Briefings in Bioinformatics, 22(4), 2021, 1–13

https://doi.org/10.1093/bib/bbaa357
Method Review

Topological network measures for drug repositioning

Apurva Badkas, Sébastien De Landtsheer and Thomas Sauter

Corresponding author: Thomas Sauter, Department of Life Sciences and Medicine, University of Luxembourg, Belval Campus, 2, avenue de l'Université L-4365 Esch-sur-Alzette, Luxembourg. Tel.: (+352) 46 66 44 6296; E-mail: Thomas.Sauter@uni.lu

Abstract

Drug repositioning has received increased attention since the past decade as several blockbuster drugs have come out of repositioning. Computational approaches are significantly contributing to these efforts, of which, network-based methods play a key role. Various structural (topological) network measures have thereby contributed to uncovering unintuitive functional relationships and repositioning candidates in drug-disease and other networks. This review gives a broad overview of the topic, and offers perspectives on the application of topological measures for network analysis. It also discusses unexplored measures, and draws attention to a wider scope of application efforts, especially in drug repositioning.

Key words: drug repositioning; networks; topological network measures; topology; computational methods

Introduction

The recent COVID-19 pandemic is a convincing case for exploring and expanding drug repositioning studies. Given the rapid and massive scale spread of the pandemic, time was of essence, and this pandemic saw concerted efforts to reposition anti-viral, anti-malarial drugs to treat patients—take the example of Remdesivir. One of the drugs included in the standard treatment regimen of COVID-19 currently, it was developed as a broad-spectrum anti-viral, which was repositioned for COVID [1]. In the face of the world-wide spread and the extent of population affected, the need for reliable treatment options in a short time brought into focus the criticality of drug repositioning studies on a larger scale.

Drug repositioning (or drug repurposing) is the use of drugs approved for one condition in treating another condition that may be related or unrelated to the original. History is peppered with examples: discovery of DNA alkylating activity of nitrogen mustard (used for warfare during World War II [2]), Sildenafil citrate (Viagra, originally indicated for hypertension, repositioned for treating erectile dysfunction [3, 4]) and Tamoxifen (originally designed as a contraceptive pill, now one of the leading drugs for breast cancer [5]), are some of the better known examples. In contrast with new chemical entities (NCEs), pharmaceuticals already approved for one condition do not have to go through extensive assessment of preclinical behavior, pharmacokinetics, pharmacodynamics and toxicological profiling before they are approved for a new indication. Availability of preliminary data for drug repositioning candidates decreases the time required for bringing them into clinical use for other conditions, reduces the risk of failure in clinical studies and offers a vast cost benefit to the health care system.

Indeed, time and cost have a direct impact on both the producers and consumers, as a major challenge in pharmaceutical industry is the large cost and extensive period required for the development process. From initial lab scale experiments through the three phases of clinical trials and final approval for patient-use, it takes 10–15 years, and about 2–3 billion dollars [6]. Approval rate for drugs entering clinical trials is about 10% [7, 8]. Another concern is the high attrition rate in clinical trials.

Apurva Badkas is pursuing her doctoral studies in the Systems Biology Group at the University of Luxembourg. Her project focuses on drug sensitivity prediction using topological network measures.
Sébastien De Landtsheer obtained his PhD in Systems Biology at the University of Luxembourg and is a postdoctoral researcher in Prof. Thomas Sauter group. His research interests include network simulations and machine-learning.
Thomas Sauter is professor of Systems Biology and Study Director at the University of Luxembourg. His research is focused in the area of molecular systems biology, particularly in Data Integration and Analysis of Cancer specific Networks.
Submitted: 12 August 2020; **Received (in revised form):** 3 November 2020

Drugs already approved for other indications would be less likely to be rejected in Phase II and Phase III due to toxicity issues, unless very high doses are required for the new indication; the majority of drug failures in Phase III come from both, lack of efficacy and safety issues (adverse events, toxicity issues that were missed in animal models) [9–11]. Re-positioned drugs have been contributing to a significant percentage of drug approvals in recent times. It has been reported that, in comparison to new entities, 30% of the repositioned candidates are approved [12, 13]. It is likely that repositioned drugs require a smaller sample size for clinical trials as compared to NCEs. The size of the market for repositioned drugs in 2020 was projected to be 31.3 billion USD [12].

Over the past decade, various complex conditions, especially cancer, have seen rise in targeted therapies. Cancer is characterized by a very high level of heterogeneity in its cause, prognosis and patient response profiles to therapies. Given the diverse landscape of genetic variations in cancer, developing specific drugs for mutations occurring in low frequencies may be cost prohibitive. Even as newer therapies are being introduced, their application may be limited over time due to resistance setting in. Cost and time for development of targeted therapies often out-weighs the time of its efficacy due to rapid resistance development. Drug repositioning could lead to cost savings in pharmaceutical industry and improved therapeutic choices for patients. Some of the current chemotherapeutic drugs provide minimal benefit in prolongation of survival, with a severely reduced quality of life. Patients relying on such drugs are also likely to benefit from alternatives.

400 million people worldwide are afflicted by conditions known as orphan diseases [14]—rare conditions, occurring in a very small percentage of population—and hence, very few studies can be undertaken for such conditions. Drugs specifically designed for these orphan diseases are usually unavailable, and hence there is a need for drug repositioning efforts to treat such diseases.

There are some other notable factors such as commercial interests and intellectual property rights. Several insightful commentaries are available on the subject [15–18].

Given the enormous quantity and varied types of data generated, computational approaches have been designed and indeed, have become the first set of tools being used for identifying potential drug candidates for conditions of interest, and vice-versa. A variety of approaches of drug repositioning have been adopted, and several excellent reviews have covered their development [6, 8, 13, 19–25]. Several conditions such as Alzheimer's, Parkinson's and other neurological conditions, various viral infections, conditions such as epilepsy, and cancer have been explored in computational drug repositioning studies.

The networks formed by connections and interactions among biological entities are being studied extensively for many applications, including drug repositioning studies. Network medicine leverages the fact that biological entities form interaction networks, and perturbation in one part of such a network creates systemic effects. Networks are a useful platform not only for visualizing large scale, heterogeneous data on one platform, but also for obtaining mechanistic insights and predictive properties, using graph theory for analyzing their structures and behaviors. Graph theory and other mathematical methods of network analysis have been applied in diverse areas such as sociology, ecology, linguistics, communication, logistics and biology. The rise in ubiquity of the use of social media has led to incredible quantities of data, and graph methods are being specifically developed and applied to analyses of these platforms such as Facebook, Twitter, YouTube, Instagram, etc. for extracting and predicting-from marketing and voting preferences to likely spread of infectious diseases. In biology, networks have seen applications in studying spread of epidemics, predicting disease genes, gaining mechanistic insights underlying various diseases, etc. [26]. Interest for complex networks analysis and application of different network measures also emerged from the need to uncover brain structural and functional properties, trying to understand the structure–function relationship [27]. Network approaches are being integrated into computational pipelines for drug discovery and drug repositioning to highlight previously unknown connections among the interacting entities. In their excellent review on the subject, Csermely *et al*. [28] provide a comprehensive rationale, and in-depth discussion on the use of networks in drug discovery and other drug-related topics, such as drug repositioning and the study of side-effects. The purpose of this review at hand is to supplement the existing literature with an overview of some of the applications of networks to drug repositioning, and in particular, to discuss different network measures applied to drug repositioning.

This work, therefore, gives a brief overview of the network measures that have been adopted in drug repositioning studies. We will see how some of the measures have been frequently applied in such studies, and that there is an argument for broadening the set of measures used to analyze networks and gain additional insights. We will also discuss some context to be considered while leveraging networks for drug repositioning.

Overview: networks in brief, and networks for drug repositioning

The two components of any graph are nodes and edges. Nodes are the interacting entities. Edges are the interactions denoting a relationship between two nodes. A network could be made of a single data-type (and hence a single type of node), such as all the interactions in a protein–protein interaction (PPI) network. It could be bi-partite—drug-disease, drug-target—and also multi-partite, with three or more types of nodes. Networks are either directed (gene-regulatory networks consisting of transcription factors and the genes they act upon, or signaling networks, where signals have a defined direction of flow) or undirected (PPI networks). They can also be weighted (edges have different strengths of interactions) or unweighted (all edges have equal strengths).

For drug repositioning, various types of networks can yield insights. A drug–drug network with edges indicating shared targets can suggest candidates for repositioning. A disease–disease network can help indicate similarities and differences between different conditions, and hence suggest candidates based on disease similarity. However, some of the successfully repositioned candidates have been found in case of very diverse conditions, hence this approach has limitations. Another limitation of this approach is that this approach cannot give insights into required dosing for drugs. While the drug candidate being repositioned may be common, diverse conditions may have condition-specific quantitative and dynamic dosing requirements. A drug–drug similarity network based on drug structure can also offer insights into structural analogs, and perhaps be a starting point for further investigations. Different networks typically constructed for drug repositioning studies are chemical compound networks, chemical reaction networks, protein-structure networks, PPI networks, drug–drug interaction networks, drug-target networks, disease–drug networks, etc.

