the database.

The following results were obtained. The GO was searched to a depth of three levels, using the biological process (BP) section of GO database. As seen in table VII the results for the Donepezil group of 16 proteins are displayed. Only the first 10 are shown, in descending order of importance e.g. the first entry GO:0044763 which corresponds to *single-organism cellular process* all 16 proteins are known to cooperate in this activity, the last entry in the table refers to *response to chemical* with 13 of the 16 proteins participating in this function. In total there were 121 identified activities in BP for the Donepezil group.

TABLE VII. GENE ONTOLOGY ENRICHMENT FOR BIOLOGICAL PROCESS ASSOCIATED WITH DONEPEZIL AND IT'S INTERACTING PROTEINS

ID	Description	Count	Ratio
GO:0044763	single-organism cellular process	16	16/16
GO:0044237	cellular metabolic process	15	15/16
GO:0044238	primary metabolic process	14	14/16
GO:0071704	organic substance metabolic process	14	14/16
GO:0050794	regulation of cellular process	14	14/16
GO:0050789	regulation of biological process	14	14/16
GO:0051716	cellular response to stimulus	13	13/16
GO:0044700	single organism signaling	13	13/16
GO:0044707	single-multicellular organism process	13	13/16
GO:0042221	response to chemical	13	13/16

For the other drug to protein groups, Galantamine and Rivastigmine had 66 and 31 biological processes respectively associated with them. Figure 7 shows the histogram of protein matched up biological terms from GO, it is clear that the Donepezil group is quite active with many of its proteins engaged in numerous cellular activities. This is termed a *promiscuous* drug since it affects many proteins and is involved in may processes, as expected such drugs cause numerous side-effects. It should be noted that not all of the processes identified by the GO enrichment are of direct use in drug repositioning, since the majority of identified processes occur in most cells and are not specific to the drug targets. However, this process provides a level of biological understanding and explanation that can augment statistical methods used so far.

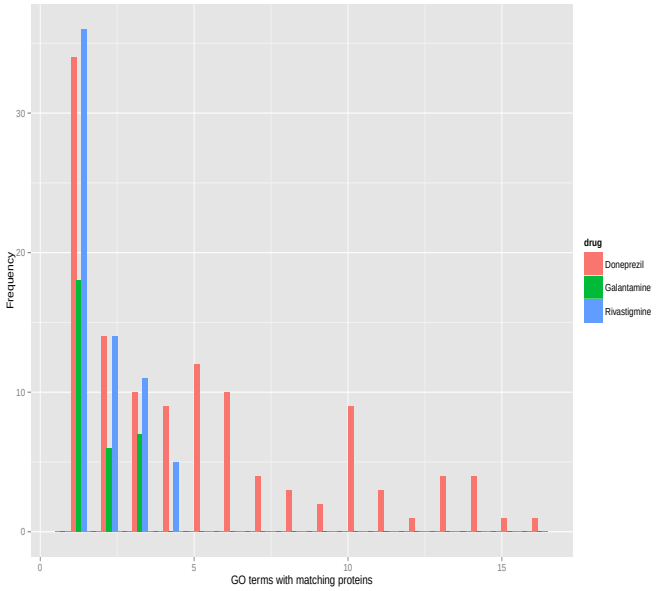


Fig. 7. Histogram plot of frequency of proteins matched with enrichment terms

V. CONCLUSION

In this paper we have investigated a systematic method of exploring multiple datasets dealing with drug side-effects; key proteins known to be responsible for diseases; to identify potential candidate drugs based on these shared side-effects and cooperating protein interactions. We have also shown that the similarity of adverse drug effects (side-effects) can be reasoned by the common protein-subnetworks that they interact with. We have demonstrated the usefulness of the side-effect approach to screen candidate drugs for repositioning and the use of ontologies such as DO and GO provide a level of biological explanation. The ultimate test of course is have these drugs tested, at least possibly *in-vitro*. Future work will further investigate the chemical properties of the candidate drugs and to develop a deeper theory of how the affected proteins cooperate within Alzheimer's disease with greater emphasis on use of the disease ontology and gene ontology for enrichment. We will furthermore differentiate between targets and off-targets, plus incorporate frequency counts of each of the side-effects caused by the drugs.

