Victor-Bogdan Popescu

Computational Network Analytics for Applications in Biomedicine

Åbo Akademi University

Victor-Bogdan Popescu

Born in 1992 in Târgoviște, Romania

Previous studies and degrees

Master in Electronic and Telecommunication Engineering
University "Politehnica" of Bucharest, Bucharest, Romania, 2016

Doctoral Dissertation Defense
Åbo Akademi University, Turku, Finland, 2022

Computational Network Analytics for Applications in Biomedicine

Victor-Bogdan Popescu

Computer Science
Faculty of Science and Engineering
Åbo Akademi University
Turku, Finland, 2022

Supervisors

Prof. Ion Petre
Department of Mathematics and Statistics
University of Turku
Finland

Dr. Eugen Czeizler
Faculty of Medicine
University of Helsinki
Finland

Prof. Jan Westerholm
Faculty of Science and Engineering
Åbo Akademi University
Finland

Reviewers

Prof. Alessandra Carbone
Laboratory of Computational and Quantitative Biology
Sorbonne University
France

Dr. Jun Pang
Department of Computer Science
University of Luxembourg
Luxembrug

Opponent

Prof. Andre Sanches Ribeiro
Faculty of Medicine and Health Technology
Tampere University
Finland

ISBN: 978-952-12-4204-5 (printed)
ISBN: 978-952-12-4205-2 (digital)
Painosalama, Turku, Finland 2022

Abstract

Network medicine has recently emerged as a field of research focusing on the analysis of networks modelling complex biological systems, for a better understanding of diseases and corresponding treatment. Building on results from graph theory, it provides a network-oriented approach for the identification of potential points of interest in these systems. Within this context, diseases can be regarded as systemic dysregulations in a patient's specific interaction network, while drug therapeutics represent the external interventions aiming to offset the effects of the disease. The disease data, which can include disease-drivers, typical genetic and functional dysregulations, or prospective drug and drug-target details, can be integrated into comprehensive networks that can help with the identification of targeted drugs and combinations thereof. There are multiple approaches to the study of these networks, such as through topological analysis or time-based dynamics. The recent availability of high-quality biological data and improvements in algorithmics and computational techniques reinforce the strong potential of the methods and their immediate applicability in the biomedical domain.

The first part of the thesis focuses on network controllability, which pertains to the ability to guide a network to a desired state through minimal external interventions through the identification of nodes of interest within. We provide a brief theoretical background for the structural controllability problem and several of its approaches, such as target- or input-constrained structural controllability, together with an overview of the existing algorithms and tools and followed by a short a discussion on the necessity of developing approachable software implementations for both existing and novel efficient algorithms. We prove that the target variant of the problem is hard to approximate and fixed-parameter tractable, and we introduce several algorithms aimed at solving it: an exhaustive search algorithm bounded by naturally constrained limits, and an approximation genetic algorithm which uses an algebraic approach. Moreover, we extend these algorithms to efficiently solve the input-constrained variant of the problem.

The second part of the thesis focuses on the applicability of the structural controllability approaches in biomedicine. We talk about the generation of

personalized protein-protein interaction network around disease-, patient-, or drug-specific proteins of interest, and their subsequent analysis, together with possible interpretations for the results. We apply this framework for the identification of potential repurposable drug suggestions for COVID-19 and breast, ovarian, and pancreatic cancer, as well as for the suggestion of personalized treatment for three multiple myeloma patients. Following these case studies and motivated by the lack of dedicated tools and the multitude of available biological databases, we introduce a novel, free, and open-source web-based platform allowing for the generation and structural controllability analysis of customized protein-protein interaction networks. This novel and, to the best of our knowledge, unique software integrates the controllability algorithms and multiple biological data sources and aims to enable direct, easy, and widespread usage of the presented methods.

This thesis focuses on developing a framework for the application of structural controllability in biomedicine. The work encompasses theory, data, algorithms, and their implementation for the identification of potential novel drugs and drug combination suggestions, drug-repurposing candidates, and treatment lines, moving forward towards personalized approaches to therapeutics.

Sammanfattning

Nätverks medicin är ett nytt forskningsområde vars fokus är analys av nätverk som motsvaras av komplexa biologiska system, vars syfte är att utvidga förståelsen av sjukdomar samt sjukdomarnas behandling. Grafteori ger ett nätverksorienterat tillvägagångssätt för identifiering av potentiellt viktiga intressepunkter i dessa system. I detta sammanhang kan sjukdomar anses vara systematiska dysregleringar i en patients specifika interaktions nätverk, medan läkemedelsterapier utgör externa åtgärder var målet är att motverka effekterna av sjukdomen. Sjukdomsdata, som kan omfatta sjukdoms drivande faktorer, typiska genetiska och funktionella dysregleringar eller detaljer om läkemedel och läkemedelsmål, kan integreras i omfattande nätverk som kan bidra till att identifiera riktade läkemedel och kombinationer av dessa. Det finns flera tillvägagångssätt för att studera dessa nätverk, t.ex. genom topologisk analys eller tidsbaserad dynamik. Tillgängligheten av biologisk data av hög kvalitet och förbättringar av algoritmer och beräkningstekniker förstärker metodernas potential och deras omedelbara tillämpbarhet inom det biomedicinska området.

Den första delen av avhandlingen handlar om kontrollerbarhet i nätverk, vilket avser förmågan att styra ett nätverk till ett önskat tillstånd med minimala externa ingrepp genom att definiera viktiga noder. Vi går kort igenom den teoretiska bakgrunden för strukturell kontrollerbarhet samt flera tillvägagångssätt för detta, t.ex. mål- eller ingångs begränsad strukturell kontrollerbarhet, tillsammans med en översikt av de befintliga algoritmerna och verktygen, följt av en kort diskussion om nödvändigheten av att utveckla lättförståeliga programvaru implementationer för både befintliga och nya effektiva algoritmer. Vi bevisar att mål varianten av problemet är svår att approximera och att den är möjlig att hantera med fasta parametrar, och vi introducerar flera algoritmer som kan lösa det: en sökalgoritm som begränsas av naturligt begränsande gränser, och en genetisk algoritm som använder en algebraisk metod för approximation. Dessutom utökar vi dessa algoritmer för att effektivt lösa den inmatnings begränsade varianten av problemet.