Another layer of complexity networks can present are entities with different edge types. While a simple graph has only one edge connecting two nodes, edges of multiple types can exist, leading to multi-edged graphs. For example, for a PPI network, edges may denote physical interactions, but can also be used to denote co-expression networks, or protein complexes and their interactions [29]. Most methods that were used so far had been for graphs homogenous edge types, however, recent efforts have been focused on analyzing multi-edged graphs, or 'multi-graphs.' For example, a multi-edged graph arose from imaging and cognitive data in a study of autism and schizophrenia, on which unsupervised clustering was performed to obtain subpopulations [30]. Extraction of data from biomedical literature yields multi-edged graphs or 'Knowledge graphs' [31, 32]. Increasing attention to these multigraphs will bring in better insights into the interplay between various interacting layers we encounter in biology.

Drug repositioning efforts have often been serendipitous, but markedly intentional in the recent times, driven by establishment of several public databases catering to providing the kind of qualitative and quantitative data required to drive these efforts. Jin and Wang [13] have classified these methods as blinded (or screening methods) where candidates are identified by-chance or by their effect on phenotypes; Target-based screening for entities binding to a target of interest used to screen for candidates; Knowledge-based methods; Signature-based methods; Pathway/network-based methods, Targeted mechanism-based methods. The starting point of drug repositioning is a known feature. For example, if the drug structure is known, other drugs with similar structure can be explored as repositioning candidates. Or, in many cases, its mechanism of action is known. A detailed understanding of mechanism of action and targets of known drugs can help identify other targets near the pathways involved. Thus, in an 'on-target' repositioning application, the same mechanism of action is involved in selecting other targets. One example is Digitoxin, which was prescribed for treatment of heart failure and arrhythmia. Its mechanism of action is blocking Na^+/K^+-ATPase pumps. This blocking mechanism is also proposed to be effective in inducing apoptosis, thus suggesting a role of digitoxin in cancer treatment [33]. An alternative mechanism of action of a known drug may also lead to repositioning opportunities, in what is known as 'off-target repositioning.' An old drug may be used for a different condition based on the new mechanism. Thalidomide is such an example of off-target repositioning [33]. Originally marketed as a sedative and a drug to treat morning sickness in pregnant women due to its observed anti-emetic effects, it was later withdrawn because of its teratogenic effects causing severe birth defects. It is now being used, because of its anti-angiogenic, anti-inflammatory and anti-myeloma activities, to treat leprosy and myeloma [34]. Several databases are now recording gene expression signatures of known/approved drugs to document the extent of drug action. Some methods look for drugs that show gene expression signature reversals of those in diseased states, so as to turn the disease phenotype into a healthy one [35]. Literature mining to build relationship graphs [36], deep learning using gene expression profiles on perturbagen treatments [37], using structural similarity between proteins, drugs, molecular docking techniques [25], pathway-based target identification [38], are among the many approaches seen. The variety also comes from data used to build networks (chemical structures, genetic data, proteomic data, transcriptomic profiles, epigenetic and phenotypic data, metabolic data, etc.). Permutations of methods and databases have led to a wide range of literature on the topic [13, 24, 39–42]. Topological measures have been used as features for the analysis of networks for drug repositioning research.

Before beginning a detailed discussion on various network metrics, a brief discussion on evolving graph transformation and analysis techniques is warranted. In the recent years, graph and node embedding techniques are being explored for handling large scale, complex and heterogenous networks. These methods are interested in dimensional reduction of these large networks, while aiming to capture both node information and topology of the networks, mapping them to lower dimensional spaces. Integration of the transcriptomic signatures across many genes and experiments poses a dimensionality problem, which can be addressed by finding low-dimensional latent representations of the compounds measured effects, with the assumption that these embeddings display a useful structure that is biologically meaningful. They can be computed in various ways. For example, Iorio *et al.* [43, 44] extract embeddings from the transcriptomic signatures of drugs across multiple experiments by prioritizing the most often perturbated genes. In this case, the embeddings are data-driven, weighted Prototype Ranked Lists which have been used to build searchable databases for drug repositioning [45]. Another approach is to use dimensionality reduction techniques to learn a low-dimension latent vector for each node. Various methods exist to find such embeddings, notably neural networks [46–48], diffusion component analysis [49] or simply matrix factorization [50]. This kind of method has been used to classify chemotherapeutic compounds with respect to their mechanism of action [51].

Increasingly, graph convolutions [52], a type of embedding taking full advantage of the topological structure of multi-modal graphs, have been used with different types of data sources in an effort to repurpose drugs. By using a large knowledge graph of drugs, diseases and biological entities, Sosa *et al.* [53] leverage complex semantic relationships between compounds to generate novel repositioning hypotheses. Similarly, Zong *et al.* used DeepWalk [54] to predict novel drug-target associations, based on a multi-modal network of drugs, diseases and target genes [55]. The same embedding method has also been used to infer disease-related miRNAs [56]. Graph convolutional neural networks (GCNs) have also shown interesting results for the problems of prediction of the side-effects of drug combinations [57, 58] but also novel PPIs [59], drug-target interactions [60] or gene function [61]. Given the pace of development of deep learning and graph embedding techniques in the fields of artificial intelligence, it is likely that GCNs will be instrumental to drug discovery and repositioning in the future.

As mentioned previously, graph theory provides the theoretical foundation to analyze topological properties of networks. The topology of a network refers to the placement or layout of the nodes and their connecting edges [62]. Numerous studies across various disciplines have led to the development of over 200 measures characterizing networks, in particular, highlighting important nodes, identifying sub-graphs (modules), and uncovering key channels of network communication. Biological networks encode information in their geometry, and several biologically relevant inferences have been made based on topological network features. Some of the frequent measures seen in literature are illustrated in Figure 1. Degree of a protein in a protein–protein network is generally an indication of its essentiality [63]. Various measures such as betweenness, closeness, load, degree centralities, connectivity, etc. were used in a study to stratify diseases [64]. Four-node motifs have been seen to occur frequently in metabolic networks [65]. It was shown that four-node motifs occurred in regions where nodes have high

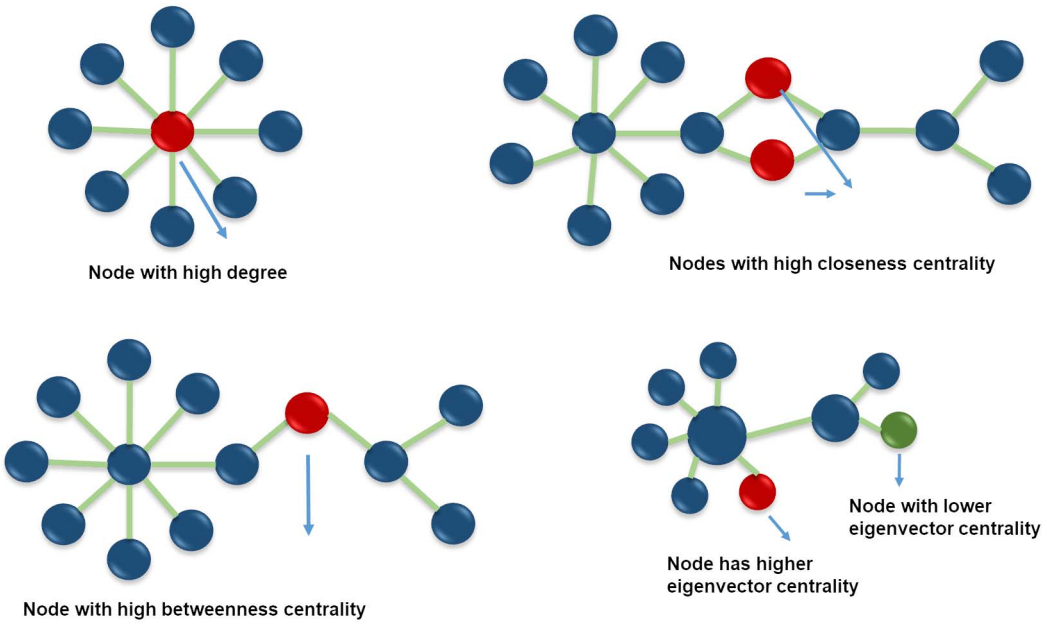

Figure 1. Some of the commonly used measures: degree (top, left)—describes the number of connections of a node, closeness centrality (top, right) highlights nodes in the network closest to other nodes (and can be easily reached), betweenness centrality (bottom, left)—nodes which channel communication in the network, and eigenvector centrality (bottom, right)—which is based on connections with highly connected neighbors.

betweenness centralities [65]. Tissue specific proteins have been shown to have high betweenness centrality [66]. Zhang *et al.* [67] analyzed the network of orphan diseases, and observed that proteins mutated in orphan diseases tend to have higher degree and betweenness centrality, but are not hub proteins in the overall human PPI network. Based on an analysis of human proteins targeted in viral infections, several hubs and bridging proteins have been found to be targeted, and it has been suggested that bridging proteins would make better targets, from the side-effect point of view [68]. Mora *et al.* [69] commented that using high-centrality targets for drugs may lead to more drug withdrawals due to side-effects. Da Silva Lopes *et al.* [70] applied four network measures (degree, betweenness, closeness and Burt's constraint) to analyze characteristics of drug targets of failed versus approved drugs. They report that targets of failed drugs have higher degree, and lower values of closeness and Burt's constraint, but no significant difference in betweenness. Their analysis, which was based on protein-structure similarity network and PPI network, also indicates that approved drugs have specific structural domains that are less frequently seen in others, while the targets of failed drugs contain domains that are ubiquitous. The metrics degree and betweenness centrality were also used to examine the addictive drugs network with their targets [71], and proposes candidates for repositioning to treat addictions. A few of the commonly found metrics are described in Table 1.

