DRUG REPOSITIONING BASED ON CENTRALITY MEASURES OF DRUG SIDE EFFECTS NETWORK

KENG YING YING

FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2021

DRUG REPOSITIONING BASED ON CENTRALITY MEASURES OF DRUG SIDE EFFECTS NETWORK

KENG YING YING

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

INSTITUTE OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2021

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **KENG YING YING**

Registration/Matric No.: **17035915**

Name of Degree: **MASTER OF SCIENCE**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

DRUG REPOSITIONING BASED ON CENTRALITY MEASURES OF DRUG SIDE EFFECTS NETWORK

Field of Study: **APPLIED MATHEMATICS (MATHEMATICS)**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the Universiti Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:
Designation:

DRUG REPOSITIONING BASED ON CENTRALITY MEASURES OF DRUG SIDE EFFECTS NETWORK

ABSTRACT

Drug repositioning, the process of discovering new therapeutic uses of existing drugs, is an alternative to the risky, costly, and time-consuming traditional drug discovery process. Based on the hypothesis that drugs with similar side effects may share similar therapeutic indications, a network of drugs is constructed based on their side-effect similarities. Then, for a target disease of concern, the potential drug candidates that may be repurposed for treating it can be identified from among the network neighbors of the existing drugs approved for the disease. By applying three classic centrality measures – degree, closeness, and betweenness – to rank the drugs in the network, we observe that the drug candidates identified from among the neighbors of the top central drugs approved for our target diseases are more consistent with clinical interests as indicated by the record of clinical trials related to the diseases, when compared to the candidates identified from the neighbors of their random and peripheral counterparts. The present work indicates that network positions of drugs have a role in repurposing their neighbors and hence should be taken into account in finding new uses of drugs.

Keywords: Network analysis, centrality, drug repositioning.

REPOSISI UBAT BERDASARKAN UKURAN SENTRALITI UNTUK RANGKAIAN UBAT

ABSTRAK

Reposisi ubat, proses mencari kegunaan baru bagi ubat-ubat yang sedia ada, merupakan alternatif bagi proses pembuatan ubat baru yang berisiko, mahal dan memakan masa. Berdasarkan hipotesis bahawa ubat dengan kesan sampingan yang serupa dapat menunjukkan kesan terapeutik yang serupa, rangkaian ubat dibina berdasarkan kesamaan kesan sampingan. Kemudian, calon pengganti ubat yang berpotensi untuk merawat penyakit yang tertentu boleh dikenal pasti daripada jiran ubat yang telah disahkan untuk penyakit tersebut. Dengan menggunakan tiga ukuran sentrali yang klasik – *degree*, *closeness*, dan *betweenness* – untuk menilai ubat-ubat di rangkaian tersebut, kami memerhatikan bahawa calon ubat yang dikenal pasti daripada jiran ubat yang sentral yang telah disahkan untuk penyakit sasaran kami adalah lebih konsisten dengan minat penyelidikan klinikal yang ditunjukkan melalui catatan ujian klinikal yang berkaitan dengan penyakit tersebut, apabila dibandingkan dengan calon yang dikenal pasti daripada jiran ubat yang periferal dan rawak. Kesimpulannya, kami menunjukkan bahawa kedudukan dalam rangkaian ubat mempunyai peranan dalam mengenali calon reposisi ubat dan justeru faktor ini harus dipertimbangkan ketika mencari penggunaan baru untuk ubat.

Kata kunci: Analisis rangkaian, sentraliti, reposisi ubat.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my highest gratitude to my supervisors, Dr. Kwa Kiam Heong and Prof. Kurunathan Ratnavelu, for their valuable guidance, positive encouragement and advice during the course of this research. They have encouraged me to develop my academic independence, which makes it a rewarding learning experience for me.

Besides, I would like to thank the panels who offered constructive feedback on this study. Their questions and comments have pushed me to think more critically and carefully about the limitation of this study.

Also, I would like to show my gratitude to the staff in Faculty of Science for their guidance on the preparation of this dissertation.

I am also grateful for the research funding provided by University of Malaya, which allowed me to pursue my research wholeheartedly without worrying about financial issues.

Besides, I would like to acknowledge Dr. Yoon Tiem Leong from Universiti Sains Malaysia for providing us computing resources to carry out part of the calculations done in this work.

Last but not least, I would like to thank my family and friends for their moral support throughout the completion of this study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ADR	:	adverse drug reaction.
AML	:	acute myeloid leukemia.
CID	:	Compound Identification Number.
CLL	:	chronic lymphocytic leukemia.
CML	:	chronic myelogenous leukemia.
COSTART	:	Coding Symbols for a Thesaurus of Adverse Reaction Terms.
GBA	:	“guilt by association".
MCL	:	mantle cell lymphoma.
MDS	:	myelodysplastic syndromes.
MedDRA	:	Medical Dictionary for Regulatory Activities.
NER	:	Named Entity Recognition.
NLP	:	Natural Language Processing.
SIDER	:	Side Effect Resource.
STITCH	:	Search Tool for Interactions of Chemicals.
UDCA	:	ursodeoxycholic acid.
UMLS	:	Unified Medical Language System.

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Background

Network science is a growing research field that focuses on studying the structure of relationships between discrete objects or entities. A network is basically represented as a graph consisting of vertices (or nodes), which denote objects or entities, connected by edges (or links) through certain types of relations or interactions between the objects or entities. Mathematically, a graph is denoted as $G = (V, E)$ where V is the set of vertices and E is the set of edges.

A network can be directed or undirected depending on the types of relations represented by the edges. In an undirected network, the edges are symmetrical, representing symmetric relations between pairs of vertices such as friendship between individuals and collaboration between researchers. By contrast, in a directed network, the edges have directions (usually represented by lines with arrows) and are not inherently symmetric, such as in a citation network where an edge is formed from paper A to paper B if A cites B in its bibliography, and in a network of webpages where an edge represents a hyperlink from one webpage to another.

In some networks, the edges can be associated with numerical values indicating their strengths, providing a more insightful network analysis. For example, the edges in a friendship network may be weighted according to the frequencies of interactions between pairs of individuals, and the edges in a research collaboration network may carry weights representing the numbers of collaborations between pairs of researchers. Such networks are known as weighted networks.

One of the most interesting aspects of network analysis is the identification of vertices that are more “central" or “important" than others in a network. Many centrality measures

have been proposed to evaluate the structural importance of vertices in a network, with each of them looking into different topological aspects of the network as criteria for assessing the vertices (Bloch & Jackson, 2016; Boldi & Vigna, 2014; Lü et al., 2016). The three classic centrality measures in network science literature are degree, closeness and betweenness centralities (Borgatti & Everett, 2006; Freeman, 1978). Degree indicates the importance of a vertex based on the number of edges connected to it. Closeness centrality captures another notion of importance by considering the proximity of a vertex to all other vertices in a network. Betweenness centrality offers a different perspective that looks into the frequency in which a vertex lies on the shortest paths between any two other vertices. All these centrality measures, despite being defined differently, assert that a central vertex plays a significant role in a network and should provide some valuable insight for network analysis.

Indeed, the study of networks provides a powerful tool for understanding and modelling the behaviors of a wide variety of complex real world systems, leading to many useful applications in various disciplines including sociology, economics, biology, computer science, epidemiology and transportation systems (Easley & Kleinberg, 2010; Lewis, 2009; Newman, 2010; Wasserman & Faust, 1994).

1.2 Problem statement

The pioneering process of discovering and developing a completely new drug involves a substantial amount of costs and time, and is full of uncertainties. It takes about 10 to 15 years, at an approximate expense of 500 to 2000 million US dollars, for a drug to get through from the initial phase of preclinical testing to the final stage of successful launch in market (DiMasi et al., 2003; Pammolli et al., 2011). With concern on this productivity issue in pharmaceutical sector, there has been a growing interest among researchers to

explore an alternative drug development strategy known as drug repositioning or drug repurposing. Drug repositioning is the process of discovering new therapeutic uses of existing or approved drugs. Since existing drugs have well-established safety profiles, this can accelerate the process of drug development and reduce the associated risks. In recent years, drug repositioning has given a fresh impetus to a number of drugs. For example, minoxidil which is originally intended for treating hypertension is later repurposed to treat hair loss, and avastin which is initially used as a treatment for metastatic colon cancer and non small-cell lung cancer is later repositioned for treating metastatic breast cancer (Dudley et al., 2011).

Along with the advancement in technology, many drug- and disease-related data have become widely accessible, leading to the introduction of various computational approaches for drug repositioning over the last decade. In view of the versatility of network analysis across a wide variety of disciplines, a broad range of network-based drug repositioning approaches have been proposed (Shahreza et al., 2017; Xue et al., 2018). Depending on the type of data used as inputs, various types of networks can be constructed. For instance, vertices may represent biological elements such as proteins, genes, drugs, or diseases, while edges may represent relations or interactions between the vertices. The inference of drug repositioning opportunities then relies heavily on a key principle known as “guilt by association" (GBA) (Gillis & Pavlidis, 2012). In the context of drug repurposing, GBA principle can be translated into assumptions such as drugs with similar chemical structures are likely to share similar therapeutic roles, drugs with similar gene expression profiles tend to target similar proteins, target proteins with similar genomic sequence have a higher chance to interact with similar drugs, diseases with similar molecular pathology may be treated by similar drugs, and etc (Shahreza et al., 2017).

Thus, assessing similarity between drugs is one way to predict potential drug reposi-

tioning candidates. In principle, if two drugs share some similar features, the indication associated to only one of them may be considered as a new therapeutic function of the other drug (Dudley et al., 2011). By constructing a network of drugs to capture the relations between them, network-based theories and methodologies can then be applied to systematically screen drug pairs that could point towards potential drug repositioning candidates.

From the network perspective, a vertex can be associated with certain topological properties (which can be assessed using various centrality measures) based on its network position and connection pattern with other vertices in the network. Can these structural attributes be applied to a network of drugs to provide meaningful clues to the discovery of promising drug repositioning candidates? Are drugs in a more central network position associated with a more significant role in the drug repositioning task? These are the questions we aim to explore in this work.

1.3 Research objectives

1. To demonstrate the connection between drugs and the centralities of drugs through their positions in a network of shared side effects.
2. To manifest the possibility of using centrality measures to outline a network-based drug repositioning approach.

1.4 Outline and structure of thesis

This thesis is organized as follows. Chapter 1 describes the background of our research topic, problem statement and research objectives. Literature review is given in Chapter 2. We then present the data sets used in this study and describe our proposed methodology in Chapter 3. The results are discussed in Chapter 4. We then conclude by stating some limitations of this study as well as some possible future work in Chapter 5.

CHAPTER 2: LITERATURE REVIEW

One of the typical strategies to predict potential drug repositioning candidates is to infer them based on drug-drug similarity following the GBA principle which assumes that drugs with similar properties may share similar therapeutic indications. Similarities between drugs can be captured from several different aspects, such as their chemical structures and properties, interactions with target proteins, molecular activities (e.g., gene expression profiles), side effects and so on (Dudley et al., 2011; Shahreza et al., 2017).

In particular, side effects and therapeutic indications are both phenotypic outcomes resulting from a drug's underlying mechanism of action, where the desired effects are regarded as the drug's indications and the unexpected consequences constitute the drug's side effects. Hence, assessing drug similarities from the perspective of side effects, when compared independently to other drug-related properties, seems to offer more relevant insights into identifying drugs that may be linked to similar therapeutic functions. In fact, a few studies have compared different inputs for drug similarities on their abilities in predicting drugs with similar indications and their results suggest that side effects and therapeutic effects of drugs are strong predictors for one another (Wang et al., 2014; Zhang et al., 2013).