ACKNOWLEDGMENT

This work was supported by a grant from the University of Sunderlands Digital Innovation Research Beacon. We are also grateful to Daniel Himmelstein for his advice regarding SIDER2. We also wish to thank the anonymous reviewers for their comments for improving the quality of the paper. The R code and data files can be made available upon application to the corresponding author.

REFERENCES

[1] T. Ashburn and K. B. Thorl. Drug repositioning: identifying and developing new uses for existing drugs. *Nature Reviews Drug Discovery*, 3:673–683, 2004.

[2] S. Bandyopadhyay and J. Rogers. Alzheimer's disease therapeutics targeted to the control of amyloid precursor protein translation: Maintenance of brain iron homeostasis. *Biochemical Pharmacology*, 88(4):486–494, 2014.

[3] A. Barabasi and Z. Oltvai. Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5:101–113, 2004.

[4] AL Barabasi, N Gulbahce, and J Loscalzo. Network medicine: a network-based approach to human disease. *Nat Rev Genet*, 12:56–68, 2011.

[5] F. Barrenas, S. Chavali, P. Holme, R. Mobini, and M. Benson. Network properties of complex human disease genes identified through genome-wide association studies. *PLoS ONE*, 4(11):e8090, 11 2009.

[6] H. Bisgin, Z. Liu, K. Reagan, H. Fang, X. Xiu, and W. Tong. Investigating drug repositioning opportunities in fda drug labels through topic modeling. *BMC Bioinformatics*, 13(S-15):S6, 2012.

[7] L. Brouwers, M. Iskar, G. Zeller, V. vanNoort, , and P. Bork. Network neighbors of drug targets contribute to drug side-effect similarity. *PLoS ONE*, 6(7):e22187, 07 2011.

[8] M. Campillos, M. Khun, A. Gavin, L. Jensen, and P. Bork. Drug target identification using side-effect similarity. *Science*, 321:263–266, 2008.

[9] J. Cheng, L. Yang, V. Kumar, and P. Agarwal. Systematic evaluation of connectivity map for disease indications. *Genome Medicine*, 6(12), 2014.

[10] A. Chiang and A. Butte. Systematic evaluation of drug-disease relationships to identity leads for novel drug uses. *Clinical pharmacology and therapeutics*, 86(5):507–510, 2009.

[11] He D, Liu ZP, and Chen L. Identification of dysfunctional modules and disease genes in congenital heart disease by a network-based approach. *BMC Genomics.*, 12, 2011.

[12] J. Dudley, D. Tarangini, and A. Butte. Exploiting drug-disease relationships for computational drug repositioning. *Briefings in Bioinformatics*, 12(4):303–311, 2011.

[13] L. Freeman. Centrality in social networks I: Conceptual clarification. *Social Networks*, 1:215–239, 1979.

[14] G. Yan G. Yu and Q. He. DOSE: an R/Bioconductor package for disease ontology semantic and enrichement analysis. *Bioinformatics*, 31(4):608–609, 2015.

[15] Kwang-Il Goh, Michael E. Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-Lszl Barabsi. The human disease network. *Proceedings of the National Academy of Sciences*, 104(21):8685–8690, 2007.

[16] J. Gonalves, A. Francisco, Y. Moreau, and S. Madeira. Interactogeneous: Disease gene prioritization using heterogeneous networks and full topology scores. *PLoS ONE*, 7(11), 2012.

[17] A. Gottlieb, G. Stein, E. Ruppin, and R. Sharan. Predict: a method for inferring novel drug indications with application to personalized medicine. *Molecular Systems Biology.*, 7(496), 2011.

[18] G. Hu and P. Agarwal. Human disease-drug network based on genomic expression profiles. *PLoS ONE*, 4(8):e163, 2009.

[19] P. Imming, C. Sinning, and A. Meyer. Drugs, their targets and the nature and number of drug targets. *Nature Reviews Drug Discovery*, 5:821–834, 2006.