Some examples of the application of network measures for repositioning are described below in more detail.

How are network/centrality measures applied for drug repositioning

In some of the approaches seen in literature, network measures-based analysis is the key component of the method. In Udrescu *et al.* [76], the authors built a drug–drug interaction network, where the weighted edges denoted a link between two drugs if they shared both a target and mode of action (agonist/antagonist). The drug–drug similarity networks formed were subjected to community clustering, and betweenness to degree ratio was determined. They proposed that high degree nodes may have saturated their potential for repositioning, while the high betweenness but low degree nodes may show potential for repositioning. This study provides validation, in the way of molecular docking studies, of their predicted repositioning candidates, and is an interesting integration of two computational areas of networks and molecular docking simulations. The group had previously used other similarity measure [77]—applying average path length, clustering coefficient, degree distribution, network density, modularity, closeness, PageRank and eigenvector centrality—along with community clustering. They used topology and modularity to identify nine clusters/communities of drug classes. In this work, an overlap between anti-epileptic drugs and immune system related drugs is indicated, thus, providing an example of drug repositioning opportunities for the two indications.

A network proximity measure has been defined based on proximity between a drug target and disease protein (average shortest path and average closest path) [78]. Based on network proximity, the authors made several observations: effective drugs to be those targets are close to disease genes, and there were some drugs for which the effect could not be explained by proximity, possibly because of incomplete network structure. They also explained the link between two conditions, type 2 diabetes and Alzheimer's disease, based on a common pathway between two drugs indicated for each condition. This approach was used to suggest associations between several approved drugs and diseases, validating two of them via patient records and *in vitro* experiments [79]. A similar approach was used by Guney [80], to analyze side-effect modules, and for predictions of side-effects not present in the database SIDER. The PageRank score and an algorithm, Netscore, were also used to analyze the side-effect modules. Such systematic observations on side-effects are important resources for drug repositioning efforts.

Table 1. Some of the most commonly found network measures in literature applied to biological networks

Measure	Explanation	Reference
Degree	Number of connections of a node in a network.	Boccaletti *et al.* [72]
Betweenness centrality	For a given node in the network, the fraction of shortest paths through the node.	Newman [73]
Closeness centrality	For a node in the network, inverse of the average distance from all other nodes.	Boccaletti *et al.* [72]
Eigenvector centrality	Eigenvector centrality of a node is its weight in the network, based on its connections to other important nodes (i.e. their connectivity in the network).	Lohmann *et al.* [74]
PageRank	PageRank of a node is determined by the number and quality of its connections. Variant of eigenvector centrality, applicable to directed networks.	Liu *et al.* [75]
Clustering coefficient	Also known as transitivity, refers to proportion of connected triangles in the network.	Boccaletti *et al.* [72]

Zhou *et al.* [81] used the same approach to investigate host–virus interactome for drug repositioning suggestions for COVID-19. They used 119 known human interactors of corona viruses (from four previous outbreaks), to identify 47 human proteins (GSK3B, DPP4, SMAD3, PARP1 and IKBKB, among others), which could be targets of an approved or experimental drug. The application of this method to a variety of different networks and datasets is an asset, especially given the heterogeneity inherent in different types of network building strategies adopted in various studies. However, since calculation of *z*-score assumes normal distribution of data, a brief discussion on the applicability of the method to the underlying data vis-à-vis validity of the assumptions for the method would be useful to the readers.

Cancer has been a key focus area of drug repositioning studies. For drug repositioning for triple negative breast cancer (TNBC), Vitali *et al.* [82] constructed a PPI network around disease proteins, and used bridging centrality to highlight target proteins since the expectation is that these bridging nodes are structurally important nodes with low degrees and thus would make better targets. Hubs were discarded as they are genes with higher connectivity, and thus can introduce more side-effects. Drugs against these targets were obtained from CTD and DrugBank. Boolean models were built using disease specific pathways, and were simulated to test the effect of selected drugs, using a Monte-Carlo approach to assign missing values for nodes in the Boolean network. This work suggests repositioning several drugs such as Imatinib for TNBC. While the integration of network analysis and Boolean methods looks to be a promising approach, one additional step may perhaps strengthen this particular study. The study looks into genes that show changes in expression and take the causal drugs into consideration as interesting candidates. For cancer studies, the effect of the drug on the growth of the cell will still be an additional assessment criterion, which is not inherent in the computational pipeline.

Network measures are also used in combination with other approaches. Jadamba and Shin [83] used degree centrality, betweenness centrality and PageRank to analyze the topological features of a drug–gene network for breast cancer. They reason that using only gene expression profiles of drug perturbations may include not just drug activity, but also off-target, non-specific perturbations. Hence, their approach was to build a network using disease specific pathways based on gene expression profiles, and drugs associated with these pathways as determined by drug-perturbed gene expression profiles in the Connectivity Map. They then used a semi-supervised algorithm with network propagation for repositioning unmapped drugs to predict 17 candidates, out of which 10 were already reported to be in use. To computationally validate the seven new candidates, they constructed a network consisting of drug–drug, drug-target gene and gene–gene interactions. Based on the highest degree centrality and PageRank scores, they validated three high scoring drugs as the strongest candidates for repositioning. Using both network-based and literature-based validation, four of the new candidates were recommended for further investigation. One of the assumptions in this study is that drugs approved for the same disease target most of the same pathways (and hence molecular targets). This is a strong assumption, and could use more justification.

Analysis of a cancer co-expression network (for seven types of cancers) with degree, betweenness and closeness centrality was undertaken by Bourdako and Spyrou [84] for comparison to a method called 'Informed walks' developed by them. They used their algorithm to identify mechanism specific sub-networks, identify driver genes for the sub-network and employed a drug-repositioning pipeline to suggest potential drugs for the identified drivers. They subsequently showed that their informed walk approach identified a higher number of significant genes as compared to the three centrality measures. Identifying specific sub-networks for each cancer has important applications from the perspective of which of them show specific markers, and thus, are likely to respond to treatments. Community classification also offers opportunities to study some of the less well-known genes in such clusters, and derive better insights into cancer-specific signatures.

An example of centrality measures used in the context of cardiovascular diseases is Huang *et al*'s. [85] study to analyze a gene association network of vascular smooth muscle cells (VSMC). They used a set of topological measures (closeness centrality, degree centrality, eccentricity centrality, betweenness centrality, bridging centrality, clustering coefficient, brokering coefficient, local average connectivity). In an effort to investigate repositioning of cancer drugs to cardiovascular diseases caused by the effect of stress on VSMCs, these topological features were used to help identify a set of differentially expressed genes (DEGs) based on the time course study on VSMCs. Interestingly, in this study, they also used statistical tests, and cluster analysis, and identified a set of DEGs using all three approaches to obtain a combined final set of DEGs. This study is an excellent example of the complementary role topological analysis can play in identification of targets. This study went on to predict 30 drugs

as potential candidates for CVD, confirming the activity of three via *in vitro* experiments.

Along with closeness and betweenness, Mortezaei *et al.* [86] used a measure called control centrality to characterize nodes that control sub-networks in a network of somatic and germline mutant genes and drug targets involved in neurodegenerative diseases (ND). They demonstrated that drug targets have the highest closeness and betweenness centralities. Control centrality analysis of the network revealed that the drug targets displayed a higher control centrality, followed by somatic mutations, and then the germline mutations. This analysis allowed them to propose candidates for repositioning for ND. One of the significant point in their study is the specificity of their results, which they obtain by comparing their ND network to degree-preserved random networks.

While studying and highlighting some novel pathways involved in schizophrenia, Gaspar and Green [87] used betweenness centrality to identify some of the hub genes that have been known to affect the nervous system and behavior. Some of the novel pathways to which the implicated genes belong to, that show enrichment, offer opportunities for drug repositioning.