Following studies which show that side-effect information may be utilized for discovering novel indications of drugs (Campillos et al., 2008), Side Effect Resource (SIDER) – a public online database of drugs' side effects – was built by Kuhn et al. (2010). With 140,064 associations between 1430 drugs and about 5880 side effects in the latest version of the database, SIDER has been a main source of phenotypic information of drugs (Kuhn et al., 2016). For example, Yang and Agarwal (2011) used the side-effect information of drugs from SIDER in demonstrating their proposed methodology of Drug Repositioning based on

the Side-Effectome (DRoSEf). The method is founded on the core principle that if a drug shares side effects with many other drugs relating to a similar therapeutic indication, then this drug should be considered for being repurposed for the same therapeutic indication. Besides, Ye et al. (2014) also proposed a network-based drug repositioning approach which leverages the side-effect data of drugs. They constructed a drug network where a pair of drugs is connected by an edge if and only if their side-effect similarity score (measured by Jaccard index) surpasses a preset threshold. A drug's novel indications can then be inferred from its neighbors in the network.

Recently, Ali et al. (2017) also proposed a drug repositioning method based on analysis of drug networks, in which several centrality measures were incorporated for selecting drugs with better network positions known as prominent drugs. Those prominent drugs possess certain topological features which make their roles more significant than others in the networks. Hence, they were used as a benchmark for optimizing a threshold for drug similarities, for which new repositioning opportunities can be envisaged from drugs whose pairwise side-effect similarity scores meet the defined threshold.

The applications of centrality measures can be found in many domains (Das et al., 2018). Using some network topological properties as criteria, centrality measures are defined to rank vertices based on their relative structural prominence in a network, allowing identification of the top central vertices in the network which may have a significant contribution to some network applications. For example, in the biological domain, centrality measures have been found useful for identifying essential proteins which play indispensable roles in biological processes of organisms and are crucial for survival (Jalili et al., 2016; Jeong et al., 2001; Koschützki & Schreiber, 2008), leading to further extension to predict genes that may be related to a disease (Ozgur et al., 2008).

To date, over two hundreds of centrality measures have been proposed in literature

(Jalili et al., 2015; Oldham et al., 2019). The choice of which centrality measure to use is often subjected to the context of the network being analyzed. However, this study is not intended to compare or review different centrality measures in terms of their relevance to drug networks or for the application of drug repositioning [1]. Instead, our main focus here is to investigate whether the centrality rankings of drugs in a network reflect the significance of their roles in drug repositioning. In other words, is there any notable distinction between the central, peripheral and random drugs in a network in terms of their contribution to drug repositioning?

Note that in the previous study done by Ali et al. (2017) which suggests that prominent drugs can be used as a predictor of drug repositioning candidates, no findings were reported on the effect of using peripheral drugs to infer drug repositioning candidates. Also, it is not clear whether there is any difference in the prediction performances when prominent drugs are used and when they are not. Hence, we wish to illuminate the significance of using central drugs in the drug repositioning process by making a comparison to their peripheral and random counterparts.

[1] Readers who are interested in a comprehensive overview and discussion of centrality measures applied to drug repositioning may refer to Badkas et al. (2020).

CHAPTER 3: METHODOLOGY

We now introduce the methodology used for investigating our hypothesis, which states that if a drug known for treating a particular disease has a more central network position than others known for providing the same therapeutic effect, then the drug's network neighbors stand a higher chance to be indicated for the disease than the rest of the drugs in the network.

3.1 Data sets

The data used in this study were obtained from SIDER which is a free online database that provides information on drugs' side effects and therapeutic indications. The data of drugs' side effects were used for computing side-effect similarity scores between drugs, which were then used for linking drugs in the form of network. On the other hand, the data of drugs' therapeutic uses were used in the prediction step where we tried to predict new drugs that can be possibly linked to an indication by referring to the network neighbors of drugs already known to be associated with the indication.

The side-effect data in SIDER contains frequency information which reflects how likely a side effect will be induced in a patient treated with a drug. As a remark, we only included in our analysis the drug-side effect relations whose frequencies of occurrence are greater than or equal to 1%, or belong to the "common" and "very common" categories, assuming that side effects that occur rarely are underrepresented in a drug's profile.

Further details on these data sets are provided in Appendix A.1.

3.2 Network construction

We began by constructing a network of drugs based on similarities in their side-effect profiles. The most straightforward way to compare side-effect similarities among drug

pairs is through their numbers of shared side effects, which can be re-scaled to the range of 0 to 1 using the metric below:

$$N_{se}(d_i, d_j) = \frac{|se(d_i) \cap se(d_j)|}{\max_{k \neq l} |se(d_k) \cap se(d_l)|}, \tag{3.1}$$

where $se(d_i)$ and $se(d_j)$ represent the sets of side effects of drugs d_i and d_j respectively. Note that the denominator in N_{se} is a constant equivalent to the maximum number of side effects shared between the drug pairs. Thus, the greater the number of side effects shared between a pair of drugs, the higher the value of N_{se} for the drug pair. Particularly, in the scenario where $N_{se}(d_i, d_j) = 1$, the pair of drugs d_i and d_j outperforms all other drug pairs in the dataset in terms of their numbers of shared side effects.

Another way to capture side-effect similarity between two drugs is by using the Jaccard index:

$$\begin{aligned} J_{se}(d_i, d_j) &= \frac{|se(d_i) \cap se(d_j)|}{|se(d_i) \cup se(d_j)|} \\ &= \frac{|se(d_i) \cap se(d_j)|}{|se(d_i)| + |se(d_j)| - |se(d_i) \cap se(d_j)|}. \end{aligned} \tag{3.2}$$

Note that the metric J_{se} measures the portion of overlap between the sets of side effects of two drugs. For example, $J_{se}(d_i, d_j) = 1$ indicates that the side-effect profiles of the two drugs d_i and d_j are completely identical, even though the number of side effects shared between them may be less than those of other drug pairs.

The two metrics relate to each other. Nevertheless, rather than yielding a common similarity ranking, they complement each other to provide a more fine-tuned similarity score. For instance, consider the three pairs of drugs in Table 3.1 whose information on side effects were obtained from SIDER database. The first and second drug pairs have the

Table 3.1: Comparison of the similarity metrics N_{se}, J_{se}, and SIM_{se} for three pairs of drugs.

Drug pair, (d_i, d_j)	Elvitegravir, Etravirine	Acebutolol, Bromocriptine	Ropinirole, Sertraline
$\lvert se(d_i) \cap se(d_j) \rvert$	7	7	48
$\lvert se(d_i) \cup se(d_j) \rvert$	9	35	240
$N_{se}(d_i, d_j)$	$\frac{7}{48}$	$\frac{7}{48}$	1
$J_{se}(d_i, d_j)$	$\frac{7}{9}$	$\frac{7}{35} = \frac{1}{5}$	$\frac{48}{240} = \frac{1}{5}$
$SIM_{se}(d_i, d_j)$	0.11343	0.02916	0.2

same number of shared side effects, so the metric N_{se} does not distinguish the similarity of the first drug pair from that of the second drug pair, but the Jaccard index indicates that the first drug pair is more similar than the second one. On the other hand, N_{se} implies that the third pair of drugs is much more similar than the second pair, while the Jaccard index fails to do so.

In view of these observations, we consider a combination of the two metrics as the drug side-effect similarity score, i.e.,

$$SIM_{se}(d_i, d_j) = N_{se}(d_i, d_j) \times J_{se}(d_i, d_j) \tag{3.3}$$

Once the pairwise side-effect similarity scores for all drugs in the dataset were computed, we then formed a network G_{se} with 864 drugs depicted as vertices, and 303417 edges carrying weights equivalent to SIM_{se} between the connected drugs. In particular, the edges only exist between drugs whose SIM_{se} is non-zero. Note that G_{se} is a weighted network with edges attached to weights that correspond to their strengths of relations.

3.3 Centrality measures

In this section, we introduce the precise definitions of the three fundamental centrality measures that we applied to the analysis of our weighted drug network G_{se}.

3.3.1 Degree centrality

Typically, the degree centrality of a vertex is defined as the number of edges linked to it. In other words, it is usually a local measure which simply captures the size of the immediate neighborhood of a vertex.

In the case of a weighted network, the weights of edges carry valuable information, and hence centrality measures can reveal more appropriate insights into the importance of vertices if the weights of edges are taken into account in the calculations.

Thus, a straightforward way to extend the definition of degree centrality to a weighted network is by summing up the weights of edges linked to a vertex. An even more comprehensive definition of degree centrality which encompasses both the number of edges and the total weights of edges connected to a vertex was proposed by Opsahl et al. (2010). Their proposed definition for the degree centrality of a vertex v is given precisely as follows:

$$C_{Deg}^{\alpha}(v) = n(v)^{1-\alpha} \times s(v)^{\alpha}, \tag{3.4}$$

where $n(v)$ represents the number of edges connected to vertex v, $s(v)$ represents the sum of weights of edges connected to vertex v, and α is an adjustable parameter which can be set according to the network setting to determine the relative weights of $n(v)$ and $s(v)$ in defining the centrality. For instance, if $\alpha = 0$, then $C_{Deg}^{\alpha}(v) = n(v)$ simply ranks the vertices according to the number of edges linked to each of them, which is essentially just the unweighted version of degree definition. On the contrary, if $\alpha = 1$, then $C_{Deg}^{\alpha}(v) = s(v)$ measures a vertex's centrality by only taking into consideration of the total weights of

edges linked to the vertex.

For example, consider the weighted network G in Figure 3.1 which consists of 5 vertices and 7 weighted edges. By setting α to three different values, i.e. $\alpha = 0$, $\alpha = 0.5$, and $\alpha = 1$, in Eq. 3.4, we observe that each of them results in different rankings of vertices. The centrality values for each vertex, when α is set to different values, are presented in Table 3.2. For instance, consider the vertex B in the network which receives the highest ranks in all three different settings for degree centrality. When $\alpha = 0$, $C^{\alpha}_{Deg}(B) = n(B) = 4$ because vertex B is connected to 4 vertices in the network; when $\alpha = 1$, we have $C^{\alpha}_{Deg}(B) = s(B) = 4 + 5 + 1 + 3 = 13$ which is the sum of weights of edges connected to vertex B; when $\alpha = 0.5$, we obtain $C^{\alpha}_{Deg}(B) = n(B)^{0.5} \times s(B)^{0.5} = 4^{0.5} \times 13^{0.5} = 7.2111$. Besides, note that when $\alpha = 0.5$, the centrality separates the five vertices to different ranks in the order of B, A, C, E and D from the most central to least central. In contrast, when $\alpha = 0$ and $\alpha = 1$, the centrality is not able to distinguish some vertices in terms of their centrality ranking. For example, vertices A and C have equal centrality value when $\alpha = 0$ since each of them is connected to 3 edges in the network. Likewise, vertices A and B are ranked together as the most central when $\alpha = 1$, since the total weights of edges connected to each of them are equivalent.

Table 3.2: Degree centralities of vertices in the network G shown in Figure 3.1 according to different values of α.