Den andra delen av avhandlingen fokuserar på tillämpning av strukturell kontrollerbarhet inom biomedicin. Vi går igenom genereringen av personalis-

erade protein-till-protein interaktionsnätverk kring sjukdoms- , patient- eller läkemedelsspecifika proteiner av intresse, och deras efterföljande analys, tillsammans med möjliga tolkningar av resultaten. Vi använder detta ramverk för identifiering av kandidater för läkemedels omvandling för COVID-19 och bröst-, äggstocks-, och bukspottkörtelcancer, samt för förslag till personlig behandling för tre patienter med multipelt myelom. Efter dessa fallstudier, motiverade av bristen på verktyg och den stora mängden av tillgängliga biologiska databaser, introducerar vi en ny, gratis plattform med öppen källkod, som möjliggör generering och analys av den strukturella kontrollerbarheten hos skräddarsydda protein-till-protein interaktionsnätverk. Denna nya och, såvitt vi vet, unika programvara integrerar algoritmer för kontrollerbarhet och flera biologiska datakällor och syftar till att möjliggöra en direkt, enkel och utbredd användning av de presenterade metoderna.

Avhandlingen fokuserar på att utveckla ett ramverk för applicering av strukturell kontrollerbarhet i biomedicin. Arbetet omfattar teori, data, algoritmer, och deras implementering för identifiering av potentiellt nya läkemedel och förslag av läkemedelskombinationer, kandidater för läkemedels omvandling och behandlingar, för att gå vidare mot en personaliserad strategi för terapi.

Acknowledgements

This thesis could not have been completed without the stars aligning and the invaluable support of many people. This is a shout-out to all of you that assisted me, in ways small and big, over the past years.

My supervisor, Ion Petre, is the one that would always come first to mind. Words can't describe how grateful I am for our countless talks and meetings, with advice and encouragement, and for the exceptionally prompt replies to any of my questions or doubts. I feel very lucky to have had such a mentor, who cared deeply about both me and my work. The feelings of gratitude extend just as much to my first co-supervisor, Eugen Czeizler, for all the support over the years with both academic and non-academic issues. I especially appreciated all of our idea-bouncing sessions, where I've always felt my opinion valued and considered. Last, but definitely not least, I'm very grateful to my second co-supervisor, Jan Westerholm, for agreeing to see me through and helping me over the last steps and hurdles. It is great to be looked after, just as much at the end as in the beginning.

I am sincerely thankful to the reviewers, Alessandra Carbone and Jun Pang, for their valuable insights and comments, and for their significant part in improving my thesis. Additional gratitude goes to the opponent, Andre Ribeiro, for their commitment and careful scrutiny of my work. Furthermore, all of the publications found within these pages have been a team effort. I would like to offer extensive thanks to my co-authors, Krishna Kanhaiya, José Ángel Sánchez-Martín, Nicoleta Siminea, and all others whose contributions made it all possible.

As concerns more worldly matters, I am grateful to Åbo Akademi and the Faculty of Science and Engineering for providing a safe, quiet, and cosy place to do my research, and to the Graduate School for funding my doctoral studies. I am also grateful to Sepinoud Azimi and Sebastien Lafond for their support, especially during the final months. In addition, there are a lot of people around the department who made the everyday life easier and better in so many ways. I particularly want to thank Minna Carla and Christel Engblom for always being kind and helpful with any number of practical things, and Karl Rönholm for all the laughs and technological support ever

since the beginning. Special thanks also go to Joel Sjöberg, for translating the abstract in Swedish. I can't as well forget to mention all of the others on our floor in Agora who have brightened many of my lunches, breaks, and outings, during the years here. In alphabetical order, Andrei, Frankie, Jesús, Krishna, Mikhail, and Usman, I am extremely grateful for all the talks, stories, and jokes, and for keeping me sociable and hopeful during the more difficult times.

Finally, I want to thank my girlfriend, my friends, and my family, near and far, for having my back, believing in me, and being ready to listen and to help at any hour of the day or the night. You know who you are, and you know I couldn't have done any of this without you!

Victor Popescu
Turku, April 2022

List of publications

1. Elio Nushi, **Victor-Bogdan Popescu**, Jose Angel Sanchez-Martin, Sergiu Ivanov, Eugen Czeizler, and Ion Petre. *Network modeling methods for precision medicine.* In Systems Biology Modelling and Analysis: Formal Bioinformatics Methods and Tools, ed. Elisabetta De Maria, Wiley (2022).

2. Eugen Czeizler, Alexandru Popa, **Victor-Bogdan Popescu**. *Fixed parameter algorithms and hardness of approximation results for the structural target controllability problem.* In Algorithms for Computational Biology, ed. Jesper Jansson, Carlos Martin-Vide Miguel and A. Vega-Rodriguez, Springer International Publishing (2018): 103-114.

3. **Victor-Bogdan Popescu**, Krishna Kanhaiya, Iulian Năstac, Eugen Czeizler, and Ion Petre. *Network controllability solutions for computational drug repurposing using genetic algorithms.* Scientific Reports, vol. 12, no. 1437 (2022).

4. **Victor-Bogdan Popescu**, Jose Angel Sanchez-Martinez, Daniela Schacherer, Sadra Safadoust, Negin Majidi, Andrei Andronescu, Alexandru Nedea, Diana Ion, Eduard Mititelu, Eugen Czeizler, and Ion Petre. *NetControl4BioMed: A web-based platform for controllability analysis of protein-protein interaction networks.* Bioinformatics, vol. 37, no. 21 (2021): 3976-3978.