Other examples include using degree and betweenness centrality to analyze a network of proteins involved in Alzheimer's disease, to highlight the hubs and central nodes such as EGFR, JUN and YAP1 [88]. They also analyzed subnetworks which yielded Ubiquitin C (UBC) and YAP1 as hubs for identified modules. These gene signatures were used as inputs to different drug repurposing tools. The study goes on to propose 27 repositioning candidates for Alzheimer's. This study utilizes multiple datasets, differential gene expression analysis methods, and multiple drug repurposing tools, to select the final list of drugs. However, from our own experience, we find many of the genes highlighted in this study (as Alzheimer's associated genes) being highlighted in other conditions such as metabolic diseases. Given the complexity of these neurodegenerative conditions, such studies need to be expanded to factor in features such as epigenetic changes, etc. to get a more comprehensive picture. Currently, very few therapeutic options are available for treating Alzheimer's disease, and several candidates have failed over the years.

Repositioning drugs for Alzheimer's disease was also demonstrated by McGarry *et al.* [89] using their tool called RESKO (REpositioning drugs using Side-Effects and Knowledge from Ontologies). They used closeness centrality and clustering coefficient, among other considerations, to rank drugs. Out of 77 potential candidates, 25 were examined. For each of the 25 drugs, its interaction network was built and analyzed for hubness, closeness and betweenness. Also, studying the PPI interactions of proteins in enriched pathways, it was observed that the disease genes in these networks have a high degree but low clustering coefficient.

Lv *et al.* [90] used network measures to suggest repositioning candidates for Autism. Drugs, drug interactions and hierarchical anatomical therapeutic chemical classification codes from literature were used to construct networks for centrality analysis. Drug associations in these networks were found using weak component analysis. Shortest path algorithm was used to find connections occurring between entities of the resulting drug similarity networks. For the drug network, cliques were identified using the clique percolation method yielding 34 cliques, and PageRank was used to identify 10 significant drugs in each clique.

Topological network analysis has also been applied to investigate skin conditions and tropical diseases. A network-based repositioning study for psoriasis was described by Manczinger *et al.* [91]. They used a support vector machine learning model on intracellular PPI network. They predicted 51 drugs, and closely examined 3 for which *in vitro/in vivo* data were available. To identify most effective targets in the PPI/drug-target network, they used betweenness centrality, which highlighted TNF-α and NFκB as top-ranking nodes in the network. Further *in vitro* experiments targeting these nodes via varying dosages of the three drugs confirmed the inhibitory effects of the drugs on the NFκB activity, and on TNF induced cytokines, thus highlighting the potential of the drugs in psoriasis treatment. They also carried out *in vivo* studies. This study proposed a novel algorithm to simulate drug effects on PPI networks. It also demonstrated the workflow from computational predictions to *in vivo* validation. The authors mention the drawback that the method is limited to drugs that interact with proteins.

Borba *et al.* [92] used a repositioning strategy using kinome network map to prioritize genes and propose seven drugs to treat the tropical disease Leishmaniasis. In analyzing the network, they used two approaches: essentiality-based approach (identifying essential genes via RNAi-based validations), and centrality-based approach: closeness and betweenness centrality to highlight 30 targets. 11 of these 30 targets were found to be targeted by 42 drugs, 15 of which had been reported as being in use for Leishmaniasis, and 27 were novel candidates. To repurpose drugs to treat sepsis, Han *et al.* [93] used degree and betweenness centrality to identify important nodes in the sepsis network.

While several of the above described methods attempt to improve upon the previously reported literature in the field, some of the common problems in these studies is the reliance on a single database, different methods of network reconstruction and data selection making it difficult to compare methods. Several of the networks covered in these studies are small networks, and particularly vulnerable to missing or inaccurate data. Perturbation studies and a confidence interval for different kinds of network measures would help gain insight into the effect of bias and noise present in the data in the analyses. Much of the data and literature have been collected for cancer studies, and hence there is a heavy bias, while literature for other rarer diseases such as tropical diseases is limited.

Considerations

While the studies described above are some relevant examples, and this coverage is not exhaustive, they provide an overview of the specific application of topological network measures for drug repositioning. Networks can be an extremely useful platform for integrating a wide variety of data. They can be a vital aid in making unintuitive connections, and helping converge diverse observations into a hypothesis building exercise. However, there are several considerations that need to be kept in sight.

Accuracy of computational predictions depends heavily on the underlying data

Quality of the data is crucial for the development of reliable hypotheses. Many different types of experiments have been reported over the years, and databases, even manually curated ones, depend heavily on these experimental reports as sources for the data they include. Differences in experimental conditions, experimental artifacts and non-reproducible observations, inconsistent/non-standardized nomenclature, add to noise. Changing technology has overcome several challenges in experimental data collection, however, the data reported using old methods remain a part of literature, and in our

databases. Additions to literature are being incorporated into databases on continuous basis—some of the databases are updated monthly—however, propagation of new information, especially where manual curation efforts are required, becomes a bottleneck. While multiple databases have been established for various data-types, there has been limited overlap of information between them. Thus, there is also an element of dataset bias in studies. Efforts are ongoing to address these concerns. For example, network building and reinforcing algorithms have been designed for drug repositioning, which use literature confirmation from multiple sources to build a consensus network [94]. Nevertheless, incomplete databases will mean that we would have to keep reviewing and updating findings to feed back into the knowledge web.

While biological networks have been seen as scale-free (few nodes with very high connectivity, while a majority of them with low connections), study biases result in disproportionate distribution of literature, especially in intensively studied diseases like cancer. In rare diseases, less frequently studied genes may also link to major hubs, and thus drive complex conditions, and could be drug targets. Uncovering the missing links remains crucial. The issue of missing data and bias directly impacts quantitative evaluation of networks, causing cases of class imbalance commonly seen in applications of machine learning for drug classification/repositioning [70]. Some of the network measures have been shown to be very sensitive to network structure.

It would be useful to bear in mind that while these computational approaches are extremely helpful in gaining insights into connections and patterns—spread across literature, from different studies, different assays and different contexts—it will not produce new discoveries out of nothing. For drug repositioning, computational efforts are based on existing information. To maximize knowledge discovery, prior information must be built on accurate data.

Building networks-different strategies, different outcomes

Several different approaches to construction of networks are seen in literature. One aspect is scope of networks. One could build cell-specific, tissue-specific, organ-specific, system-specific network, or consider a whole PPI network for analysis. If a cell-specific network is built, systemic effects would be missed. Since biological entities have to be in proximity to interact, network perturbation studies would require specific networks for predictive analyses. A network perspective of the successful and withdrawn/failed drugs has drawn attention in the recent times. Several observations regarding network properties of such drugs have been made based on their drug target and side-effects data [66, 70]. In one of the studies, the authors highlight that tissue specific proteins are underused as therapeutic targets [66]. These are desirable targets as far as fewer side-effects are expected, however, these proteins are not as conserved as more ubiquitous proteins. Their network analysis results show that these tissue-specific proteins have higher betweenness centrality values. Biological networks contain modules formed by groups of functionally linked entities. Indeed, as Kitano [95] points out, modularity is an excellent mechanism that limits spread of perturbations in the system, and addresses the local environment. However, in certain cases, such as targeting viral proteins, a more systemic effect of drugs would be preferred. Hence, network building requires context specificity, though it might make generalization/comparison of different approaches difficult. Based on whether local or global topological measures are being used for drug repositioning analysis, applications will need consideration. Another aspect is different network building strategies for network expansion. Starting with the known information, for example, a set of mutated genes in cancer, networks can be expanded by including only the first neighbor interactions, or several step neighbors, or using propagation algorithms. For the former, a drawback would be that peripheral genes are known to play a role in disease propagation, and would be missed, while for the latter, the fact that non-specific interactions would add to noise.

Reliable predictions of drug repositioning for known drugs and conditions based on network measures require a reliable network, in terms of data quality and size of the underlying network. Network building considerations directly affect conclusions and predictions of analyses. Mora *et al.* [69] explored the role of network size and completeness in evaluating drug target predictions, based on degree and betweenness centrality. They show that using degree as a reporting measure for drug targets is unreliable, as it heavily depends on the choice of dataset. They also reported that using subsets of data, in their specific study, to use a subnetwork of PPIs based on different types of reliability scores/criteria, reduces the reliability of using degree or centralities (they investigated betweenness and closeness) as predictors of drug targets. Indeed, there are practical difficulties in balancing quality and quantity of information used to build reliable networks. Data selection criteria can have a direct impact on the topology of the network constructed, and thus directly influence the outcomes of analyses and predictions (Figure 2). There are some excellent studies on the impact of network sampling, edge deletions, node deletions/additions on centrality measures [97–100]. Thus, any missing or additional data in the network will affect the outcomes; the extent to which it affects the accuracy of predictions depends on the network.