[20] M. Kuhn, M. Al Banchaabouchi, M. Campillos, L. Jensen, C. Gross, A. Gavin, and P. Bork. Systematic identification of proteins that elicit drug side effects. *Molecular Systems Biology*, 9(1), 2013.

[21] M. Kuhn, M. Campillos, I. Letunic, LJ.Jensen, and P. Bork. A side effect resource to capture phenotypic effects of drugs. *Molecular Systems Biology*, 6(343), 2010.

[22] V. Law, C. Knox, and Y. Djoumbou et al. Drugbank 4.0: shedding new light on drug metabolism. *Nucleic Acids Research*, 42:D1091–D1097, 2014.

[23] Jiang Li, Binsheng Gong, Xi Chen, Tao Liu, Chao Wu, Fan Zhang, Chunquan Li, Xiang Li, Shaoqi Rao, and Xia Li. Dosim: An r package for similarity between diseases based on disease ontology. *BMC Bioinformatics*, 12(1):266, 2011.

[24] Ruifeng Liu, Narender Singh, Gregory Tawa, Anders Wallqvist, and Jaques Reifman. Exploiting large-scale drug-protein interaction information for computational drug repurposing. *BMC Bioinformatics*, 15(1):210, 2014.

[25] K. McGarry, J. Chambers, and G. Oatley. Graph based analysis of protein interaction for diabetes research. *Artificial Intelligence in Medicine*, 41(2):129–144, 2007.

[26] K. McGarry and U. Daniel. Computational techniques for identifying networks of interrelated diseases. *The 14th UK Workshop on Computational Intelligence, UKCI-2014*, 2014.

[27] K. McGarry and U. Daniel. Data mining open source databases for drug repositioning using graph based techniques. *Drug Discovery World*, 16(1):64–71, 2015.

[28] K. Michael, D. Szklarczyk, A. Franceschini, C. von Mering, L. Jensen, Lars Juhl, and P. Bork. Stitch 3: zooming in on proteinchemical interactions. *Nucleic Acids Research*, 40(D1):D876–D880, 2012.

[29] G. Morris, I. Clark, and B. Vissel. Inconsistencies and controversies surrounding the amyloid hypothesis of alzheimer's disease. *Acta Neuropathol Commun*, 2(1), 2014.

[30] M. Oh, J. Ahn, and Y. Yoon. A network-based classification model for deriving novel drug-disease associations and assessing their molecular actions. *PLOSOne*, 9(10), 2014.

[31] J. Overington, B. AlLazikani, and A. Hopkins. How many drug targets are there? *Nature Reviews Drug Discovery*, 5:993–996, 2006.

[32] M. Re and G. Valentini. Network-based drug ranking and repositioning with respect to drugbank therapeutic categories. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 10(6):1359–1371, 2013.

[33] Lynn Schriml, Cesare Arze, Suvarna Nadendla, Yu-Wei Chang, Mark Mazaitis, Victor Felix, Gang Feng, and Warren Kibbe. Disease ontology: a backbone for disease semantic integration. *Nucleic Acids Research*, 40:D940 – D946, 2012.

[34] K. Shameer, B. Readhead, and J. Dudley. Computational and experimental advances in drug repositioning for accelerated therapeutic stratification. *Current Topics in Medicinal Chemistry*, 15(1):5–20, 2015.

[35] Y. Wang, S. Chen, N. Deng, and Y. Wang. Network predicting drug's anatomical therapeutic chemical code. *Bioinformatics*, 29(10):1317–1324, 2013.

[36] Jihong Yang, Zheng Li, Xiaohui Fan, and Yiyu Cheng. Drug disease association and drug-repositioning predictions in complex diseases using causal inference probabilistic matrix factorization. *Journal of Chemical Information and Modeling*, 54(9):2562–2569, 2014.

[37] H. Ye, Q. Liu, and J. Wei. Construction of drug network based on side effects and its application for drug repositioning. *PLoS ONE*, 9(2), 2014.