5. Nicoleta Siminea, **Victor-Bogdan Popescu**, Jose Angel Sanchez-Martin, Ana-Maria Dobre, Daniela Florea, Georgiana Gavril, Corina Ițcuș, Krishna Kanhaiya, Octavian Pacioglu, Laura Ioana Popa, Romica Trandafir, Maria Iris Tușa, Manuela Sidoroff, Mihaela Păun, Eugen Czeizler, Andrei Păun, and Ion Petre. *Network controllability analysis for drug repurposing in COVID-19.* Briefings in Bioinformatics, vol. 23, no. 1 (2022).

Contents

Chapter 1

Introduction

Many biological systems can be represented using a network-based approach, from the signaling protein-protein interaction networks and metabolic networks at cell-level, up to neuronal networks or food-web networks. Applied network science within computational systems biology aims to help with the modelling and analysis of any such complex system with highly intricate relations involving multiple actors that are also linked and acting together as a whole. Network medicine has recently emerged as a field of research with a focus on the analysis of biological networks towards a better understanding in identifying and treating diseases [1]. This field aims to apply network modelling methods to metabolic pathway networks or protein-protein interaction networks towards diagnosing diseases and discovering novel treatments [2, 3]. Many of these network-based methods have seen recent successful applications and experimental validations, such as for the identification of mechanisms [4], effects [5], or potential of treatment [6] for specific medicine and diseases.

Network science consists of the study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena [7]. Thus, the understanding of network theory, the study of such representations from a mathematical point of view as part of graph theory, is central to the effective applications of these approaches to real-life cases. Within this context, a complex network corresponds to a graph that has one or more non-trivial topological features, that is, features that do not usually occur in regular or random graphs, but consistently do in the graph representations of real-life systems [8]. There are many ways in which the models of such complex networks can be analyzed, from a topological point of view (e.g., the structure of the network and the statistical analysis of its nodes and edges) to a more dynamical approach (e.g., taking into account possible changes in the network over time). Each modelling method can be employed in a wide variety of fields, with the field-specific details

and the research aims highly influencing the application of the method and the interpretation of the results. For example, the centrality measures of the nodes in a network can be used to identify congestion points in a road network [9], as well as for drug-discovery purposes in a drug network [10]. These more traditional network modelling methods have also been integrated with artificial intelligence and machine learning techniques to form the emerging area of graph representation learning. The applications can include node classification, relation prediction, or cluster detection, such as for recommending content to users in a social network [11], or identifying fraudulent activity in financial transaction networks [12]. A main focus of this field of research is further represented by its applicability in bioinformatics and biomedicine, for example, for the identification of protein functions [13], or the prediction of drug side-effects [14].

Another promising approach to the analysis of complex networks, which has seen renewed interest within the past years, is represented by the structural network controllability. Its goal is to identify suitable external interventions that would influence the state of a network in a desired way (that is, on an intuitive level, to "control" it). While the concept and the theoretical framework for the structural control of dynamical systems have been known for several decades [15, 16], they have recently re-emerged as a field of interest thanks to several developments in the approach and algorithmics of controlling complex networks [17]. Network controllability has seen recent applications in a number of different research areas [18], with a focus on biological networks [19], including protein-protein interaction networks [20], gene regulatory networks [21], and brain networks [22]. Moreover, these results have been confirmed by experimental validation in several studies, such as for the identification of neuron contribution in the locomotion of C.elegans [23], or for drug-repurposing in leukemia [24], breast cancer, and COVID-19 [25].

The complex systems that are encoded by such complex networks are often difficult to model because of the non-linear manner in which their components interact, and the behavior of the entire system cannot be directly inferred from the separate analysis of its individual interactions [26]. For example, protein-protein interactions represent contacts between two or more protein molecules as part of a biochemical event [27]. As a protein's function tends to be heavily regulated and influenced by the function of other components within a cell, the system could be modelled by a network where the nodes represent proteins and the edges the interactions between them for the particular cell. In biology, the development and use of algorithms that are modelling such systems are central to the field of computational systems biology [28]. Recent advances in biology, in the form of both experimental and computational methods, currently allow for the availability of high-

quality data, key to the modelling of complex systems. For example, the multitude of methods employed for the identification of all protein-protein interactions in the human interactome [29] lead to the recent availability of several corresponding databases [30].

The interactome represents the complete set of interactions within the cell [31], together with their underlying specialized subnetworks, such as the functional signaling pathways [32]. These signaling pathways through the protein-protein interactions in a cell trigger molecular events which can control the cell's growth or proliferation. Consequently, some perturbations can be linked with a number of diseases, such as cancer [33] or Alzheimer [34]. Thus, due to their essential role within the cellular processes, protein-protein interactions have recently been of high research interest, with a large number of methods aiming to identify and analyze them from various perspectives towards different goals [29]. Although a complete set is difficult to obtain and the thoroughness of the current methods can be debated [35, 36], the multitude of approaches led to a large availability of data, aggregated into several databases containing both experimentally obtained data [37], as well as computationally predicted data [38]. The size of the interactome is thought to correlate with the biological complexity of the corresponding organism, with estimates placing the human interactome at around 650,000 interactions [39]. This highlights the need for efficient and reliable algorithms for the modelling and analysis of such large data.