Integrating different approaches for network building in view of data-availability

As Jin and Wong discuss in their work [13], several methods of varying levels of complexity and sophistication have been proposed over the years, which can be integrated into efficient pipelines depending upon the condition of interest and data available. Several conditions, such as cancer, are well studied, and studies have generated a multitude of datasets of different data-types. However, for diseases such as cardiovascular diseases, the amount of data and data-types available, is currently limited. Hence, methods requiring several data-types may not work for such conditions. While topological analysis is highly context dependent, and may require extensive computational assessment for its application to diverse networks, it could prove helpful in such scenarios, where other data-types are unavailable. It would be a cost-effective first-approach in using available data to obtain insights on repositioning opportunities, based on shared molecular mechanisms, or shared genes/pathways between rare and well-studied diseases. In cases of limited patient numbers, or diseases such as cancer where certain populations are generally responsive to treatments, other kinds of data such as electronic health records (EHR) can be integrated. Based on EHRs and omics profiles, clustering and sub-population identification can help recruit patients most likely to respond, and would allow to better transition from computational predictions to clinical trials.

Benchmarking efforts for reliability of network measures in context of their network structures can add confidence to efforts of applying these measures for predictive analysis. Given the

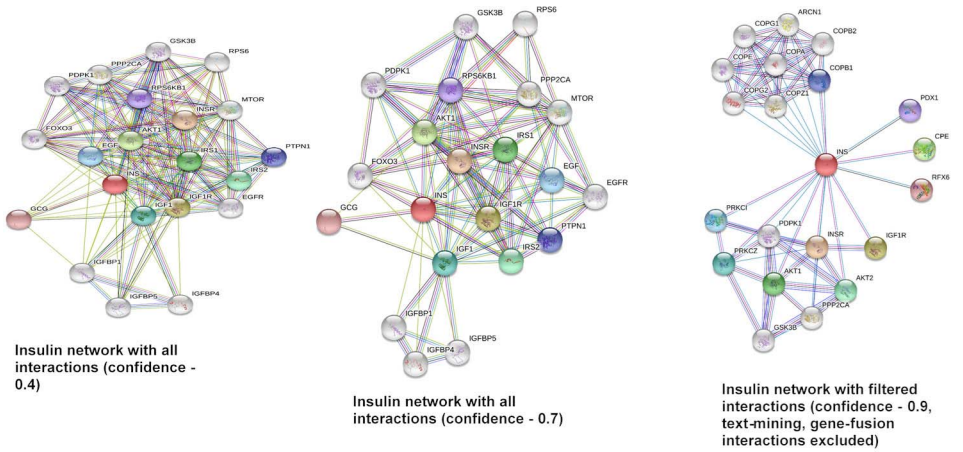

Figure 2. Effect of data selection on network topology: seen here are the networks of the protein insulin with different criteria for selection of its interactions, built using the STRING [96] database. While one can chose interactions with lower confidence to allow scope for exploration of possible links, it leads to a denser structure. As more stringent criteria is applied, while the quality of interactions is higher, possibilities of discovering potentially novel interactions decreases. The effect on topology is prominent, and would directly affect analysis and prediction based on these different networks.

vast amount of data and several approaches/measures, the number of permutations that can be attempted may be constrained. Advances in machine learning/artificial intelligence may lead to more concerted efforts in this space. One of the applications is, of-course, biomedical data extraction and network creation. Another area where advances in ML/AI could help is to learn to assess quality and quantity of input data, and choose/combine different methods depending on the data available. Different methods have their own strengths and weaknesses. Given the variations in both the data available and the number of methods/metrics proposed, manually running these permutations may not be feasible. Advances in ML/AI techniques for decision making and implementation of these methods may help researchers explore various dimensions of data available, and establish comprehensive analyses.

Exploring other network measures—opportunities and risks

While several measures have been proposed, very few measures have been applied in these numerous studies. For example, betweenness centrality has been used extensively, the rationale being that these nodes are crucial for communication between modules. Some of the other frequently seen measures are degree, closeness centrality, PageRank, eigenvector centrality and clustering coefficient. However, these represent a small percentage of the proposed measures. Whether there are other measures that can be applied to such networks that also yield insights into key nodes as drugs/drug targets should be systematically explored. For example, motif-based centrality has been applied to gene-regulatory networks [101]. If analysis of such motifs in network structures provides insights into side-effects/toxicity information for drug targets, it could help anticipate adverse reactions. Indeed, network topology has been reported to influence synergistic/antagonistic effects of drug combinations [102]. The study reported, among other observations, that motifs that are involved in positive feedback, and are also drug targets, could be involved in antagonistic (buffering) effects of drug combinations. A wider range of topological measures can be applied to drug repositioning studies to highlight unintuitive relationships between drugs, targets and various conditions.

However, the type of network under consideration should be contextualized. Are all measures applicable to all types of networks? Several points emerge and should be consciously analyzed such as applying a measure designed for an undirected network to directed ones. For example, one would need to assess if motif-based centrality or PageRank centrality yield similar interpretations for gene-regulatory, protein–protein and metabolic networks. In a gene regulatory or a metabolic network, where reaction specificity is seen between specific substrates and enzymes, explicit interpretation of random walk-based centrality may be required. It would be helpful to know under which conditions certain measures can be applied, and in which ones they cannot. Robustness of these measures over different networks is also a consideration. If a measure is a predictive feature for one kind of network, would it be a reliable feature for other networks? The nature of underlying network (field of study, degree distribution, topological characteristics) becomes significant in order to apply these metrics. Dependence of network measures on the size, type and topology of the networks is well documented [99, 103]. In the context of analyzing brain connectivity datasets, Rubinov and Sporns differentiate between anatomical and functional networks, and highlight the differences in applying and interpreting such measures to them [27]. Comparison of different networks becomes difficult. The same measure can yield different outcomes in different networks, which was the case in the study presented by Wang *et al.* [101]. They reported applying in-degree, out-degree, total degree, PageRank and betweenness to five different directed biological networks, and found a lack of robustness in these measures in identifying nodes that occur frequently in motifs (and thus assumed to be important) across the networks, and indeed, proposed a newer, more robust measure based on frequency of node appearances in motifs, in their paper. Betweenness centrality, which is one of the most frequently seen measure in literature, was shown also to be unstable [104]. While addressing the reliability of centrality measures for drug repositioning, Da Silva Lopes *et al.* [70] conclude, based on their study, that centrality measures can be used to classify problematic drugs based on protein-structure similarity networks and PPI networks, though with lower reliability. However, they used a limited number of measures, hence similar studies will have to be carried out for other measures to generalize this statement. Indeed, their study

included the successfully repositioned drug Thalidomide, which the study found difficulty to classify. Udrescu *et al*. [77] found modularity (clustering) to be a better predictor of drug properties than topology-based clustering. Given the heterogeneity of networks and measures, some studies have proposed integrated measures such as 'integrated hubness score' [105] to overcome the limitations of using a single metric. Thus, perhaps such an integrated score considering centralities that may reflect specific centrality scores for assessing re purposing potential can be devised.

Theoretical foundation of different centrality measures may be different, but may yield the same outcomes. Correlations among these measures are dependent on the topology of the underlying network. For example, on the basis of a study analyzing 212 networks from different fields, Oldham *et al*. [103] observed correlations between 17 centrality-based measures. They showed presence of clusters across networks that showed high correlations between random-walk closeness centrality and information centrality, and between Katz centrality and total communicability centrality, but showed that participation coefficient and bridging centrality had the lowest average correlation with other measures. However, besides these core clusters, other clusters across network showed a wide variation, indicating network specific roles. In general, all studies have indicated that there is yet to be consensus on several aspects of application of network measures across different fields, most studies alluding to the dependence on the underlying network. This affects the applicability of topological analysis. Take the case of machine learning, where topological measures are used as features. It is likely that across different networks, the machine learning algorithm will be required to vary the features it deems useful/redundant, thus introducing a layer of variability, for example, for patient-specific network analysis. While it may not necessarily be a disadvantage, the variability in application of network measures is certainly a detail one should keep in sight.

While designing these measures, several assumptions are made about the networks, depending on the nature of nodes and edges, the nature of their interactions, the kind of external inputs received. One must ensure these assumptions are comparable while applying measures from one field to the other. For example, communication in social networks may not be comparable to a biological input in the form of a ligand binding to a receptor, given that ligands have finite quantities, while speech has no such limitations. Bringmann *et al*. [106] questioned, recently, the applicability of betweenness centrality and closeness to psychological networks, and cautioned against the use of one type of network measure in a context where it may not be applicable. Take the simplest case of betweenness centrality. The underlying assumption of the nodes being central because a high fraction of shortest paths pass through them could be evaluated through the lens of biological context. Does the assumption hold ground? It is unlikely that biological entities interact only via 'shortest distances' or direct interactions, indeed because of redundancies encoded in biological systems, there are usually several layers to interactions between two entities. Other versions of betweenness centrality (the random walk betweenness centrality, flow betweenness centrality, communicability betweenness centrality) have been proposed to overcome the shortest path restriction. The question of which one would better apply to the biological system under study must be evaluated. Another application of graph methods under study today is link prediction to determine possible connections between entities. But for a case of drug–protein interaction, it would be useful to bear in mind the biological constraints: conditions such as proximity (presence of a protein in nucleus, cytoplasm or in the cell membrane), active transport of drug into the system, lysosomal degradation, drug solubility, presence of competitive inhibitors, structural integrity of the drug and the protein, active efflux of drugs out of the cell, etc. Thus, while social networks today are unconstrained by absence of physical connections between potential interactors, the degree of freedom available for applying methods directly from social network analysis to biological networks may be much more restricted. The nature of flow of information, perturbations and their effects on the neighboring nodes in diverse networks (social, communication networks such as internet and telephone, transport and logistics, and biological) need to be evaluated before such methods can be used between disciplines.