Vertex	C^{α}_{Deg}, $\alpha = 0$	C^{α}_{Deg}, $\alpha = 0.5$	C^{α}_{Deg}, $\alpha = 1$
A	3	6.2450	13
B	4	7.2111	13
C	3	5.4772	10
D	2	3.1623	5
E	2	3.7417	7

The definition of degree centrality stated in Eq. 3.4 can be translated in a context relevant to our drug network G_{se}, whereby $n(v)$ represents the number of drugs sharing

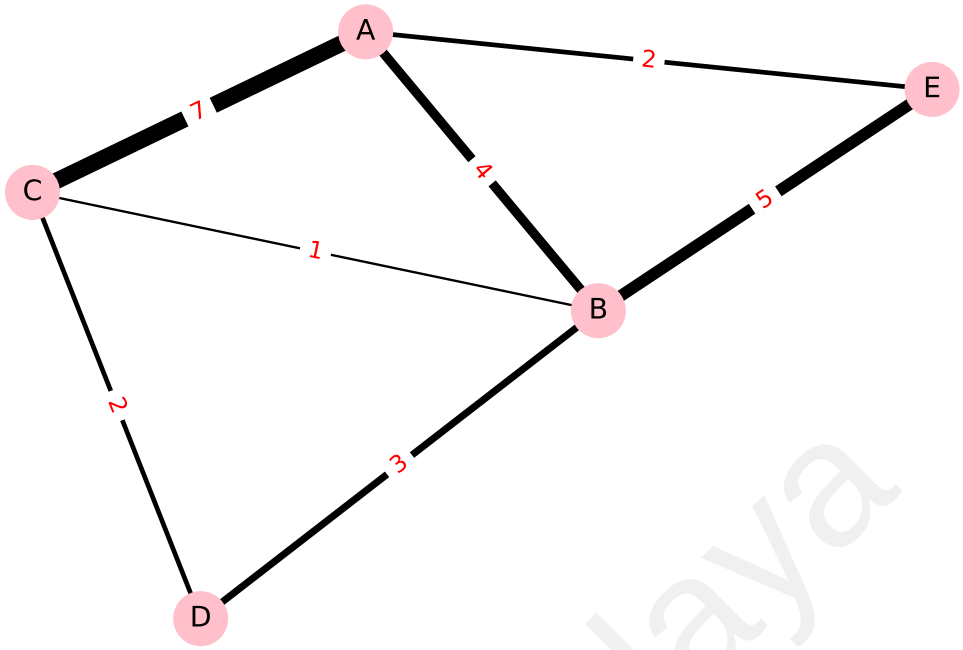


Figure 3.1: An example of a weighted network G, where edge weight corresponds to the strength of relation such that greater edge weight indicates stronger link between the vertices.

some side effects with drug v, and $s(v)$ indicates the extent to which drug v is similar to other drugs in the network. In our opinion, $n(v)$ and $s(v)$ are both important aspects that are worth the same amount of attention when deciding a drug's local prominence in the network. Hence, the degree centrality defined in Eq. 3.4 is applied to our drug network G_{se}, with α set at the value of 0.5 [1].

3.3.2 Closeness centrality

Closeness centrality essentially rates a vertex by its proximity to each of the other vertices in a network. The distance between a pair of vertices in a network is calculated based on the paths connecting them. A path between two distinct vertices u and v is a sequence of adjacent vertices (i.e. vertices that are connected by an edge) that starts with u

[1] We do not claim that the choice of $\alpha = 0.5$ for C_{deg} is the most suitable one or is statistically significant for capturing the local prominence of each vertex in our drug network. The particular value of $\alpha = 0.5$ is chosen to only demonstrate the case when both factors $n(v)$ and $s(v)$ are considered equally for measuring the local prominence of a vertex v in the weighted network.

and ends with v, in which no vertex in the sequence is repeated.

The length of a path in a weighted network is usually measured by the sum of weights of edges along the path. In a network G where all edges carry non-negative weights, the shortest paths can be found by Dijkstra's algorithm (Dijkstra, 1959). Such paths are appropriate for connecting vertices in a network where edge weight represents a form of cost such that greater edge weight indicates a weaker or costlier connection, since vertices are more likely to be connected through paths with minimum total costs (weights). However, in some other networks (for instance, our drug network), edge weight may represent the strength of relation such that greater edge weight indicates a stronger link, and thus a path with maximum total strengths (weights) may be preferable for connecting a pair of vertices. In that case, we can alter each edge weight in the original network G to its reciprocal and then obtain a network G' in which the shortest paths identified using Dijkstra's algorithm are the ones with maximum strength in the original network G. That is, in a network G where greater edge weight indicates closer relationship, the length of a path can be calculated as the sum of inverted weights of its constituent edges. Formally, suppose $[u, h_1, h_2, \ldots, h_k, v]$ is a sequence of vertices on a path p, and let $w_{i,j}$ represent the edge weight between any two adjacent vertices i and j. Then the length of the path p is calculated as follows:

$$l_p = \frac{1}{w_{u,h_1}} + \frac{1}{w_{h_1,h_2}} + \cdots + \frac{1}{w_{h_k,v}}. \tag{3.5}$$

Note that there may be more than one path linking the two vertices u and v (see Table 3.3). In particular, the shortest one among them is known as a geodesic path between u and v,

and its length gives the (geodesic) distance between the two vertices, i.e.

$$d(u, v) = \min_{p \in P(u,v)} l_p, \tag{3.6}$$

where $P(u, v)$ is the set of all paths between u and v. See Table 3.4 for the geodesic distance computed for each pair of vertices in the network G in Figure 3.1.

Finally, the closeness centrality of a vertex v in G is computed as the inverse of the sum of its geodesic distances to all other vertices in a network as follows:

$$C_{Clo}(v) = \left[\sum_{u \in V(G)-v} d(u, v) \right]^{-1}, \tag{3.7}$$

where $V(G)$ is the set of vertices in G, and $d(u, v)$ is the geodesic distance between vertices u and v. As an example, Table 3.5 shows the closeness centrality of each vertex in the network G shown in Figure 3.1.

Table 3.3: The length of each path between vertices A and E in the network G shown in Figure 3.1. The length of each path is calculated as the sum of inverted weights of the edges along the path. Note that $[A, B, E]$ is the geodesic path between vertices A and E, so its length gives the geodesic distance between A and E.

Path	Path length
$[A, E]$	0.5
$[A, B, E]$	0.45
$[A, C, B, E]$	1.3429
$[A, C, D, B, E]$	1.1762

3.3.3 Betweenness centrality

The definition of betweenness centrality also involves the concept of geodesic path. As in the case of closeness centrality, we identified a geodesic path between a pair of distinct vertices in our drug network as a path that connects the vertices via edges with the least sum of inverted weights. Suppose g_{ij} is the number of geodesic paths connecting two

Table 3.4: Geodesic path and distance for each pair of vertices in the network G shown in Figure 3.1.

Vertex pair	Geodesic path	Geodesic distance
(A, B)	$[A, B]$	0.25
(A, C)	$[A, C]$	0.1429
(A, D)	$[A, B, D]$	$0.58\overline{3}$
(A, E)	$[A, B, E]$	0.45
(B, C)	$[B, A, C]$	0.3929
(B, D)	$[B, D]$	$0.\overline{3}$
(B, E)	$[B, E]$	0.2
(C, D)	$[C, D]$	0.5
(C, E)	$[C, A, B, E]$	0.5929
(D, E)	$[D, B, E]$	$0.5\overline{3}$

Table 3.5: Closeness centrality of each vertex in the network G shown in Figure 3.1.

Vertex	Closeness centrality, C_{Clo}
A	0.7012
B	0.8502
C	0.6140
D	0.5128
E	0.5630

different vertices i and j, and $g_{ij}(v)$ is the number of such paths that pass through a vertex v, where $v \notin \{i, j\}$. Then the betweenness centrality of v in network G is defined as:

$$C_{Btw}(v) = \sum_{i,j \in V(G)-v, i \neq j} \frac{g_{ij}(v)}{g_{ij}}, \tag{3.8}$$

where $V(G)$ is the set of vertices in the network G that v belongs to.

Take the weighted network shown in Figure 3.1 as an example. From Table 3.4, we can see that vertex A appears as the intermediary vertex in the geodesic path $[B, A, C]$ between the pair of vertices B and C, and also in the geodesic path $[C, A, B, E]$ between the vertices C and E. Since there is only 1 geodesic path for each pair of vertices, hence we have $C_{Btw}(A) = 1/1 + 1/1 = 2$. See Table 3.6 for a complete list of betweenness centralities of vertices in the same network.

Table 3.6: Betweenness centrality of each vertex in the network G shown in Figure 3.1.

Vertex	Betweenness centrality, C_{Btw}
A	2
B	4
C	0
D	0
E	0

3.3.4 Convex combinations of centrality measures

In general, each centrality measure focuses on a particular structural attribute of vertices and provides unique insights into the significance, influence and roles of each vertex in a network. However, this also implies that each centrality measure has its own limiting perspective when evaluating the vertices. For instance, when using Kendall's tau-b correlation measure to compare the ranking of vertices in G_{se} yielded by betweenness centrality to that given by closeness and degree centralities, the values obtained are 0.4106 and 0.4239 respectively, out of the perfect positive correlation value of 1. (See Table 3.7.)

This can be resolved by considering combinations of several centrality measures. For example, it was shown that combinations of at least two centrality measures achieved a more reliable prediction as compared to the use of single centrality measures when predicting essential genes from genetic networks (del Rio et al., 2009).

In this study, we present a simpler case of using convex combinations of two centrality measures in analyzing our drug network, and then compare the results with the ones obtained by their constituent centrality measures. As demonstrated in a recent study by Keng et al. (2020), a convex combination of two arbitrary centrality measures C_0 and C_1 can be formulated for any vertex v in a network as follows:

$$C_t(v) = (1-t)C_0(v) + tC_1(v), \tag{3.9}$$

where t is some constant in the range from 0 to 1 which decides how much each of the individual centralities C_0 and C_1 contributes to the combination.

An appropriate t-value is then determined based on Kendall's tau-b correlation coefficient (τ_b) to form an optimum convex combination that resembles both of its constituent centralities in terms of their ways of ranking the vertices. The correlation coefficient τ_b quantifies the similarity between two rankings in the range from -1 to 1, where a value of -1 indicates that the two rankings are completely different, a value of 1 indicates that the two rankings are perfectly identical, and a value of 0 indicates that there is no association between the two rankings (Kendall, 1938).

Let $\tau_b(C_s, C_t)$ be the Kendall's correlation between the two centrality rankings induced by C_s and C_t for any $s, t \in [0, 1]$. One way to form an optimum convex combination C_t which resembles the two original centralities C_0 and C_1 as much as possible is to find a $t \in [0, 1]$ such that both $\tau_b(C_0, C_t)$ and $\tau_b(C_1, C_t)$ are as close as can be to their maximum possible value, i.e. 1. This can be obtained by finding the t-value which minimizes the following function:

$$f(t) = [1 - \tau_b(C_0, C_t)]^2 + [1 - \tau_b(C_1, C_t)]^2. \tag{3.10}$$

Note that such a t-value is guaranteed to exist, since there are only finitely many possible values of $\tau_b(C_s, C_t)$ for any $s, t \in [0, 1]$ (Keng et al., 2020).

A point t is regular if C_t separates all the vertices in a network by centrality rankings, except for the vertices that are already ranked equally by both the constituent centralities C_0 and C_1. A convex combination C_t induced by a regular point t has several desired characteristics. Firstly, it follows from the definition of a regular point that C_t is able to distinguish as many vertices as possible by their centrality ranks based on the ones induced

by the constituent centralities C_0 and C_1. Besides, small changes in the t-value do not affect the centrality ranking yielded by C_t. Lastly, regular points abound in the interval $[0, 1]$ in the sense that they can be found with probability 1.