While the study of the protein-protein interactions is a prerequisite, further comprehension of the detailed role and effects of each protein and its encoding gene is required for a better understanding of the cell and its behavior, towards the application of the corresponding methods in biomedicine. This approach would allow for the model integration of data on specific genes of interest in a cell, such as disease- or patient-specific related data. Indeed, this area of research has seen great interest over the last years, with multiple studies dedicated to the identification of essential genes (that is, genes whose loss of function can lead to a compromise in the viability of the cell or organism) [40] or disease-specific essential genes (that is, genes whose loss of function can cause apoptosis in diseased cells, but not in healthy cells) [41], as well as mutated genes on a disease-specific level [42] and patient-specific level [43]. Additionally, the availability of detailed data on drugs and corresponding drug-target genes [44], together with data on the more recent treatment through targeted therapy [45], highlights a prospective different area of focus and has the potential to lead towards a personalized approach to biomedicine.

The research presented in this thesis focuses on developing a comprehensive framework for the application of network-based modelling in biomedicine, through the development of theoretically sound and efficient algorithms, their

implementation for easy and widespread usage, as well as illustrating examples of their real-life employment towards personalized medicine. Starting from the results on the control of linear dynamical systems in [15] and the more recent advances on network controllability [17] and target network controllability [46, 47], we extended the theoretic framework towards real-life application in biomedicine with two fixed-parameter tractable algorithms, performing an exhaustive search and constrained by natural bounds. Furthermore, we improved the greedy algorithm presented in [47] and we developed a novel genetic algorithm that can outperform the previously existing target network controllability algorithms. We also proved the feasibility of these algorithms by implementing corresponding command line-, desktop-, and web-based solutions that allow for the controllability analysis of any network. In addition, we aggregated and merged existing biological data from multiple databases on proteins, protein-protein interactions, disease-specific essential and mutated genes, and drugs with their drug-targets. We then used the software implementations and the data to illustrate the generation and analysis of customized protein-protein interaction networks for several multiple myeloma patients, and, separately, for COVID-19, towards the identification of interaction paths of interest and a personalized approach to treatment. Moreover, to facilitate the immediate applicability in biomedicine, we introduced a novel, free to use, and open-source web-based platform which encompasses the complete framework. Its novelty stems from the capability to both easily integrate readily available biological data for the generation of personalized protein-protein interaction networks, and to directly analyze any such networks using structural controllability. At the same time, the platform also provides integration with external databases and tools, allowing for easy importing and exporting of biological data and immediate evaluation of the results.

The thesis is structured as follows: in the second chapter we discuss the background of the network controllability framework, together with corresponding algorithms and implementation; in the third chapter, we shift towards the application of network controllability methods in biomedicine, together with the biological data required for, and the output of, such computational analyses; in the fourth chapter, we briefly present the summaries of the included research papers; and lastly, in the fifth chapter, we provide a short discussion on the methods described in the paper, including potential benefits and limitations, and we draw on future research directions.

Chapter 2

Structural network controllability

In this chapter, we discuss the structural controllability framework as a tool for the analysis of complex networks. We introduce the background, formal definition, and algorithms for the full structural network controllability, together with its extended versions of target structural network controllability and input-constrained target structural network controllability. Lastly, we compare briefly structural controllability to other methods for the modelling and control of networks, and we provide an overview of our contribution to the field.

2.1 Network modelling and controllability

Real-life systems are often complex, with many components that are closely interacting with each other and with the environment. Modelling the behavior of such a complex system is considerably difficult due to the non-trivial way in which its components are connected, such as non-linear connections or the presence of feedback loops [26]. In many cases, the system can be represented in a more simplified way as a network with the nodes representing its components, and the edges the interactions between them. These simplifications, however, are only approximations and can't always incorporate all details of inherently complex real-life phenomena. In turn, the modelled networks can be analyzed from a mathematical point of view using network and graph theory. Network modelling and analysis has seen applications in a wide variety of domains, including electrical engineering [48] and telecommunications [49], sociology [50], or biology [51].

An interesting approach to the analysis of complex networks refers to

the ability to control them, i.e., to drive them from any initial state to a specific final state of choosing through suitable external interventions and within finite time. This matches well the intuitive idea of control. Known as network controllability, this approach has been successfully employed in various applications within different research areas, such as power system networks [52], interbank networks [53], neural networks [23], or protein-protein interaction networks [25].

The network controllability problem aims to find the minimum external interventions required to control a network, i.e., drive each of its nodes to a desired state. The structural controllability approach concerns itself only with finding the nodes in the network on which external interventions need to be exerted in order to control it, and not on what or how these interventions should be [17]. This represents a shift towards a qualitative approach, where only the structure of the network matters (e.g., if a specific edge between two nodes exists or not), as opposed to a quantitative one (e.g., the actual weight of the edge is not important, as long as it is non-zero, thus it exists).

In many real-life scenarios, however, it is not necessary to control the entire network. Additionally, in [17] it is shown that, in order to achieve full control over a network, a high percentage of its nodes might have to be directly controlled through external interventions, which renders the approach unfeasible. Instead, it might be sufficient to achieve control over only specific nodes of interest in the network, called targets. Thus, target network controllability aims to find the minimum external interventions required to control a specific subset of target nodes in a network, i.e., drive each of them to a desired state [46]. Similarly, there exists a structural counterpart to this approach, which again focuses only on the structure of the network, i.e., on identifying the nodes in the network on which external interventions need to be exerted.

Furthermore, another often encountered real-life constraint concerns the ability to efficiently control through external interventions the nodes of the network. In many cases, only specific nodes are known to be directly controllable, which can render many obtained solutions difficult to be efficiently applied from a practical perspective. To increase the applicability of the approach in real-life, we can focus our search on these specific nodes in the network, called preferred inputs, leading to a double optimization strategy. Thus, input-constrained target network controllability aims to find the minimum external interventions required to control a specific subset of target nodes, while maximizing the use of preferred inputs, i.e., drive each of the targets to a desired state through paths that would ideally start in preferred inputs. The structural counterpart to this approach can be defined just as before.