There is indeed scope and need for identifying specific topological features that can be applied to drug/target/disease networks from among hundreds of measures proposed. For example, topological coefficient gives the number of neighbors shared between two nodes. For a drug target network in the case of cancer, this could help identify other candidates in the neighborhood of the original target, in case resistance sets in. Applying this analysis to patient-specific gene expression/protein-interaction networks could help personalize treatment, and thereby increase the possibility of patient response to alternative therapies. Another interesting measure is Burt's constraint. Briefly put, it is an indication of the influence of position in accessing information. It is related to another concept of 'weak ties' or positional advantage in social networks. Nodes which are in a position to exchange non-redundant information are valuable [107]. Higher values of Burt's constraint would imply, for drug repositioning studies, identifying drugs working on targets sharing stronger connections. Estrada *et al*. [108] had proposed bipartivity for identifying essential proteins in a protein-interaction network, which is the membership of a protein in two or more clusters. While essentiality would require low values of bipartivity, perhaps applying this measure to drug networks may help identify drugs with a high co-efficient participating in multiple clusters, and thus, candidates with higher repositioning potential.

While some measures may not be relevant or applicable, literature bias toward a select few metrics makes it difficult to discuss or compare the potential of some of the other proposed metrics in topological analysis. Many of the classical measures such as betweenness centrality have been explored further and alternative metrics have been proposed to overcome some of the inherent limitations of these metrics. Many of these measures remain unexplored, and thus, opportunities for exploring these measures can indeed be helpful to the community by studying them and generating literature for comparison in line with the popular metrics in the field.

As the quantity of available data, and therefore network size increases, computational resources and time also add constraints [103], and thus influence which measures would be preferred. Computational cost has also prompted researchers to propose alternate metrics, which offer reduced computational times, an example being that of localized bridging centrality (LBC) in place of bridging centrality [109]. LBC can be computed, according to its authors, using parallel processing.

Caveats: limitations, experimental design biases and on the importance of definitions

Topological analysis may help highlight previously unknown relationships between targets, drugs, diseases computationally, and indeed, provide a sound theoretical grounding for

experimental/clinical validations for repositioning candidates. However, in practice, additional considerations may hamper repositioning efforts. While the main advantage of repositioning may be the availability of pre-clinical data, different dosing may be required to observe the effect of a known drug in a different disease setting. In case of higher dosing requirements, toxicity issues may lead to unsuccessful repositioning attempts. Hence, while a network-based analysis may indicate previously unconsidered links, additional dosage/toxicity assessment may add limitations.

In order to control for experimental/methodological biases, the design and selection of appropriate control graphs is an important aspect. As in experimental biology, which uses untreated/wild-type controls, random graphs are used as controls in network biology to determine specificity/significance of the observations. Considering the noise present in data, identifying disease/drug/target specific effects improves the translatable potential of predictions, and reduces the risk of off-target effects.

One of the serious challenges is the inconsistent use of terms in the field of topological network analysis. Originally, graph theory evolved from the solution to the problem of the seven bridges of Königsberg, and had geometrical basis. Its application to sociological research led to the development of a host of measures, and terminologies changed contexts in different disciplines. For example, in most cases, 'hubs' are taken to be nodes with high degree, but at times, loosely taken to be highly central/topologically important nodes. The resource Centiserver [110], lists several measures on their website (https://www.centiserver.org/centrality/) such as betweenness centrality and bottleneck, bridgeness, bridgeness centrality and bridging. Betweenness is defined based on shortest paths, while bridging is described as nodes connecting high degree connections, and bridging centrality takes into account both communicability and degree. Thus, nodes with high bridging centrality connect highly connected modules. However, in practice, betweenness centrality is often used interchangeably with bridging/bridging centrality. Hence, adapting definitions from various field with appropriate changes to ensure correct interpretation with context specificity is important, especially when the predictions are applied to complex fields such as the drug discovery and repositioning.

Perspective/conclusions

This discussion of network measures for repositioning focuses on computational insights into new relationships between drugs/diseases. It does not consider other ways of repositioning, such as changes in formulation, changes in route of administration or new combinations of previously known drugs. In essence, drug re positioning is a broad topic, and several terminologies and use cases exist—coming from pharma companies, doctors and health care professionals, patients and regulatory bodies.

Drug repositioning has an enormous potential and could result in large benefits, especially for patients at the end of the healthcare chain. While tremendous leaps have been made in computational approaches to drug repositioning, the interplay between experimental, clinical data and optimal computational methods for network building will need to be a continuous, iterative process for some time to come.

Networks have the potential to offer non-intuitive insights for drug repositioning. They enable exploration of large-scale, dependent, heterogeneous data that are especially useful in biology, given that this data represent dynamic, interacting layers of a living and evolving system. Several methods and measures exist that aim at understanding the key nodes, important interactions, structure and communication in networks. The field of drug repositioning has also leveraged biological network analysis via different network measures to explore new opportunities. Several studies have mentioned using various measures for analysis and prediction of candidates for repositioning. There remains much to explore, in terms of the types of measures used, and a thorough understanding of context of applicability of these measures across diverse network types. However, because the applicability and predictive abilities of different measures are heavily dependent on the topology of the underlying network, there is an element of variability inherent to the approach. Finally, despite having good quality inputs, and sound analysis methodology, candidates proposed may not actually be viable for a given condition. Patient specific genetics, food habits and environment, response to specific drugs, and incomplete understanding of causality are a few variables that may not be accounted for, and end up having an un-anticipated effect on the predicted candidates. A prediction may also be an artifact due to the use of limited or conflicting data, or a single data-type. An example is the case where a repositioned drug was suggested as a therapeutic for irritable bowel disease (IBD), but the drug has diarrhea as a side-effect, which is a symptom of IBD [111, 112]. Thus, even after such predictions are available, additional feasibility analyses need to be included in the repositioning pipeline. Existing safety and efficacy data should be revisited, so that drugs already in use for other conditions can benefit from availability of patient records which can be analyzed to assess safety and efficacy outcomes, and check for indications of efficacy for new conditions (from side-effects data, for example). If patient data are available, simulation on patient specific networks may be attempted. Multiple computational methods based on diverse data inputs should be combined for robust predictions. New technology (mobiles, fitness monitoring apps, etc.) is being leveraged to collect real-time data and conduct virtual trials. These approaches can help identify specific patient populations, and monitor the outcomes of prescribing repositioned drugs. Thus, computational predictions need to be integrated with *in vitro, in vivo*, and patient data to assess feasibility and efficacy of the proposed candidate. Such combination of approaches will allow us to better address the gap between predictions and clinical trial success of candidates.

While limitations in this approach do exist, it is nevertheless useful in establishing connections between diverse datasets, and providing new insights using available data, which make it a pertinent approach for drug repositioning studies.

Key Points

- Drug repositioning efforts have been contributing to a number of successful drug approvals in the past decade.
- Network analysis using graph theory principles offers insights by integrating diverse data types, large amounts of data and several sophisticated metrics highlighting important nodes and connections.
- Drug repositioning studies have seen inputs using several popular metrics such as node degree, betweenness and closeness centrality, PageRank and eigenvector centralities, etc.
- A large number of measures exist, coming from a variety of fields. However, only few have been applied

to biological networks, and especially in drug repositioning studies.
- A wider, systematic exploration of different measures, with a comprehensive discussion on their dependence on network topology and applicability in drug repositioning, may contribute to drug repositioning studies and its clinical translation.