We then incorporate the algorithm outlined by Keng et al. (2020) to find regular points that give rise to optimum pairwise combinations of the centralities C_{Deg}, C_{Clo} and C_{Btw} for our drug network. As a result, we obtain the following pairwise combinations

- $0.5236C_{Clo} + 0.4764C_{Deg}$
- $0.8855C_{Btw} + 0.1145C_{Clo}$
- $0.9C_{Btw} + 0.1C_{Deg}$

for our network.

As is pointed in Keng et al. (2020), an efficient algorithm to find optimum convex combinations of more than two centrality measures for large networks is yet to be constructed. This is due to the dimensions of the simplices involved. Hence, we are contented with combinations of two centrality measures for the moment.

3.3.5 Centrality analysis

We now provide a more detailed analysis of centralities for our drug network G_{se}. Firstly, we note that closeness and degree are strongly correlated with $\tau_b(C_{Clo}, C_{Deg}) = 0.9079$. Remarkably, their optimum convex combination has even stronger correlation with each of the constituent centralities, with $\tau_b \approx 0.95$.

In contrast, betweenness does not correlate as much with each of closeness and degree, with $\tau_b(C_{Btw}, C_{Clo}) = 0.4106$ and $\tau_b(C_{Btw}, C_{Deg}) = 0.4239$ respectively. However, the optimum combination of betweenness and closeness correlates with each of its atomic centralities with $\tau_b \approx 0.7$ – a significant increase from the correlation between the two atomic centralities themselves.

Similarly, the optimum convex combination which consists of 90% of betweenness and 10% of degree also has a correlation of $\tau_b \approx 0.7$ with each of its constituent centralities – again indicating an improvement from the correlation between the constituent centralities.

The pairwise Kendall's correlations for the three centralities and their optimum convex combinations in pairs are shown in Table 3.7.

Table 3.7: Kendall's τ_b correlations between the centralities applied in the analysis of G_{se}.

	Degree	Closeness	Betweenness	Closeness & Degree	Betweenness & Closeness	Betweenness & Degree
Degree	1	0.9079	0.4239	0.9540	0.6815	0.7122
Closeness	0.9079	1	0.4106	0.9539	0.7056	0.6724
Betweenness	0.4239	0.4106	1	0.4194	0.7179	0.7243
Closeness & Degree	0.9540	0.9539	0.4194	1	0.6996	0.6931
Betweenness & Closeness	0.6815	0.7056	0.7179	0.6996	1	0.9238
Betweenness & Degree	0.7122	0.6724	0.7243	0.6931	0.9238	1

3.4 Prediction of drug repositioning candidates

In this section, we present a computational procedure to uncover potential drug repositioning candidates for a disease based on the centralities of drugs associated to the disease. Following the principle that drugs sharing similar side effects may share similar therapeutic indications, we assume that the drugs that are directly connected to a drug d in the network G_{se} (i.e. direct neighbors of d in G_{se}) are more likely to share the same indications with d than those that are not.

To get a hint on how centralities of drugs can indicate the drugs' abilities to repurpose their neighbors for treating the same disease, take a look at three subgraphs of G_{se} shown in Figure 3.2, which depict parts of the neighborhoods of three different drugs known for

treating the disease leukemia. The three drugs have the highest, median and lowest degree centralities respectively among all drugs associated with leukemia. From the figures, we see that the one with higher degree centrality has higher SIM_{se} with each of its ten most similar neighbors in the network that are not known for treating the disease. Since similar drugs are likely to share a similar therapeutic effect, we hypothesize that the neighbors of the more central drug are more likely to be repurposed successfully for treating the same disease than that of the less central one.

Let D_s be the set of existing drugs known for treating a disease s. Our aim is to find a drug (other than those in D_s) that could be repositioned as an alternative drug for the disease. We first identify the top $k\%$ of the most central drugs in D_s using a centrality measure, where k is any value that satisfies $1 \leq k\%|D_s| \leq |D_s|$.

Let $T_s(c)$ be the set of top $k\%$ of the most central drugs in D_s chosen with respect to a centrality measure c. The collection of neighbors of each drug in $T_s(c)$, excluding those that are already included in D_s, are then considered as candidate drugs in which their relations to disease s will be further evaluated. These candidate drugs can be generally found as follows:

$$P_s^*(c) = \bigcup_{d \in T_s(c)} nb(d) - D_s, \tag{3.11}$$

where $nb(d)$ is the set of direct neighbors of drug d in the network G_{se}. Then each of the candidate drugs is assigned a score that represents the likelihood that the drug is a valid repositioning candidate for disease s. In particular, the score for a candidate drug d^* in relation to treating disease s is computed as its average similarity score with the drugs in $T_s(c)$:

$$Score(d^*, s) = \frac{\sum_{d \in T_s(c)} SIM_{se}(d^*, d)}{|T_s(c)|}. \tag{3.12}$$

The higher the score, the more likely the drug can be used for treating disease s, since it

has overall higher similarity score with the central drugs associated with the disease.

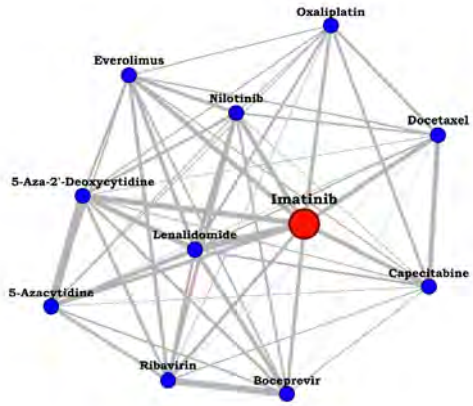
Let $P_s(c)$ represent the list of top-20 drug candidates in $P_s^*(c)$ whose scores are the highest. These drugs were then predicted as the final list of the most potential alternative drugs for treating disease s. To measure the clinical relevance of our prediction, we compared the predicted drugs with the existing clinical trials registered in ClinicalTrials.gov database. For a conservative evaluation, we assume our prediction of a drug-disease relation is "positive" if and only if it has been investigated in at least three clinical trials in the database [2]. Finally, we quantified the positive predictive value of our prediction in $P_s(c)$ as follows:

$$PPV_s(c) = \frac{|V_s(c)|}{|P_s(c)|} = \frac{|V_s(c)|}{20}, \tag{3.13}$$

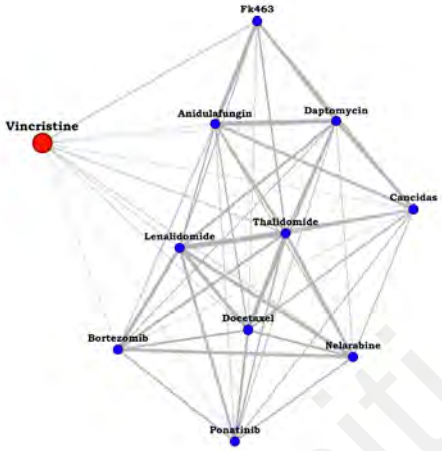
where $V_s(c)$ is the set of drugs in $P_s(c)$ in which their relations with disease s have been studied in at least three clinical trials.

For comparison purpose, we carried out a similar procedure of prediction using the bottom k% of the least central drugs and a randomly chosen k% of drugs from D_s respectively. The predictions achieved by k% of central, peripheral and randomly chosen drugs were then compared while varying the values of k, and the results are presented in the next chapter. The outline of our procedure is summarized in Figure 3.3.

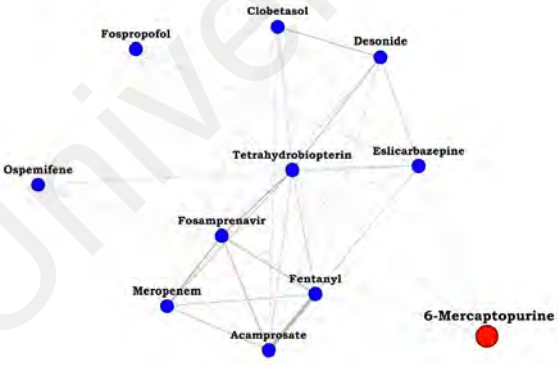
[2] As a remark, we only include the clinical trials from 1st January 2000 to 1st January 2020.



(a) Neighborhood of the drug Imatinib, which has the highest degree centrality among all the drugs approved for treating leukemia.



(b) Neighborhood of the drug Vincristine, in which its degree centrality is in the median among all the drugs approved for treating leukemia.



(c) Neighborhood of the drug 6-Mercaptopurine, which has the lowest degree centrality among all the drugs approved for treating leukemia.

Figure 3.2: Subgraphs of G_{se}. The red node represents a drug known for treating leukemia, whereas the blue nodes represent the drugs that have the highest SIM_{se} with the red node, after excluding those approved for treating leukemia. The weights of the edges, which represent SIM_{se} between the drugs, can be visualized based on the thickness of the edges in the graphs, such that a thicker edge has greater edge weight.

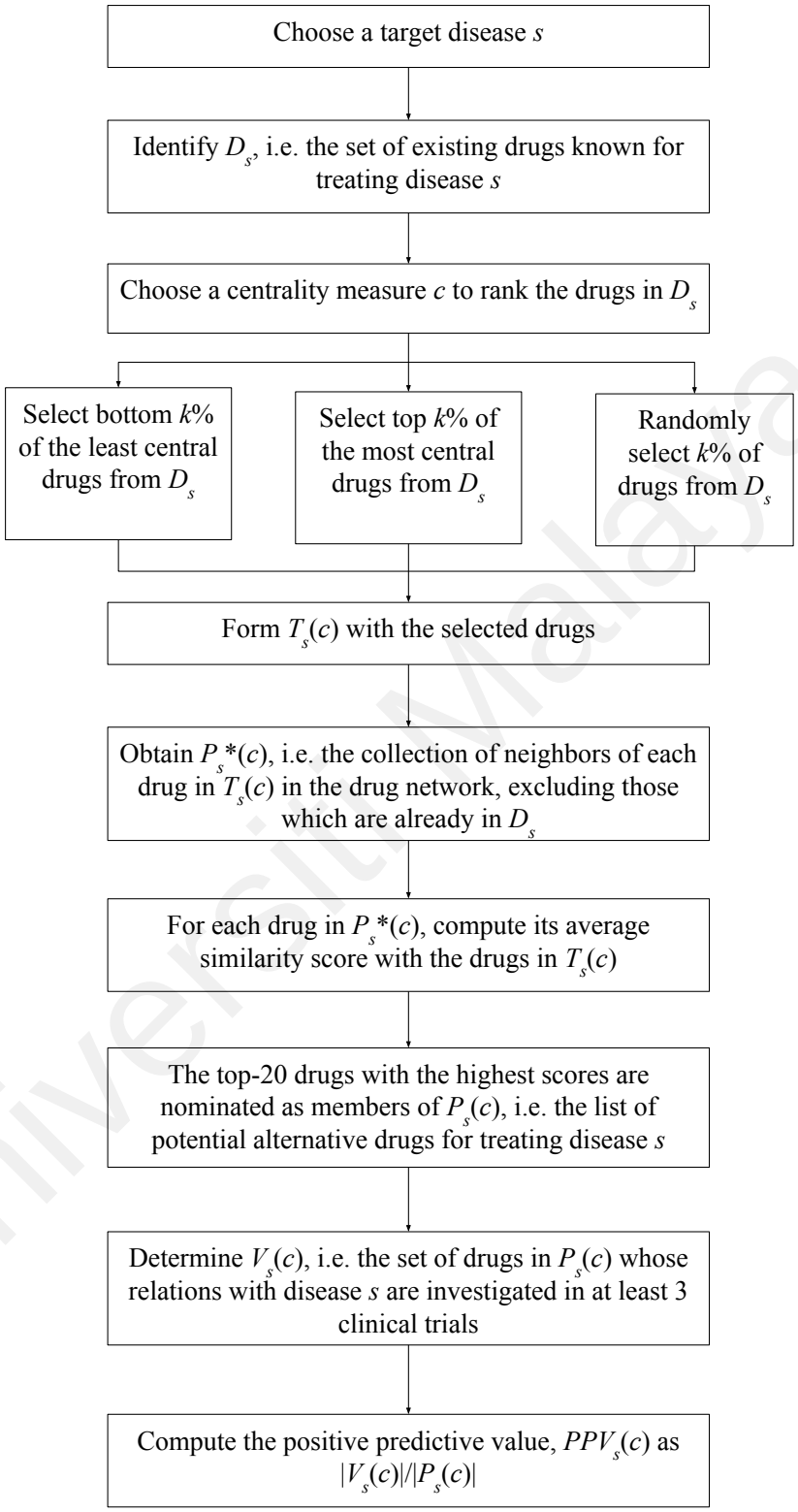


Figure 3.3: An outline of our procedure for predicting potential drug repositioning candidates for a target disease and assessing the resulting prediction performance.