On the other hand, a purely structural approach to the identification of

efficient control of complex networks might prove to be insufficient when additionally considering the network dynamics [54]. While structural methods consider all connections in a network as always equally and fully capable of contributing to control, in a dynamical system the state of these connections may change, and their redundancy becomes important. Therefore, effective control of such dynamical networks may require further domain knowledge or alternative techniques [55]. In turn, many real-life dynamical networks exhibit self-regulating capabilities that allows them to adapt when faced with internal or external perturbations [56]. This allows for a shift of focus back to the structure, as dynamics of the system would tend to return it to the default state. Indeed, when applied on complex networks in dynamical environments, structural controllability methods have been shown to both underestimate and overestimate the number of required interventions for control [54]. Thus, they should altogether provide a good approximation, especially for networks whose precise dynamics are too complex or largely unknown.

In this thesis and in the related original publications, we focus mainly on the study of target structural controllability (and its input-constrained variant) analysis of directed networks.

2.2 Structural controllability framework

A linear, time-invariant dynamical system of size n, where $A \in R^{n \times n}$ represents the time-invariant state transition matrix describing how each state affects the system and $x(t) = (x_1(t), \ldots, x_n(t))^T$ represents the state vector of the system at time t, can be defined by

$$\frac{dx(t)}{dt} = Ax(t). \tag{2.1}$$

If the system is influenced by an external source of size m, where $B \in R^{n \times m}$ is the time-invariant input matrix describing how the m inputs are affecting the n variables and $u(t) = (u_1(t), \ldots, u_m(t))^T$ represents the input vector at time t, then it becomes

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t). \tag{2.2}$$

Additionally, if the system exports an output of size l, where $C \in R^{l \times n}$ is the output matrix describing how the n variables are affecting the l outputs and $y(t) = (y_1(t), \ldots, y_l(t))^T$ represents the output vector at time t depending on $x(t)$, then it would be additionally described by

$$y(t) = Cx(t). \tag{2.3}$$

For ease of future reference notation, we will denote by X, I, and T the set of system variables, inputs, and outputs (or targets) respectively. Furthermore, such a system with $|X| = n$, $|I| = m$, and $|T| = l$, corresponding to $A \in R^{n \times n}$, $B \in R^{n \times m}$, and $C \in R^{l \times n}$ will be denoted as (A, B, C), or, in the particular case where $l = n$ and $X = T$, thus with the outputs the same as the system variables, the system will be denoted simply as (A, B).

A system (A, B, C) is said to be output controllable if there exists a suitable combination of inputs in I that can drive the system from any initial state to any particular desired state within finite time, for any numerical setup. We define the output controllability matrix of the system as

$$OC(A, B, C) = [CB, CAB, CA^2B, \ldots, CA^{n-1}B]. \tag{2.4}$$

As shown in [46], the system is output controllable if and only if

$$rank(OC(A, B, C)) = |T|. \tag{2.5}$$

As $OC(A, B, C) \in R^{l \times mn}$ and $l \leq n$, the equation 2.5 requires the $OC(A, B, C)$ matrix to be of full rank in order for the system to be output controllable.

In the particular case where $l = n$ and $X = T$, this corresponds to the Kalman's condition for full controllability [15]. In other words, the system (A, B) is controllable if and only if

$$rank([B, AB, A^2B, \ldots, A^{n-1}B]) = n. \tag{2.6}$$

As can be seen, this definition of control depends on the numerical values of the system variables. For a more efficient approach, we can decouple the numerical setup and focus instead on the internal structure of the system. Two equal-sized matrices $A \in R^{m \times n}$ and $A' \in R^{m \times n}$ are said to be structurally equivalent, which we denote by $A \sim A'$, if they have their zero values in the same positions, i.e., $A'_{i,j} = 0$ if and only if $A_{i,j} = 0$, for any $1 \leq i \leq m$ and $1 \leq j \leq n$. Similarly, two systems (A, B, C) and (A', B', C') are structurally equivalent if $A \sim A'$, $B \sim B'$, and $C \sim C'$.

Thus, a system (A, B, C) is said to be structurally output controllable if there exists a suitable combination of time-dependent inputs in I and a suitable numerical setup of the non-zero values in A, B, and C that can drive the system from any initial state to any particular desired state within finite time. Following up, the system is structurally output controllable if and only if there exists a combination of non-zero values for the non-zero entries in the system matrices such that the equation 2.5 holds. Furthermore, if a system is structurally controllable, then it is controllable in almost all numerical setups for its non-zero entries [57, 58].

A linear system can be represented as a directed weighted graph, by considering as nodes in the graph the n variables of the system, and as directed edges in the graph the non-zero entries in the system's state transition matrix, i.e., there exists a directed edge from the node representing the variable x_i to the node representing the variable x_j if and only if $A(x_j, x_i) \neq 0$. On the other hand, the m inputs of the system correspond to external input nodes, with the non-zero entries in the system's input matrix defining the directed edges between the input nodes and the nodes of the graph, i.e., there exists a directed edge from the node representing the input u_i to the node representing the variable x_j if and only if $B(x_j, u_i) \neq 0$. The nodes x_j for which there exists u_i, such that $B(x_j, u_i) \neq 0$, are directly controllable from the outside and can drive the system to the desired state. Similarly, the l outputs of the system correspond to external output nodes, with the non-zero entries in the system's output matrix defining the directed edges between the nodes of the graph and the output nodes, i.e., there exists a directed edge from the node representing the variable x_i to the node representing the output y_j if and only if $C(y_j, x_i) \neq 0$.

The system controllability problem has an equivalent network controllability problem formulation, where the variables, inputs, and outputs are nodes in a corresponding directed graph. Additionally, the structural network controllability problem has a counterpart formulation in terms of graphs [57]. With the same notations and constraints as before, the system (A, B) with n variables and m inputs is structurally controllable if and only if there exists a set of n directed paths from the input nodes to each of the nodes in the graph, such that no two paths intersect at the same distance from their ending nodes. In a similar way, for the system (A, B, C) with n variables, m inputs, and l outputs, the aim becomes finding a set of l directed paths from the input nodes to the output nodes, such that no two paths intersect at the same distance from their ending nodes. In this case, however, unlike for full structural controllability, the graph condition is necessary, but not sufficient [59].