References

1. Eastman RT, Roth JS, Brimacombe KR, *et al*. Remdesivir: a review of its discovery and development leading to emergency use authorization for treatment of COVID-19. *ACS Cent Sci* 2020;**6**:672–83.
2. Falzone L, Salomone S, Libra M. Evolution of cancer pharmacological treatments at the turn of the third millennium. *Front Pharmacol* 2018;**9**:1300.
3. Galiè N, Ghofrani HA, Torbicki A, *et al*. Sildenafil citrate therapy for pulmonary arterial hypertension. *N Engl J Med* 2005;**353**:2148–57.
4. Papapetropoulos A, Szabo C. Inventing new therapies without reinventing the wheel: the power of drug repurposing. *Br J Pharmacol* 2018;**175**:165–7.
5. Quirke VM. Tamoxifen from failed contraceptive pill to best-selling breast cancer medicine: a case-study in pharmaceutical innovation. *Front Pharmacol* 2017;**8**:620.
6. Pushpakom S, Iorio F, Eyers PA, *et al*. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* 2018;**18**:41–58.
7. Akhondzadeh S. The importance of clinical trials in drug development. *Avicenna J Med Biotechnol* 2016;**8**:151.
8. Yella JK, Yaddanapudi S, Wang Y, *et al*. Changing trends in computational drug repositioning. *Pharmaceuticals* 2018;**11**:57.
9. Hwang TJ, Carpenter D, Lauffenburger JC, *et al*. Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMA Intern Med* 2016;**176**:1826–33.
10. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. *Contemp Clin Trials Commun* 2018;**11**:156–64.
11. Mak IW, Evaniew N, Ghert M. Lost in translation: animal models and clinical trials in cancer treatment. *Am J Transl Res* 2014;**6**:114–8.
12. Hernandez JJ, Pryszlak M, Smith L, *et al*. Giving drugs a second chance: overcoming regulatory and financial hurdles in repurposing approved drugs as cancer therapeutics. *Front Oncol* 2017;**7**:273.
13. Jin G, Wong STC. Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. *Drug Discov Today* 2014;**19**:637–44.
14. Vanhaelen Q, Mamoshina P, Aliper AM, *et al*. Design of efficient computational workflows for in silico drug repurposing. *Drug Discov Today* 2017;**22**:210–22.
15. Oprea TI, Bauman JE, Bologa CG, *et al*. Drug repurposing from an academic perspective. *Drug Discov Today Ther Strateg* 2011;**8**:61–9.
16. Oprea TI, Mestres J. Drug repurposing: far beyond new targets for old drugs. *AAPS J* 2012;**14**:759–63.
17. Kato S, Moulder SL, Ueno NT, *et al*. Challenges and perspective of drug repurposing strategies in early phase clinical trials. *Onco Targets Ther* 2015;**2**:576–80.
18. Cha Y, Erez T, Reynolds IJ, *et al*. Drug repurposing from the perspective of pharmaceutical companies. *Br J Pharmacol* 2018;**175**:168–80.
19. Swamidass SJ, Lu ZY, Agarwal P, *et al*. Computational approaches to drug repurposing and pharmacology. *Pac Symp Biocomput* 2014;110–3.
20. Li P, Fu Y, Wang Y. Network based approach to drug discovery: a mini review. *Mini-Reviews Med Chem* 2015;**15**:687–95.
21. Park K. A review of computational drug repurposing. *Transl Clin Pharmacol* 2019;**27**:59–63.
22. Delavan B, Roberts R, Huang R, *et al*. Computational drug repositioning for rare diseases in the era of precision medicine. *Drug Discov Today* 2018;**23**:382–94.
23. Li J, Zheng S, Chen B, *et al*. A survey of current trends in computational drug repositioning. *Brief Bioinform* 2015;**17**:2–12.
24. Xue H, Li J, Xie H, *et al*. Review of drug repositioning approaches and resources. *Int J Biol Sci* 2018;**14**:1232–44.
25. March-Vila E, Pinzi L, Sturm N, *et al*. On the integration of in silico drug design methods for drug repurposing. *Front Pharmacol* 2017;**8**:298.
26. Liu C, Ma Y, Zhao J, *et al*. Computational network biology: data, models, and applications. *Phys Rep* 2020;**846**:1–66.
27. Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 2010;**52**:1059–69.
28. Csermely P, Korcsmáros T, Kiss HJM, *et al*. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther* 2013;**138**:333–408.
29. Pavlopoulos GA, Secrier M, Moschopoulos CN, *et al*. Using graph theory to analyze biological networks. *Bio Data Min* 2011;**4**:10.
30. Ingalhalikar M, Smith AR, Bloy L, *et al*. Identifying subpopulations via unsupervised cluster analysis on multiedge similarity graphs BT-medical image computing and computer-assisted intervention–MICCAI. 2012; 254–61.
31. Percha B, Altman RB. A global network of biomedical relationships derived from text. *Bioinformatics* 2018;**34**:2614–24.
32. Nicholson DN, Himmelstein DS, Greene CS. Expanding a database-derived biomedical knowledge graph via multirelation extraction from biomedical abstracts. *bioRxiv* 2020;730085.
33. Würth R, Thellung S, Bajetto A, *et al*. Drug-repositioning opportunities for cancer therapy: novel molecular targets for known compounds. *Drug Discov Today* 2016;**21**:190–9.
34. Vargesson N. Thalidomide-induced teratogenesis: history and mechanisms. *Birth Defects Res C Embryo Today* 2015;**105**:140–56.
35. Ferguson LB, Harris RA, Mayfield RD. From gene networks to drugs: systems pharmacology approaches for AUD. *Psychopharmacology (Berl)* 2018;**235**:1635–62.
36. Gramatica R, Di Matteo T, Giorgetti S, *et al*. Graph theory enables drug repurposing–how a mathematical model can drive the discovery of hidden mechanisms of action. *PLoS One* 2014;**9**:e84912.
37. Donner Y, Kazmierczak S, Fortney K. Drug repurposing using deep embeddings of gene expression profiles. *Mol Pharm* 2018;**15**:4314–25.
38. Tan SK, Jermakowicz A, Mookhtiar AK, *et al*. Drug repositioning in glioblastoma: a pathway perspective. *Front Pharmacol* 2018;**9**:218.
39. Li YY, Jones SJM. Drug repositioning for personalized medicine. *Genome Med* 2012;**4**(3):27.

40. Tang J, Aittokallio T. Network pharmacology strategies toward multi-target anticancer therapies: from computational models to experimental design principles. *Curr Pharm Des* 2014;**20**:23–36.
41. Zhang P, Wang F, Hu J. Towards drug repositioning: a unified computational framework for integrating multiple aspects of drug similarity and disease similarity. *AMIA Annu Symp Proc* 2014;**2014**:1258–67.
42. Sanseau P, Koehler J. Editorial: computational methods for drug repurposing. *Brief Bioinform* 2011;**12**:301–2.
43. Iorio F, Tagliaferri R, Di Bernardo D. Identifying network of drug mode of action by gene expression profiling. *J Comput Biol* 2009;**16**:241–51.
44. Iorio F, Bosotti R, Scacheri E, *et al*. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci USA* 2010;**107**:14621–6.
45. Carrella D, Napolitano F, Rispoli R, *et al*. Mantra 2.0: an online collaborative resource for drug mode of action and repurposing by network analysis. *Bioinformatics* 2014;**30**:1787–8.
46. Abdolhosseini F, Azarkhalili B, Maazallahi A, *et al*. Cell identity codes: understanding cell identity from gene expression profiles using deep neural networks. *Sci Rep* 2019;**9**:2342.
47. Filzen TM, Kutchukian PS, Hermes JD, *et al*. Representing high throughput expression profiles via perturbation barcodes reveals compound targets. *PLoS Comput Biol* 2017;**13**:e1005335.
48. Donner Y, Kazmierczak S, Fortney K. Drug repurposing using deep embeddings of gene expression profiles. *Mol Pharm* 2018;**15**:4314–25.
49. Cho H, Berger B, Peng J. Compact integration of multi-network topology for functional analysis of genes. *Cell Syst* 2016;**3**:540–8.e5.
50. Taroni JN, Grayson PC, Hu Q, *et al*. MultiPLIER: a transfer learning framework for transcriptomics reveals systemic features of rare disease. *Cell Syst* 2019;**8**: 380–94.e4.
51. Wang S, Huang E, Cairns J, *et al*. Identification of pathways associated with chemosensitivity through network embedding. *PLoS Comput Biol* 2019;**15**:e1006864.
52. Henaff M, Bruna J, LeCun Y. Deep convolutional networks on graph-structured data. *ArXiv:1506.05163v1 [cs.LG]* 2015; abs/1506.0.
53. Sosa DN, Derry A, Guo M, *et al*. A literature-based knowledge graph embedding method for identifying drug repurposing opportunities in rare diseases. *Pac Symp Biocomput* 2020;**25**:463–74.
54. Perozzi B, Al-Rfou R, Skiena S. Deepwalk: online learning of social representations. In: Proc. 20th ACM SIGKDD Int Conf Knowl Discov data Min. 2014; pp. 701–10.
55. Zong N, Kim H, Ngo V, *et al*. Deep mining heterogeneous networks of biomedical linked data to predict novel drug–target associations. *Bioinformatics* 2017;**33**:2337–44.
56. Chen Z, Wang X, Gao P, *et al*. Predicting disease related microRNA based on similarity and topology. *Cells*. 2019;**8**(11):1405.
57. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 2018;**34**:i457–66.
58. Feng Y-H, Zhang S-W, Shi J-Y. DPDDI: a deep predictor for drug-drug interactions. *BMC Bioinformatics* 2020;**21**: 419.
59. Xiao Z, Deng Y. Graph embedding-based novel protein interaction prediction via higher-order graph convolutional network. *PLoS One* 2020;**15**:e0238915.
60. Eslami Manoochehri H, Nourani M. Drug-target interaction prediction using semi-bipartite graph model and deep learning. *BMC Bioinformatics* 2020;**21**:248.
61. Fan K, Zhang Y. Pseudo2GO: a graph-based deep learning method for pseudogene function prediction by borrowing information from coding genes. *Front Genet* 2020;**11**:807.
62. Hu JX, Thomas CE, Brunak S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* 2016;**17**:615–29.
63. Ashtiani M, Salehzadeh-Yazdi A, Razaghi-Moghadam Z, *et al*. A systematic survey of centrality measures for protein-protein interaction networks. *BMC Syst Biol* 2018;**12**:80.
64. Pournoor E, Elmi N, Masoudi-Sobhanzadeh Y, *et al*. Disease global behavior: a systematic study of the human interactome network reveals conserved topological features among categories of diseases. *Informatics Med Unlocked* 2019;**17**:100249.
65. Piraveenan M, Wimalawarne K, Kasthurirathn D. Centrality and composition of four-node motifs in metabolic networks. *Procedia Comput Sci* 2013;**18**:409–18.
66. Ryaboshapkina M, Hammar M. Tissue-specific genes as an underutilized resource in drug discovery. *Sci Rep* 2019;**9**:7233.
67. Zhang M, Zhu C, Jacomy A, *et al*. The orphan disease networks. *Am J Hum Genet* 2011;**88**:755–66.
68. de Chassey B, Meyniel-Schicklin L, Aublin-Gex A, *et al*. New horizons for antiviral drug discovery from virus–host protein interaction networks. *Curr Opin Virol* 2012;**2**:606–13.
69. Mora A, Donaldson IM. Effects of protein interaction data integration, representation and reliability on the use of network properties for drug target prediction. *BMC Bioinformatics* 2012;**13**:294.
70. Da Silva Lopes TJ, Shoemaker J, Matsuoka Y, *et al*. Identifying problematic drugs based on the characteristics of their targets. *Front Pharmacol* 2015;**6**:186.
71. Sun J, Huang L-C, Xu H, *et al*. Network-assisted prediction of potential drugs for addiction. *Biomed Res Int* 2014;**2014**:258784.
72. Boccaletti S, Latora V, Moreno Y, *et al*. Complex networks: structure and dynamics. *Phys. Rep*. 2006;**424**:175–308.
73. Newman MEJ. A measure of betweenness centrality based on random walks. *Soc. Networks* 2005;**27**:39–54.
74. Lohmann G, Margulies DS, Horstmann A, *et al*. Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain. *PLoS One* 2010;**5**:e10232.
75. Liu C, Ma Y, Zhao J, *et al*. Computational network biology: data, models, and applications. *Phys. Rep*. 2020;**846**:1–66.
76. Udrescu L, Bogdan P, Chiş A, *et al*. Uncovering new drug properties in target-based drug-drug similarity networks. *Pharmaceutics* 2020, 12(**9**), 879.
77. Udrescu L, Sbârcea L, Topîrceanu A, *et al*. Clustering drug-drug interaction networks with energy model layouts: community analysis and drug repurposing. *Sci Rep* 2016;**6**: 32745.
78. Guney E, Menche J, Vidal M, *et al*. Network-based in silico drug efficacy screening. *Nat Commun* 2016;**7**:10331.
79. Cheng F, Desai RJ, Handy DE, *et al*. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun* 2018;**9**:2691.