CHAPTER 4: RESULTS AND DISCUSSION

A numerical experiment was conducted based on the presented methodology, and the results are discussed in this chapter.

4.1 Selection of target diseases

In this study, we selected the top 10 diseases or conditions with the highest numbers of clinical studies in the ClinicalTrials.gov database as our target diseases for prediction of their potential alternative drugs:

(1) cardiovascular diseases (CVDs),

(2) gastrointestinal diseases,

(3) mental disorders,

(4) dermatologic disorders,

(5) wounds and injuries,

(6) virus diseases,

(7) breast carcinoma,

(8) lymphoma,

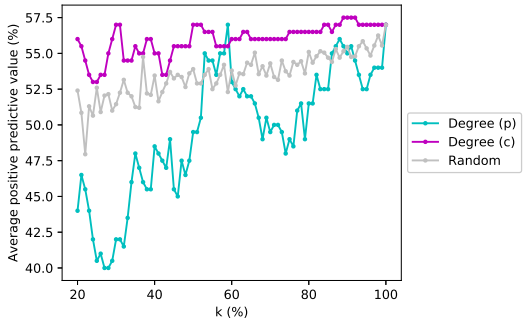
(9) acquired immunodeficiency syndrome,

(10) leukemia.

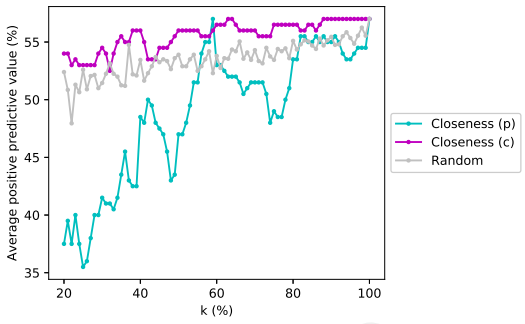
4.2 Prediction results

Let k be the percentage of drugs selected from D_s for each target disease s listed above. The drugs were selected from D_s in three ways, which consist of 1) the top k% of the most central drugs in D_s, 2) the bottom k% of the least central drugs in D_s, and 3) a randomly chosen k% of drugs from D_s. By referring to the neighborhoods of these selected drugs, we predicted the top-20 most potential alternative drugs for each target disease s. The

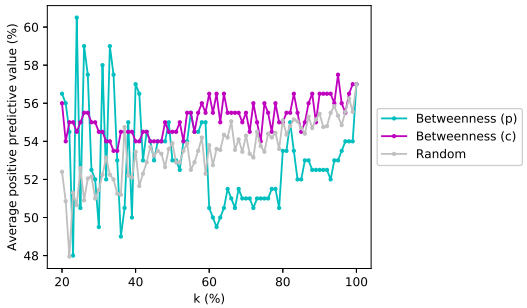
prediction results obtained by different ways of selecting k% of drugs in each D_s were then compared based on their average positive predictive values for the 10 target diseases.



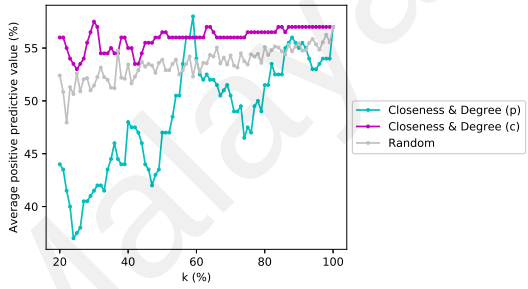
(a) Degree centrality.



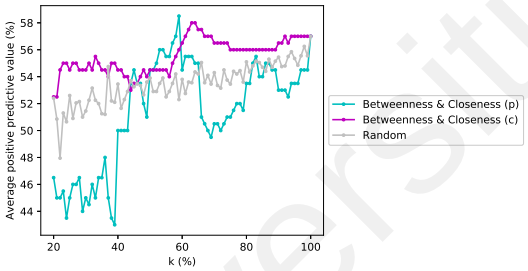
(b) Closeness centrality.



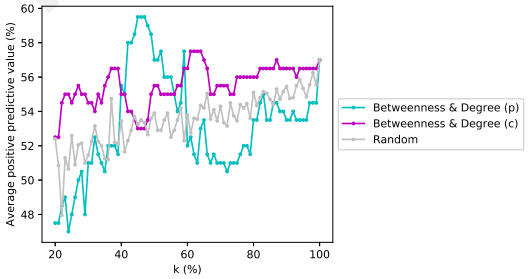
(c) Betweenness centrality.



(d) Optimum convex combination of closeness and degree centralities.



(e) Optimum convex combination of betweenness and closeness centralities.



(f) Optimum convex combination of betweenness and degree centralities.

Figure 4.1: Comparison of average positive predictive values between the top k% of central drugs, the bottom k% of peripheral drugs chosen from each D_s using different centrality measures, and the k% of drugs randomly chosen from each D_s, while varying the values of k.

Figure 4.1 shows the average positive predictive values obtained using different centrality measures when the values of k are varied. As a remark, the letter 'c' stated beside the centrality indicates that the top k% of the most central drugs are selected from each D_s, while the letter 'p' indicates that the bottom k% of the peripheral or least central drugs are selected instead. As for the case of 'random', k% of drugs are chosen randomly

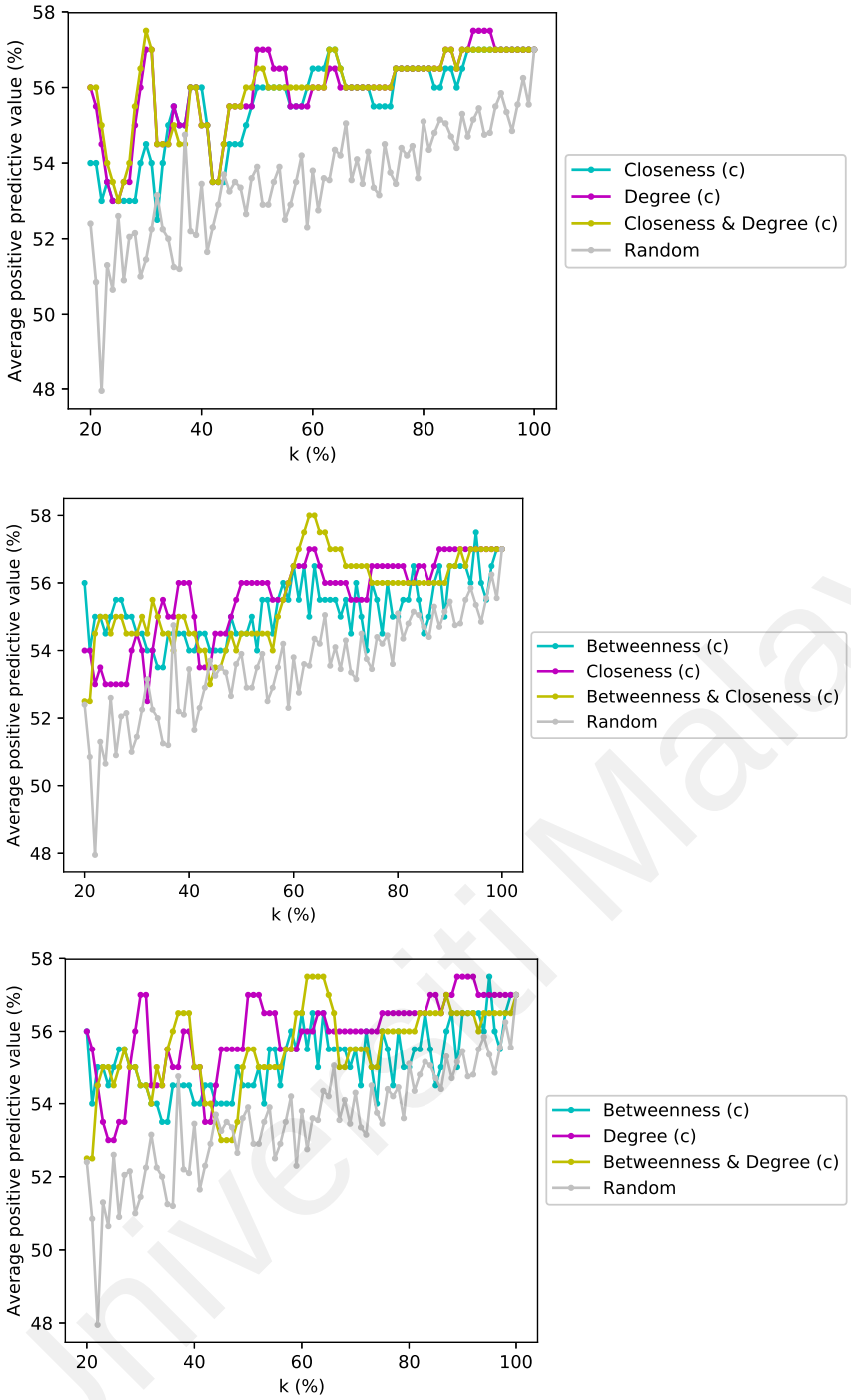


Figure 4.2: Comparison between optimum convex combinations of centralities and their constituent centralities in terms of the average positive predictive values achieved by the top k% of central drugs from each D_s while varying the values of k.

from each D_s for ten times, and the average positive predictive values are then taken over the ten trials. Apart from that, note that the k-value is chosen in a way so that at least 1 drug is chosen from each D_s for prediction, for otherwise there would be no reference point for which the candidate drugs can be predicted.

As seen from the plots in Figure 4.1, when k is varied from 20% to 100%, the average positive predictive values of our prediction when using k% of central drugs associated with the target diseases are consistently above 52% for all centrality measures including their optimum convex combinations. In contrast, the prediction using peripheral drugs achieve the lowest average positive predictive value of merely 35.5% when k% = 25% of least central drugs are chosen from each D_s with respect to closeness centrality. Besides, the average positive predictive values given by peripheral drugs fluctuate wildly as k varies, while that yielded by the central drugs perform relatively stable across all k-values. Furthermore, we observe that the central drugs identified by each centrality measure perform significantly better than the randomly chosen drugs for any value of k. Although the predictions given by peripheral drugs are not steady nor reliable, they yield "surprises" at some k-values. For instance, when k% = 24% of drugs with lowest betweenness are chosen from each D_s, it results in a surprisingly impressive prediction with average positive predictive value equal to 60.5% – which is in fact the highest average positive predictive value obtained among all the cases in our experiment.