An additional constraint can be further introduced by considering that the effect of the external control inputs is reduced in each time step. This corresponds to defining a maximum time step k beyond which the effect of an input within the system can be regarded as negligible. Thus, the output controllability matrix of the system $OC(A, B, C)$ will be reduced to its k-corresponding submatrix $[CB, CAB, CA^2B, \ldots, CA^kB]$. For the network controllability approach, this additional restriction corresponds to limiting to k the maximum length of any path between the input nodes and the nodes of the network, or the output nodes. All of the other previous considerations remain valid.

The set of external inputs describes the complexity and ability of an

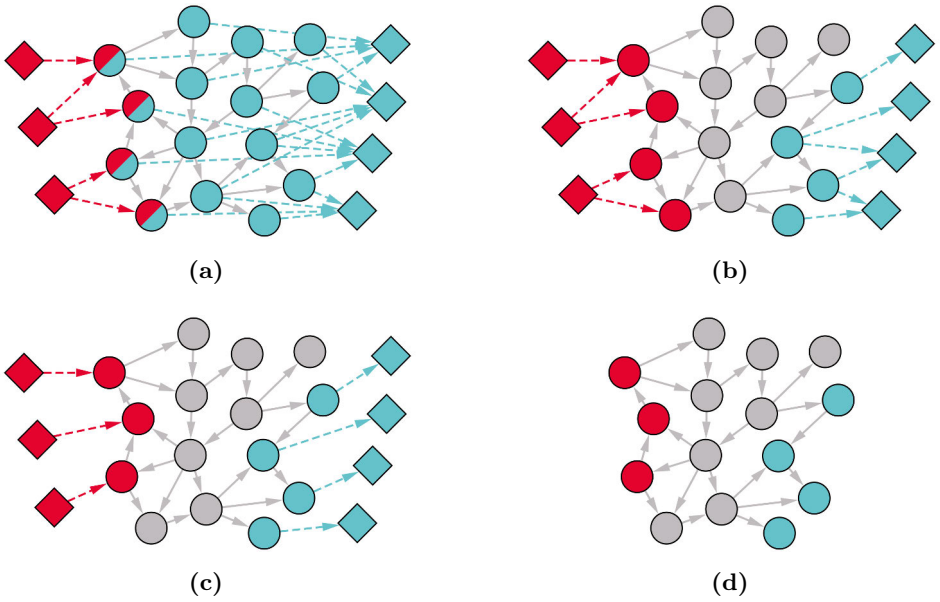


Figure 2.1: The different setups of the structural controllability problem for an example network with 15 nodes and 22 edges. **a**: full controllability, where the state of external outputs connected to all nodes in the network is influenced from external inputs; each external element can be connected to one or more nodes in the network. **b**: output controllability, where the state of external outputs connected to specific output nodes in the network is influenced from external inputs; each external element can be connected to one or more nodes in the network. **c**: simplified output controllability, where the state of external outputs connected to specific output nodes in the network is influenced from external inputs; each external element can be connected to exactly one node in the network. **d**: target controllability, where the state of specific target/output nodes in the network is influenced from control/input nodes in the network; the external elements are not considered in this approach and each input node can be directly influenced. **in red**: inputs and input nodes; **in blue**: outputs and output nodes; **in gray**: other notes. **as diamonds**: elements external to the network; **as circles**: elements internal to the network. **with dashed lines**: connections external to the network; **with solid lines**: connections internal to the network.

outside entity to influence several specific nodes in a graph. Thus, one such external input can be directly connected to multiple nodes in the graph, while each node in the graph might be directly connected to multiple external inputs (Figures 2.1a and 2.1b). Within our approach to structural controllability, we are considering only the directly controllable internal nodes in the graph, as opposed to the external inputs directly controlling them. In other words, our external inputs are connected to exactly one node in the graph and vice-versa (Figure 2.1c), such that each row and each column of the system's input matrix has exactly one non-zero value. Furthermore, if we assign the same unit value for all the edges between an input node and a node in the graph, then the input matrix B becomes a submatrix of the identity matrix I_n. This allows our approach to use only the internal elements of the graph (Figure 2.1d). The same considerations are also valid for the outputs and the output matrix C.

2.3 Full structural controllability

The full structural controllability problem, i.e., determining the minimum number of input nodes required to control an entire network, has been shown in [17] to be equivalent to identifying a maximum matching within a bipartite graph corresponding to the network. Specifically, we need to find a maximum set of edges that do not share starting or ending nodes, and a node is considered unmatched if there doesn't exist any edge ending in it [60]. Thus, the network can be fully controlled if and only if we can directly control each unmatched node and there exists a path to each matched node from an input node (Figures 2.2a and 2.2b). The maximum matching in a directed network can be found in at most $O(N^{1/2}E)$ steps, where N and E represent the number of nodes, respectively edges in the network [61], so the algorithm represents an efficient solution to the problem.

2.4 Target structural controllability

Unlike full structural controllability, the more general case of output structural controllability (also called target structural controllability) has been shown to be NP-hard [47]. Thus, we are interested in finding efficient approximation algorithms that could reach a solution as close to optimal as possible within a reasonable time.

Extending on the full structural controllability algorithm in [17], a first approach for solving the target structural controllability problem was described in [46]. The proposed approach follows the results of [59], aiming to solve the problem using a linking graph structure. The key concept is that a single node can control a set of target nodes in a network if there exists

a path of unique length to each of them, and we can thoroughly identify the sets of controlled target nodes for any given node in networks with no loops. This concept forms the basis for a greedy algorithm that approximates the minimum set of input nodes sufficient for target control by performing successive maximum matchings within a bipartite graph corresponding to the network. All nodes left unmatched during each of the matchings will need to be directly controlled (Figures 2.2a and 2.2c).