80. Guney E. Investigating side effect modules in the interactome and their use in drug adverse effect discovery. *Springer Proc Complex* 2017;239–50.
81. Zhou Y, Hou Y, Shen J, *et al*. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov* 2020;**6**:14.
82. Vitali F, Cohen LD, Demartini A, *et al*. A network-based data integration approach to support drug repurposing and multi-target therapies in triple negative breast cancer. *PLoS One* 2016;**11**:e0162407.
83. Jadamba E, Shin M. A systematic framework for drug repositioning from integrated omics and drug phenotype profiles using pathway-drug network. *Biomed Res Int* 2016;**2016**:7147039.
84. Bourdakou MM, Spyrou GM. Informed walks: whispering hints to gene hunters inside networks' jungle. *BMC Syst Biol* 2017;**11**:97.
85. Huang C-H, Ciou J-S, Chen S-T, *et al*. Identify potential drugs for cardiovascular diseases caused by stress-induced genes in vascular smooth muscle cells. *Peer J* 2016;**4**:e2478.
86. Mortezaei Z, Cazier J-B, Mehrabi AA, *et al*. Novel putative drugs and key initiating genes for neurodegenerative disease determined using network-based genetic integrative analysis. *J Cell Biochem* 2019;**120**:5459–71.
87. Gaspar HA, Breen G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. *Sci Rep* 2017;**7**:12460.
88. Siavelis JC, Bourdakou MM, Athanasiadis EI, *et al*. Bioinformatics methods in drug repurposing for Alzheimer's disease. *Brief Bioinform* 2015;**17**:322–35.
89. McGarry K, Graham Y, McDonald S, *et al*. RESKO: repositioning drugs by using side effects and knowledge from ontologies. *Knowledge-Based Syst* 2018;**160**:34–48.
90. Lv Y, Ding Y, Song M, *et al*. Topology-driven trend analysis for drug discovery. *J Informet* 2018;**12**:893–905.
91. Manczinger M, Bodnár VÁ, Papp BT, *et al*. Drug repurposing by simulating flow through protein–protein interaction networks. *Clin Pharmacol Ther* 2018;**103**:511–20.
92. Borba JVB, Silva AC, Ramos PIP, *et al*. Unveiling the kinomes of *Leishmania infantum* and *L. braziliensis* empowers the discovery of new kinase targets and Antileishmanial compounds. *Comput. Struct. Biotechnol J* 2019;**17**:352–61.
93. Han H-W, Hahn S, Jeong HY, *et al*. LINCS L1000 dataset-based repositioning of CGP-60474 as a highly potent anti-endotoxemic agent. *Sci Rep* 2018;**8**:14969.
94. Nam Y, Kim M, Chang H-S, *et al*. Drug repurposing with network reinforcement. *BMC Bioinformatics* 2019;**20**:383.
95. Kitano H. A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov* 2007;**6**:202–10.
96. Szklarczyk D, Morris JH, Cook H, *et al*. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res* 2016;**45**:D362–8.
97. Costenbader E, Valente TW. The stability of centrality measures when networks are sampled. *Soc Networks* 2003;**25**:283–307.
98. Borgatti SP, Carley KM, Krackhardt D. On the robustness of centrality measures under conditions of imperfect data. *Soc Networks* 2006;**28**:124–36.
99. Smith JA, Moody J. Structural effects of network sampling coverage I: nodes missing at random. *Soc Networks* 2013;**35**:652–68.
100. Naujokaitis-Lewis IR, Rico Y, Lovell J, *et al*. Implications of incomplete networks on estimation of landscape genetic connectivity. *Conserv Genet* 2013;**14**:287–98.
101. Wang P, Lü J, Yu X. Identification of important nodes in directed biological networks: a network motif approach. *PLoS One* 2014;**9**:e106132.
102. Yin N, Ma W, Pei J, *et al*. Synergistic and antagonistic drug combinations depend on network topology. *PLoS One* 2014;**9**:e93960.
103. Oldham S, Fulcher B, Parkes L, *et al*. Consistency and differences between centrality measures across distinct classes of networks. *PLoS One* 2019;**14**:e0220061.
104. Segarra S, Ribeiro A. Stability and continuity of centrality measures in weighted graphs. *IEEE Trans Signal Process* 2016;**64**:543–55.
105. Salavaty A, Ramialison M, Currie PD. IHS: an integrative method for the identification of network hubs. bioRxiv 2020.02.17.953430 (Preprint).
106. Bringmann LF, Elmer T, Epskamp S, *et al*. What do centrality measures measure in psychological networks? *J Abnorm Psychol* 2019;**128**:892.
107. Ramadan E, Alinsaif S, Hassan MR. Network topology measures for identifying disease-gene association in breast cancer. *BMC Bioinformatics* 2016;**17**:274.
108. Estrada E. Protein bipartivity and essentiality in the yeast protein–protein interaction network. *J Proteome Res* 2006;**5**:2177–84.
109. Nanda S, Kotz D. Localized bridging centrality for distributed network analysis. In: *Proc. 17th Int. Conf. Comput. Commun. Networks 2008*; 2008; pp. 1–6
110. Jalili M, Salehzadeh-Yazdi A, Asgari Y, *et al*. CentiServer: a comprehensive resource, web-based application and R package for centrality analysis. *PLoS One* 2015;**10**:e0143111.
111. Dudley JT, Sirota M, Shenoy M, *et al*. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci Transl Med* 2011;**3**:96ra76.
112. Deftereos S. Is a single type of data sufficient for accurate computational drug repositioning? *Sci Transl Med* 2011. Letter to the editor, https://stm.sciencemag.org/content/3/96/96ra76/tab-e-letters.