Note that as the value of k increases, there is an increasing overlap between the set of top k% of central drugs associated with a target disease and the sets of their peripheral and random counterparts. Thus, it is no surprise that the average positive predictive values achieved with k% of peripheral drugs and k% of randomly chosen drugs associated with the target diseases get closer to that achieved with k% of top central drugs when the value of k increases.

On the other hand, Figure 4.2 shows the comparison between the optimum convex combinations and their constituent centralities. Overall, the central drugs identified by the optimum convex combinations rarely perform worse than that identified by their constituent centralities in capturing valid drug repositioning candidates for the target diseases. This

implies that the optimum convex combinations can be considered as a measure for a safe evaluation.

Recall that the positive predictive value is calculated based on the presence of clinical trials that investigate a drug candidate in relation to a disease. More precisely, we consider the prediction of a drug-disease relation is positive if and only if there exist at least three clinical trials (in ClinicalTrials.gov database) that investigate the association between the drug and the disease. Hence, it is important to note that an absence or lack of clinical trials associated with a drug-disease pair does not necessarily imply that the drug is invalid for treating the disease, but that the relation has not (yet) gained enough attention for clinical study. Likewise, the presence of clinical trials associated with a drug-disease pair does not necessarily imply that the drug has a positive effect on treating the disease, but simply that there has been considerable interest in studying the drug-disease relation in the clinical stage. Hence, the positive predictive values reported in this study simply reflect the alignment of our prediction with the *existing* clinical interests, and may not imply the actual validity of the predicted drug-disease relations.

4.3 Case studies

We further study the drug candidates predicted for two of the target diseases – leukemia and cardiovascular diseases. Table 4.1 shows the lists of potential drug candidates that appeared in our predictions when 50% of central drugs associated with each disease were chosen with respect to each centrality measure.

Leukemia

Here, we discuss some of the potential drugs for leukemia which are listed in Table 4.1. Leukemia is a cancer that begins in blood-forming tissue, such as bone marrow, and causes large numbers of abnormal blood cells to be produced and enter the bloodstream. There

Table 4.1: Lists of potential drug candidates for leukemia and cardiovascular diseases, when the top 50% of central drugs approved for each disease were identified using each centrality measure and used for prediction. Note that the relation between each target disease and each of its potential drug candidates is not known in the original data set we retrieved from SIDER. The number of clinical studies (as recorded in ClinicalTrials.gov) related to each predicted drug-disease relation is presented as follows.

(a) List of predicted drug candidates for leukemia.

Drug candidates	Number of clinical studies
Ribavirin CID100005064	5
5-Azacytidine CID100001805	300
Capecitabine CID100060953	1
Rivastigmine CID100005077	0
Oxaliplatin CID100004609	9
CGP 19835 A CID105479141	0
Lenalidomide CID100216326	149
Thalidomide CID100005426	22
Temozolomide CID100005394	8
Nilotinib CID100644241	105
Bexarotene CID100082146	5
5-Aza-2'-Deoxycytidine CID100016886	220
Docetaxel CID100003143	2
Everolimus CID106442177	84
Boceprevir CID110324367	0
Anagrelide CID100002182	2

(b) List of predicted drug candidates for cardiovascular diseases.

Drug candidates	Number of clinical studies
Citalopram CID100002771	23
Bupropion CID100000444	8
Imatinib CID100005291	11
Alprazolam CID100002118	0
Fenofibrate CID100003339	31
Exemestane CID100060198	0
Duloxetine CID100060834	1
UDCA CID100005645	3
K779 CID171306834	0
Paroxetine CID100004691	5
Risedronate CID100005245	0
Vacv CID100005647	0
Quetiapine CID100005002	1
Rivastigmine CID100005077	4
Meloxicam CID100004051	2
Fluoxetine CID100003386	19
Thalidomide CID100005426	158

are several types of leukemia, which are divided based mainly on whether the leukemia is acute (fast growing) or chronic (slower growing), and whether it starts in myeloid cells or

lymphoid cells (American Cancer Society, n.d.).

The drug ribavirin was first discovered in 1972 as a broad-spectrum anti-viral drug (Sidwell et al., 1972). It is conventionally used in standard treatment of Hepatitis C infection (Wohnsland et al., 2007). However, recent studies have shown that ribavirin led to clinical improvement in treating patients with acute myeloid leukemia (AML) (Assouline et al., 2009, 2014; Borden & Culjkovic-Kraljacic, 2010; Shi et al., 2015). Although 5-azacytidine and 5-aza-2'-deoxycytidine were approved by the U.S. Food and Drug Administration (FDA) for the treatment of myelodysplastic syndromes (MDS), the two drugs have been consistently used in patients with AML and chronic myelogenous leukemia (CML) (Momparler et al., 1984; Mund et al., 2005; Rohon et al., 2012; Savona et al., 2015). The drug oxaliplatin, which is an approved medication for treating advanced colorectal cancer and stage III colon cancer (National Cancer Institute (NCI), n.d.-b), also has been investigated in combination with other drugs for patients with relapsed or refractory AML (Tsimberidou et al., 2014, 2008). Besides, lenalidomide, an approved drug for treating multiple myeloma, MDS and mantle cell lymphoma (MCL), also has been clinically active in treatment of various lymphoproliferative disorders, including chronic lymphocytic leukemia (CLL) (Awan et al., 2010; González-Rodríguez et al., 2013; Itchaki & Brown, 2017). Moreover, the therapeutic effects of the drug thalidomide also have been evaluated in patients with CLL (Awan et al., 2010; Kay et al., 2009). In addition, several clinical studies showed that temozolomide demonstrates significant anti-leukemic activity in patients with relapsed and refractory acute leukemia (Gojo et al., 2016; Seiter et al., 2009, 2002). Besides, the usage of nilotinib for treating CML has been officially approved (National Cancer Institute (NCI), n.d.-a). Apart from that, the drug bexarotene has been tested in several clinical trials (NCT00316030, NCT00425477, NCT01001143) as a potential therapy for AML patients (Tsai et al., 2008). There are also several studies that

evaluate everolimus as a possible treatment for leukemia (NCT00093639, NCT01154439, NCT03740334).

Cardiovascular diseases (CVDs)

CVDs include diseases of the heart, vascular diseases of the brain, and diseases of the blood vessels ("Global Atlas on Cardiovascular Disease Prevention and Control", 2011). For example, heart attack and stroke are two common CVDs. Now, we discuss some of the predicted drug candidates listed in Table 4.1 in terms of their potential to provide clinical benefits to CVDs.

Citalopram has been suggested to be one of the first-choice antidepressant agents for patients with coronary artery disease who suffer from depressive symptoms (Lesp'erance et al., 2007; Yekehtaz et al., 2013). Besides, bupropion also may be considered as a safe antidepressant for patients with heart disease (Roose et al., 1991, 1987; Wenger & Stern, 1983). Imatinib which is originally indicated for treating leukemia, also has been actively studied for the treatment of pulmonary arterial hypertension (Farha et al., 2014; Frost et al., 2015; Ghofrani et al., 2010, 2005). Even though there is no clinical study in ClinicalTrials.gov which investigates the therapeutic effect of alprazolam for CVD, a recent study has shown that alprazolam is associated with a reduced risk of major adverse cardiovascular events and all-cause mortality among hypertensive individuals, as well as a reduced risk of hemorrhagic stroke in adults younger than 65 years old (Yeh et al., 2019). Besides, fenofibrate has been investigated as primary prevention of cardiovascular disease events (Jakob et al., 2016), and is related to reducing CVD risk in patients with metabolic syndrome (Kim et al., 2019) as well as in patients with type 2 diabetes (Elam et al., 2017). On the other hand, study shows that ursodeoxycholic acid (UDCA), which is widely prescribed in the treatment of cholestatic liver disease, can be potentially used as

standalone treatment for cardiac dysfunction (Vasavan et al., 2018). Another study also suggests that UDCA may help to improve peripheral blood flow in patients with chronic heart failure (von Haehling et al., 2012). Furthermore, the lipid-lowering effects of UDCA may also lower the risk of atherosclerotic CVD (Simental-Mendía et al., 2019). Apart from that, paroxetine, a drug usually used for treating depression and anxiety disorders, has the potential to provide therapeutic benefits on cardiac function in patients with acute myocardial infarction with depression (Tian et al., 2016). Besides, paroxetine is also recommended to be repositioned for diabetes-related CVD (Wheatcroft, 2013). Another drug, meloxicam, which is commonly used for treating Alzheimer's disease in its early stage, is also linked to a reduced risk of heart attacks and deaths from CVD (Nordström et al., 2013). Moreover, a study suggests that fluoxetine can be used in improving post-myocardial infarction depression (Strik et al., 2000). On top of that, the use of fluoxetine may reduce the cardiovascular-related morbidity and mortality rates of patients with depression who experience acute myocardial infarction (Taylor, 2005). Also, several studies propose a new role for thalidomide in treatment of heart failure in addition to traditional cardiovascular medications (Davey & Ashrafian, 2000; Gullestad et al., 2002, 2005).

CHAPTER 5: CONCLUSION

5.1 Summary

In this study, we constructed a weighted network of drugs based on their side-effect similarities. Then, based on the assumption that drugs with similar side-effects may share similar therapeutic roles, we indicated how potential drug candidates that may be repurposed for treating a target disease can be inferred from among the neighbors of the existing drugs approved for the disease. Since different drugs may have different sets of neighbors in the network, the question is then: among the existing drugs approved for a disease, whose neighbors are the most promising drug repositioning candidates for the disease? Will the centrality rankings of drugs provide some clues?

To investigate it, we selected diseases with the highest numbers of clinical trials in the ClinicalTrials.gov database as our target diseases. For each target disease, we used a centrality measure to identify the set of top $k\%$ of central drugs approved for the disease, and the set of bottom $k\%$ of least central drugs approved for the disease, for some parameter k. As a comparison, we also randomly chose a set of $k\%$ of drugs approved for the disease without using any centrality measure. The three sets of drugs were then used separately as a reference set for our prediction, in which we selected top-20 most potential drug repositioning candidates for the disease from among the neighbors of the drugs in the reference set. The prediction performances were then evaluated based on the percentage of the predicted candidates that have been investigated in relation to the disease in at least three clinical trials registered in ClinicalTrials.gov.

As a result, we observe that the predictions given by the top central drugs approved for our target diseases are more consistent with clinical interests, as compared to their peripheral and random counterparts. This suggests that similarity to the top central drugs

associated with a particular disease may be prioritized for a more accurate and effective screening of drug repositioning candidates for the disease. Our findings may serve as a basis that encourages more robust application of centrality measures in future drug repositioning efforts.

5.2 Limitation of this study

We acknowledge that the three centrality measures used in this study may not be the most suitable ones for analyzing the roles of drugs in the context of drug repositioning. We only used them as examples to showcase our investigation on how centrality measures may be of help in drug repositioning, in recognition of them as the classic centrality measures in the network science literature.

Besides, we only evaluated our predictions in this study based on how much they agree with the investigation performed on the clinical stage. It should be noted that a current lack of clinical trials for a drug-disease pair may not imply the drug's inability or irrelevance in treating a disease, as future clinical studies may provide further evidence in support of the drug-disease relation. On the other hand, not every clinical trial associated with a drug-disease pair implies that the drug has a positive effect on the disease, as some clinical trials may report failure of a drug-disease treatment relation. Although the positive predictive values reported in this study may not directly reflect the actual viability of the predicted drug repositioning candidates in treating a disease, they can be used as an indicator of the clinical relevance of our prediction, suggesting a possibility for more advanced research on the clinical potential of our predicted candidates.