Following on this algorithm, it has been shown in [47] that the proposed method fails to provide a valid solution in specific special cases, and that an additional validation step is required, by checking if the controllability matrix $OC(A, B, C)$ is indeed of full rank, a shift back from graph theory. Furthermore, several heuristics have been proposed, significantly improving the algorithm run time and the size of its returned solutions. The improvements focus on optimizing the choice of maximum matchings, taking into account the structure of the control paths obtained up to that stage.

The target structural controllability problem has been the central part of the studies (Table 2.1). In Publication 2, we proved that the problem is fixed-parameter tractable by the number of target nodes and hard to approximate at a factor better than $O(\log n)$. We also provided an additional fixed-parameter algorithm for solving the problem through an exhaustive search bounded by real-life-derived limitations, such as restricting the maximum length of a control path. In Publication 3, we developed and implemented a new approximation genetic algorithm, focusing on an algebraic approach rather than graph theory, and managing to outperform the previous algorithms in a majority of cases. In Publication 4, we introduced a web-based application which allows for the structural target controllability analysis of any network. Additionally, in all of the original publications, we also demonstrated the applicability of the proposed methods and software on multiple real-life directed complex networks, with a focus on directed protein-protein interaction networks. This will be further discussed in Chapter 3.

2.5 Input-constrained target structural controllability

An extension to the target structural controllability and motivated by the real-life applicability of the method, the input-constrained version considers an additional input, as a supplementary set of input nodes that are already known to be controllable through external interventions. Several variants of the problem exist, such as identifying if a given subset of controllable input nodes can control the entire set of target nodes [67], or trying to maximize the number of known controllable inputs nodes among the identified controlling nodes for the set of target nodes [47].

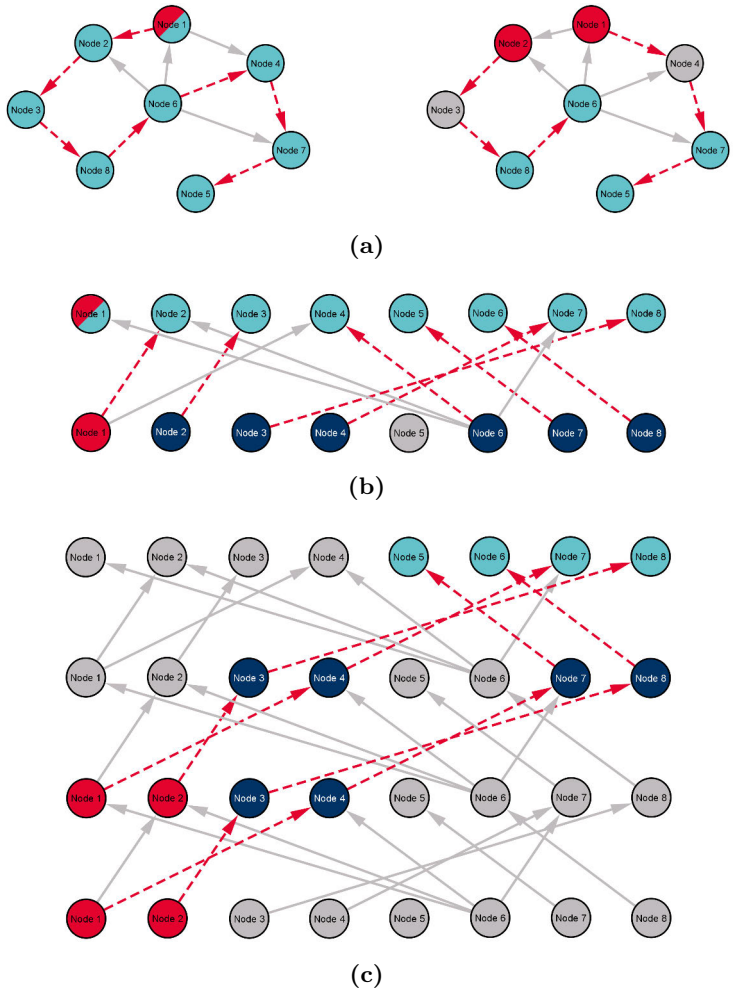


Figure 2.2: The maximum matchings required for the controllability analysis of an example network with 8 nodes, out of which 4 are target nodes, and 11 edges. **a**: the corresponding network and possible resulting control paths for the full controllability problem (left) and the target controllability problem (right). **b**: the maximum matching corresponding to the full controllability problem. **c**: the successive maximum matchings corresponding to the target controllability problem. **in red**: unmatched/control nodes; **in dark blue**: intermediary nodes; **in light blue**: target nodes; **in gray**: other nodes. **with dashed lines**: matched edges; **with solid lines**: other edges.