Apart from that, evaluating similarities between drugs solely from the perspective of side effects may limit the validity of our inference in this study. Since we only focused on analyzing the drugs' centralities and their application to drug repositioning, we admit that

we may have overlooked other pharmacological factors that may impact the drugs' abilities to treat a disease. Hence, the analysis presented may be rather simplistic for a practical application of drug repositioning. Nevertheless, the unique insights portrayed in this work regarding the centralities of drugs are likely to motivate more in-depth studies on the use of centrality measures in guiding drug repositioning efforts in the future.

5.3 Suggestions for future work

Besides drugs' side effects, other aspects of drugs such as their chemical properties and protein targets may be integrated for evaluation of drug-drug similarities. This may improve the applicability of the drug repositioning model presented in the present study. Meanwhile, disease similarity also may be considered, since studies show that similar diseases tend to be treated by common therapeutic drugs (Cheng et al., 2016; Hu et al., 2017). This way, drug candidates for a new disease in which treatment is yet to be found may be deduced from the central drugs associated with its similar diseases.

On the other hand, our work may be expanded in the future by analyzing the use and relevance of other centrality measures in discovering drug repositioning opportunities. This will illuminate and provide more justification on the significance of central drugs. Meanwhile, our centrality-guided drug repositioning approach may be assessed more critically by examining the possible biases in the prediction towards or against certain groups of drugs.

While most drug repositioning strategies focus on finding alternative uses of approved drugs, they may serve as a paradigm for exploring new therapeutic potential for natural products or bioactive compounds (Alcaro & Ortuso, 2020; DeCorte, 2016; Maruca et al., 2020; Rastelli et al., 2020). Furthermore, from a broader perspective, drug repositioning may be regarded as a recommendation process. In our case, we recommend some drugs as

alternative treatment options for some diseases. Thus, the framework presented in this study may be applied more widely to the class of recommendation problems in network science. In return, we may also gain some valuable insights into the drug repositioning problem by considering the approaches and techniques proposed for developing recommendation systems.

REFERENCES

Alcaro, S., & Ortuso, F. (2020). Multi-target drug discovery: An opportunity for novel and repurposed bioactive compounds. *European Journal of Medicinal Chemistry*, *192*, Article#112188.

Ali, Y. E. M., Kwa, K. H., & Ratnavelu, K. (2017). Predicting new drug indications from network analysis. *International Journal of Modern Physics C*, *28*(09), Article#1750118.

American Cancer Society. (n.d.). *Leukemia.* Retrieved on 22 May 2020 from https://www.cancer.org/cancer/leukemia.html.

Assouline, S., Culjkovic, B., Cocolakis, E., Rousseau, C., Beslu, N., Amri, A., . . . Borden, K. L. B. (2009). Molecular targeting of the oncogene eIF4e in acute myeloid leukemia (AML): a proof-of-principle clinical trial with ribavirin. *Blood*, *114*(2), 257–260.

Assouline, S., Culjkovic-Kraljacic, B., Bergeron, J., Caplan, S., Cocolakis, E., Lambert, C., . . . Borden, K. L. B. (2014). A phase I trial of ribavirin and low-dose cytarabine for the treatment of relapsed and refractory acute myeloid leukemia with elevated eIF4e. *Haematologica*, *100*(1), e7–e9.

Awan, F. T., Johnson, A. J., Lapalombella, R., Hu, W., Lucas, M., Fischer, B., & Byrd, J. C. (2010). Thalidomide and lenalidomide as new therapeutics for the treatment of chronic lymphocytic leukemia. *Leukemia & Lymphoma*, *51*(1), 27–38.

Badkas, A., Landtsheer, S. D., & Sauter, T. (2020). Topological network measures for drug repositioning. *Briefings in Bioinformatics*, *00*(00), 1-13.

Bloch, F., & Jackson, M. O. (2016). Centrality measures in networks. *SSRN Electronic Journal*. Retrieved on 18 March 2021 from https://dx.doi.org/10.2139/ssrn.2749124.

Boldi, P., & Vigna, S. (2014). Axioms for centrality. *Internet Mathematics*, *10*(3-4), 222-262.

Borden, K. L. B., & Culjkovic-Kraljacic, B. (2010). Ribavirin as an anti-cancer therapy: acute myeloid leukemia and beyond? *Leukemia & Lymphoma*, *51*(10), 1805–1815.

Borgatti, S. P., & Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Social Networks*, *28*(4), 466–484.

Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L. J., & Bork, P. (2008). Drug target identification using side-effect similarity. *Science*, *321*(5886), 263–266.

Cheng, L., Jiang, Y., Wang, Z., Shi, H., Sun, J., Yang, H., . . . Zhou, M. (2016). DisSim: an online system for exploring significant similar diseases and exhibiting potential therapeutic drugs. *Scientific Reports*, *6*, Article#30024.

Council for International Organizations of Medical Sciences. (1999). *Guidelines for preparing core clinical-safety information on drugs : report of cioms working groups iii and v, including new proposals for investigator's brochures.* (2nd ed.). Geneva : Council for International Organizations of Medical Sciences, c1999.

Das, K., Samanta, S., & Pal, M. (2018). Study on centrality measures in social networks: a survey. *Social Network Analysis and Mining*, *8*(1), 13.

Davey, P., & Ashrafian, H. (2000). New therapies for heart failure: is thalidomide the answer? *QJM: An International Journal of Medicine*, *93*(5), 305–311.

DeCorte, B. L. (2016). Underexplored opportunities for natural products in drug discovery. *Journal of Medicinal Chemistry*, *59*(20), 9295–9304.

del Rio, G., Koschützki, D., & Coello, G. (2009). How to identify essential genes from molecular networks? *BMC Systems Biology*, *3*(1), 102.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1), 269–271.

DiMasi, J. A., Hansen, R. W., & Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *Journal of Health Economics*, *22*(2), 151 - 185.

Dudley, J. T., Deshpande, T., & Butte, A. J. (2011). Exploiting drug-disease relationships for computational drug repositioning. *Briefings in Bioinformatics*, *12*(4), 303–311.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: : Reasoning about a*

highly connected world. Cambridge: Cambridge University Press.

Elam, M. B., Ginsberg, H. N., Lovato, L. C., Corson, M., Largay, J., Leiter, L. A., . . . an, R. P. B. (2017). Association of fenofibrate therapy with long-term cardiovascular risk in statin-treated patients with type 2 diabetes. *JAMA Cardiology*, *2*(4), 370.

Farha, S., Dweik, R., Rahaghi, F., Benza, R., Hassoun, P., Frantz, R., . . . Asosingh, K. (2014). Imatinib in pulmonary arterial hypertension: C-kit inhibition. *Pulmonary Circulation*, *4*(3), 452–455.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Frost, A. E., Barst, R. J., Hoeper, M. M., Chang, H.-J., Frantz, R. P., Fukumoto, Y., . . . Ghofrani, H.-A. (2015). Long-term safety and efficacy of imatinib in pulmonary arterial hypertension. *The Journal of Heart and Lung Transplantation*, *34*(11), 1366–1375.

Ghofrani, H. A., Morrell, N. W., Hoeper, M. M., Olschewski, H., Peacock, A. J., Barst, R. J., . . . Pascoe, S. (2010). Imatinib in pulmonary arterial hypertension patients with inadequate response to established therapy. *American Journal of Respiratory and Critical Care Medicine*, *182*(9), 1171–1177.

Ghofrani, H. A., Seeger, W., & Grimminger, F. (2005). Imatinib for the treatment of pulmonary arterial hypertension. *New England Journal of Medicine*, *353*(13), 1412–1413.

Gillis, J., & Pavlidis, P. (2012). "Guilt by association" is the exception rather than the rule in gene networks. *PLoS Computational Biology*, *8*(3), Article#e1002444.

Global atlas on cardiovascular disease prevention and control. (2011). In S. Mendis, P. Puska, & B. Norrving (Eds.), (p. 3-18). World Health Organization.

Gojo, I., Beumer, J. H., Pratz, K. W., McDevitt, M. A., Baer, M. R., Blackford, A. L., . . . Karp, J. E. (2016). A phase 1 study of the PARP inhibitor veliparib in combination with temozolomide in acute myeloid leukemia. *Clinical Cancer Research*, *23*(3), 697–706.

González-Rodríguez, A. P., Payer, A. R., Acebes-Huerta, A., Huergo-Zapico, L., Villa-

Alvarez, M., Gonzalez-García, E., & Gonzalez, S. (2013). Lenalidomide and chronic lymphocytic leukemia. *BioMed Research International*, *2013*, Article#932010.

Gullestad, L., Semb, A. G., Holt, E., Skårdal, R., Ueland, T., Yndestad, A., . . . Aukrust, P. (2002). Effect of thalidomide in patients with chronic heart failure. *American Heart Journal*, *144*(5), 847–850.

Gullestad, L., Ueland, T., Fjeld, J. G., Holt, E., Gundersen, T., Breivik, K., . . . Aukrust, P. (2005). Effect of thalidomide on cardiac remodeling in chronic heart failure. *Circulation*, *112*(22), 3408–3414.

Hu, Y., Zhou, M., Shi, H., Ju, H., Jiang, Q., & Cheng, L. (2017). Measuring disease similarity and predicting disease-related ncRNAs by a novel method. *BMC Medical Genomics*, *10*(S5).

Itchaki, G., & Brown, J. R. (2017). Lenalidomide in the treatment of chronic lymphocytic leukemia. *Expert Opinion on Investigational Drugs*, *26*(5), 633–650.

Jakob, T., Nordmann, A. J., Schandelmaier, S., Ferreira-González, I., & Briel, M. (2016). Fibrates for primary prevention of cardiovascular disease events. *The Cochrane Database of Systematic Reviews*, *11*(11), Article#CD009753.

Jalili, M., Salehzadeh-Yazdi, A., Asgari, Y., Arab, S. S., Yaghmaie, M., Ghavamzadeh, A., & Alimoghaddam, K. (2015). CentiServer: A comprehensive resource, web-based application and R package for centrality analysis. *PloS ONE*, *10*(11), Article#e0143111.

Jalili, M., Salehzadeh-Yazdi, A., Gupta, S., Wolkenhauer, O., Yaghmaie, M., Resendis-Antonio, O., & Alimoghaddam, K. (2016). Evolution of centrality measurements for the detection of essential proteins in biological networks. *Frontiers in Physiology*, *7*, 375.

Jeong, H., Mason, S. P., Barabási, A.-L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41–42.

Kay, N. E., Shanafelt, T. D., Call, T. G., Wu, W., & Laplant, B. R. (2009). N9986: a phase II trial of thalidomide in patients with relapsed chronic lymphocytic leukemia. *Leukemia & Lymphoma*, *50*(4), 588–592.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1-2), 81–93.

Keng, Y. Y., Kwa, K. H., & McClain, C. (2020). Convex combinations of centrality measures. *The Journal of Mathematical Sociology*, 1–28.

Kim, N. H., Han, K. H., Choi, J., Lee, J., & Kim, S. G. (2019). Use of fenofibrate on cardiovascular outcomes in statin users with metabolic syndrome: propensity matched cohort study. *BMJ (Clinical research ed.)*, *366*, Article#l5125.

Koschützki, D., & Schreiber, F. (2008). Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regulation and Systems Biology*, *2*, 193-201.

Kuhn, M., Campillos, M., Letunic, I., Jensen, L. J., & Bork, P. (2010). A side effect resource to capture phenotypic effects of drugs. *Molecular Systems Biology*, *6*(1), 343.

Kuhn, M., Letunic, I., Jensen, L. J., & Bork, P. (2016). The SIDER database of drugs and side effects. *Nucleic Acids Research*, *44*(D1), D1075–D1079.

Lesp'erance, F., Frasure-Smith, N., Koszycki, D., Lalibert'e, M.-A., van Zyl, L. T., Baker, B., . . . for the CREATE Investigators (2007). Effects of citalopram and interpersonal psychotherapy on depression in patients with coronary artery disease. *JAMA*, *297*(4), 367-379.

Lewis, T. G. (2009). *Network science: Theory and applications*. Hoboken, N.J: John Wiley & Sons.

Lü, L., Chen, D., Ren, X.-L., Zhang, Q.-M., Zhang, Y.-C., & Zhou, T. (2016). Vital nodes identification in complex networks. *Physics Reports*, *650*, 1–63.

Maruca, A., Rocca, R., Catalano, R., Mesiti, F., Costa, G., Lanzillotta, D., . . . Artese, A. (2020). Natural products extracted from fungal species as new potential anti-cancer drugs: A structure-based drug repurposing approach targeting HDAC7. *Molecules*, *25*(23), Article#5524.

Momparler, R. L., Bouchard, J., Onetto, N., & Rivard, G. E. (1984). 5-AZA-2'-deoxycytidine therapy in patients with acute leukemia inhibits DNA methylation. *Leukemia Research*, *8*(2), 181–185.

Mund, C., Hackanson, B., Stresemann, C., Lübbert, M., & Lyko, F. (2005). Characterization of DNA demethylation effects induced by 5-aza-2'-deoxycytidine in patients with myelodysplastic syndrome. *Cancer Research*, *65*(16), 7086–7090.

National Cancer Institute (NCI). (n.d.-a). *Nilotinib.* Retrieved on 22 May 2020 from https://www.cancer.gov/about-cancer/treatment/drugs/nilotinib.

National Cancer Institute (NCI). (n.d.-b). *Oxaliplatin.* Retrieved on 22 May 2020 from https://www.cancer.gov/about-cancer/treatment/drugs/oxaliplatin.

Newman, M. E. J. (2010). *Networks : an introduction*. Oxford New York: Oxford University Press.

Nordström, P., Religa, D., Wimo, A., Winblad, B., & Eriksdotter, M. (2013). The use of cholinesterase inhibitors and the risk of myocardial infarction and death: a nationwide cohort study in subjects with alzheimer's disease. *European Heart Journal*, *34*(33), 2585–2591.

Oldham, S., Fulcher, B., Parkes, L., Arnatkeviciute, A., Suo, C., & Fornito, A. (2019). Consistency and differences between centrality measures across distinct classes of networks. *PLoS ONE*, *14*(7), Article#e0220061.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251.

Ozgur, A., Vu, T., Erkan, G., & Radev, D. R. (2008). Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics*, *24*(13), i277–i285.

Pammolli, F., Magazzini, L., & Riccaboni, M. (2011). The productivity crisis in pharmaceutical R&D. *Nature Reviews Drug Discovery*, *10*(6), 428–438.

Rastelli, G., Pellati, F., Pinzi, L., & Gamberini, M. C. (2020). Repositioning natural products in drug discovery. *Molecules*, *25*(5), Article#1154.

Rohon, P., Vondrakova, J., Jonasova, A., Holzerova, M., Jarosova, M., & Indrak, K. (2012). Treatment of chronic myelomonocytic leukemia with 5-azacytidine: Case reports. *Case Reports in Hematology*, *2012*, 1–4.

Roose, S. P., Dalack, G. W., Glassman, A. H., Woodring, S., Walsh, B. T., & Giardina, E. G. (1991). Cardiovascular effects of bupropion in depressed patients with heart disease. *The American Journal of Psychiatry*, *148*, 512-516.

Roose, S. P., Glassman, A. H., Giardina, E. G., Johnson, L. L., Walsh, B. T., & Bigger, J. T. (1987). Cardiovascular effects of imipramine and bupropion in depressed patients with congestive heart failure. *Journal of Clinical Psychopharmacology*, *7*, 247–251.

Savona, M. R., Gore, S. D., Kolibaba, K. S., Ukrainskyj, S. M., Kumar, K., Dong, Q., . . . Garcia-Manero, G. (2015). CC-486 (oral azacitidine) monotherapy in patients with acute myeloid leukemia (AML). *Blood*, *126*(23), 452–452.

Seiter, K., Katragadda, S., Ponce, D., Rasul, M., & Ahmed, N. (2009). Temozolomide and cisplatin in relapsed/refractory acute leukemia. *Journal of Hematology & Oncology*, *2*(1), 21.

Seiter, K., Liu, D., Loughran, T., Siddiqui, A., Baskind, P., & Ahmed, T. (2002). Phase i study of temozolomide in relapsed/refractory acute leukemia. *Journal of Clinical Oncology*, *20*(15), 3249–3253.

Shahreza, M. L., Ghadiri, N., Mousavi, S. R., Varshosaz, J., & Green, J. R. (2017). A review of network-based approaches to drug repositioning. *Briefings in Bioinformatics*, *19*(5), 878–892.

Shi, F., Len, Y., Gong, Y., Shi, R., Yang, X., Naren, D., & Yan, T. (2015). Ribavirin inhibits the activity of mTOR/eIF4E, ERK/Mnk1/eIF4E signaling pathway and synergizes with tyrosine kinase inhibitor imatinib to impair Bcr-Abl mediated proliferation and apoptosis in Ph+ leukemia. *PLoS ONE*, *10*(8), Article#e0136746.

Sidwell, R. W., Huffman, J. H., G. P. Khare, L., Allen, B., J. T. Witkowski, R., & Robins, K. (1972). Broad-spectrum antiviral activity of virazole: 1-f8- d-ribofuranosyl-1,2,4-triazole- 3-carboxamide. *Science*, *177*(4050), 705–706.

Simental-Mendía, L. E., Simental-Mendía, M., Sánchez-García, A., Banach, M., Serban, M.-C., Cicero, A. F. G., & Sahebkar, A. (2019). Impact of ursodeoxycholic acid on circulating lipid concentrations: a systematic review and meta-analysis of randomized placebo-controlled trials. *Lipids in Health and Disease*, *18*(1), 88.

Strik, J. J. M. H., Honig, A., Lousberg, R., Lousberg, A. H. P., Cheriex, E. C., Tuynman-

Qua, H. G., ... Praag, H. M. V. (2000). Efficacy and safety of fluoxetine in the treatment of patients with major depression after first myocardial infarction: Findings from a double-blind, placebo-controlled trial. *Psychosomatic Medicine*, *62*(6), 783–789.

Taylor, C. B. (2005). Effects of antidepressant medication on morbidity and mortality in depressed patients after myocardial infarction. *Archives of General Psychiatry*, *62*(7), 792.

Tian, X., Wang, Q., Guo, R., Xu, L., Chen, Q., & Hou, Y. (2016). Effects of paroxetine-mediated inhibition of GRK2 expression on depression and cardiovascular function in patients with myocardial infarction. *Neuropsychiatric Disease and Treatment*, *Volume 12*, 2333–2341.

Tsai, D. E., Luger, S. M., Andreadis, C., Vogl, D. T., Kemner, A., Potuzak, M., ... Carroll, M. (2008). A phase i study of bexarotene, a retinoic x receptor agonist, in non-m3 acute myeloid leukemia. *Clinical Cancer Research*, *14*(17), 5619–5625.

Tsimberidou, A. M., Keating, M. J., Jabbour, E. J., Ravandi-Kashani, F., O'Brien, S., Estey, E., ... Borthakur, G. (2014). A phase i study of fludarabine, cytarabine, and oxaliplatin therapy in patients with relapsed or refractory acute myeloid leukemia. *Clinical Lymphoma Myeloma and Leukemia*, *14*(5), 395–400.e1.

Tsimberidou, A. M., Wierda, W. G., Plunkett, W., Kurzrock, R., O'Brien, S., Wen, S., ... Keating, M. J. (2008). Phase I-II study of oxaliplatin, fludarabine, cytarabine, and rituximab combination therapy in patients with richter's syndrome or fludarabine-refractory chronic lymphocytic leukemia. *Journal of Clinical Oncology*, *26*(2), 196–203.

Vasavan, T., Ferraro, E., Ibrahim, E., Dixon, P., Gorelik, J., & Williamson, C. (2018). Heart and bile acids - clinical consequences of altered bile acid metabolism. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1864*(4), 1345–1355.

von Haehling, S., Schefold, J. C., Jankowska, E. A., Springer, J., Vazir, A., Kalra, P. R., ... Anker, S. D. (2012). Ursodeoxycholic acid in patients with chronic heart failure. *Journal of the American College of Cardiology*, *59*(6), 585–592.

Wang, F., Zhang, P., Cao, N., Hu, J., & Sorrentino, R. (2014). Exploring the associations between drug side-effects and therapeutic indications. *Journal of Biomedical Informatics*, *51*, 15–23.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Wenger, T. L., & Stern, W. C. (1983). The cardiovascular profile of bupropion. *The Journal of Clinical Psychiatry*, *44*, 176–182.

Wheatcroft, S. B. (2013). Teaching an old drug new tricks: Can paroxetine ease the burden of cardiovascular disease in diabetes? *Diabetes*, *62*(3), 698–700.

Wohnsland, A., Hofmann, W. P., & Sarrazin, C. (2007). Viral determinants of resistance to treatment in patients with hepatitis c. *Clinical Microbiology Reviews*, *20*(1), 23–38.

Xue, H., Li, J., Xie, H., & Wang, Y. (2018). Review of drug repositioning approaches and resources. *International Journal of Biological Sciences*, *14*(10), 1232–1244.

Yang, L., & Agarwal, P. (2011). Systematic drug repositioning based on clinical side-effects. *PLoS ONE*, *6*(12), Article#e28025.

Ye, H., Liu, Q., & Wei, J. (2014). Construction of drug network based on side effects and its application for drug repositioning. *PLoS ONE*, *9*(2), Article#e87864.

Yeh, C.-B., Tsai, M.-C., Teng, Y.-H., Ku, M.-S., Huang, J.-Y., Wang, B.-Y., . . . Ho, S.-W. (2019). Association of alprazolam with major cardiovascular events in patients with hypertension. *Journal of Evaluation in Clinical Practice*, *26*(3), 983–991.

Yekehtaz, H., Farokhnia, M., & Akhondzadeh, S. (2013). Cardiovascular considerations in antidepressant therapy: an evidence-based review. *The Journal of Tehran Heart Center*, *8*, 169–176.

Zhang, P., Agarwal, P., & Obradovic, Z. (2013). Computational drug repositioning by ranking and integrating multiple data sources. In H. Blockeel, K. Kersting, S. Nijssen, & F. Železný (Eds.), *Machine learning and knowledge discovery in databases* (pp. 579–594). Berlin, Heidelberg: Springer Berlin Heidelberg.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

List of Publications

Keng, Y. Y., Kwa, K. H., & Ratnavelu, K. (2021). Centrality analysis in a drug network and its application to drug repositioning. *Applied Mathematics and Computation*, *395*, Article#125870.

Keng, Y. Y., Kwa, K. H., & McClain, C. (2020). Convex combinations of centrality measures. *The Journal of Mathematical Sociology*, 1–28.