Type	Name	N	E	AD	T	P
Erdös-Rényi	Erdös-Rényi 100 [*]	62	47	0.76	5	-
(Synthetic)	Erdös-Rényi 500 [*]	497	1,270	2.56	25	-
	Erdös-Rényi 1000 [*]	1,000	4,952	4.95	50	-
	Erdös-Rényi 1500 [*]	1,500	11,258	7.51	75	-
	Erdös-Rényi 2000 [*]	2,000	19,869	9.93	100	-
	Erdös-Rényi 2500 [*]	2,500	30,976	12.39	125	-
	Erdös-Rényi 3000 [*]	3,000	44,713	14.90	150	-
Scale-Free	Scale-Free 100 [**]	100	202	2.02	5	-
(Synthetic)	Scale-Free 500 [**]	500	1,037	2.07	25	-
	Scale-Free 1000 [**]	1,000	2,112	2.11	50	-
	Scale-Free 1500 [**]	1,500	3,202	2.13	75	-
	Scale-Free 2000 [**]	2,000	4,353	2.18	100	-
	Scale-Free 2500 [**]	2,500	5,406	2.16	125	-
	Scale-Free 3000 [**]	3,000	6,491	2.16	150	-
Small World	Small World 100 [***]	100	400	4.00	5	-
(Synthetic)	Small World 500 [***]	500	2,000	4.00	25	-
	Small World 1000 [***]	1,000	4,000	4.00	50	-
	Small World 1500 [***]	1,500	6,000	4.00	75	-
	Small World 2000 [***]	2,000	8,000	4.00	100	-
	Small World 2500 [***]	2,500	10,000	4.00	125	-
	Small World 3000 [***]	3,000	12,000	4.00	150	-
Trust	Prison Inmates [62]	67	182	2.72	14	-
(Real)	College Students [62]	32	96	3.00	7	-
Electronic	Electronic Circuit 208 [63]	122	189	1.55	25	-
Circuit	Electronic Circuit 420 [63]	208	189	0.91	42	-
(Real)	Electronic Circuit 838 [63]	512	819	1.60	103	-
Protein-	Breast DEF [64]	1,415	2,435	1.72	112	123
Protein	Breast HCC1428 [65]	1,495	2,650	1.77	126	135
Interaction	Breast MDA-MB-361 [65]	1,478	2,590	1.75	124	136
(Real)	Ovarian DEF [64]	1,047	1,579	1.51	140	100
	Ovarian O1946 [65]	1,155	1,823	1.58	159	104
	Ovarian OVCA8 [65]	1,157	1,781	1.54	161	105
	Pancreatic AspC-1 [65]	1,022	1,534	1.50	125	90
	Pancreatic DEF [64]	991	1,484	1.50	168	86
	Pancreatic KP-3 [65]	1,134	1,757	1.55	167	94
	SIGNOR BrOvPa DEF [64]	2,913	6,729	2.31	145	201

Table 2.1: An overview of networks analyzed throughout the studies and the corresponding publications. **N**: number of nodes in the network. **E**: number of edges in the network. **AD**: average degree of the network. **T**: number of target nodes in the network. **P**: number of preferred input nodes in the network. *, **, ***: generated using *networkx* [66]; *: *fast_gnp_random* with $p = 0.005$; **: *scale_free_graph* with default parameters; ***: *watts_strogatz_graph* with $k = 4$ and $p = 0.2$.

Some of the heuristics proposed in the target structural controllability algorithm of [47] are designed towards reaching the latter goal, i.e., maximizing the number of returned controllable input nodes, by selecting them with a higher priority during the maximum matching stages of the algorithm. Overall, however, the algorithms offer no guarantee that any of these controllable input nodes will be indeed present in the solution, or that using them would lead to a better solution, i.e., with a lower number of controlling input nodes.

In each of the original publications mentioned in Section 2.4, in addition to the general target structural controllability, we also investigated the input-constrained variant (Table 2.1). The two algorithms proposed in Publication 2 can perform an exhaustive search to identify the minimum set of controllable input nodes that can control a maximum number of target nodes. The genetic algorithm proposed in Publication 3 also contains additional crossover and mutation functions that can maximize the number of controllable input nodes in a solution. Lastly, the application described in Publication 4 integrates the input-constrained variant of each of the implemented target structural controllability algorithms.

2.6 Alternative approaches to structural controllability

In addition to structural controllability, there are multiple other approaches to network modelling for identifying nodes of interest in a network. The identification process depends on the corresponding method, while the designated labelling of interest varies with the application field and the goal of the modelling.

A first simple approach to measuring the importance of a node in a network is to compute its centrality measures, where the aim is to find nodes that are of significance in the context of the network's topology. Multiple centrality measures have been defined, most of them taking into account the structural properties of the network, such as node connectivity. These measures can be grouped in several categories, such as:

- degree centralities, measuring the number of incident edges for a node in the network;
- proximity centralities, measuring how close a node is to other nodes in the network;
- path centralities, measuring the impact of a node in the paths that traverse the network;

- spectral centralities, measuring the algebraic properties of the adjacency matrix of the network.

Most of these centrality measures can be computed in linear or low-polynomial time in terms of the number of nodes and/or edges in the network [68, 69, 70].

Unlike the centrality methods, the controllability approaches represent an optimization problem, where the aim is to find a minimal set of nodes of significance in the network. Similar to structural controllability, where the solution to the problem represents a minimal set of nodes that can control all the nodes in the network, a solution of the minimum dominating set (MDS) approach is represented by a minimal set of nodes that can dominate all the nodes in the network, such that any node in the network is either in the dominating set, or adjacent to a node in it. A generalization of the problem, called minimum k-dominating set (MkDS), has the more general objective of finding a minimal set of nodes that can dominate all the nodes in the network at most k steps, such that any node in the network is either in the dominating set, or connected to a node in it through a path of length at most k. Another generalization, akin to the target structural controllability, is the red-blue (k-)dominating set problem, with the objective of finding a minimum subset of blue-labelled nodes in the network that can (k-)dominate all the red-labelled nodes in the network. It is easy to see that, within this context, the blue- and red-labelled nodes would correspond to the preferred control inputs, and respectively targets within the input-constrained target structural controllability approach. All of the variants of the minimum dominating set problem are NP-hard [71], but efficient algorithms are known to exist [72].

A slightly different, but well-established, controllability method considers the modelling of complex networks as Boolean networks and analyzing them as such. This model maps the nodes of the network to binary values (i.e., 0 and 1, or false and true), and assigns to the edges of the network Boolean functions that directly influence these values. Consequently, at each step of discrete time, the value of the nodes change based on the edge functions and the node values at the previous step [73]. There are many approaches to Boolean network modelling, such as considering the network topology (e.g., random, scale-free), or the paradigm of transitions between the states of the network (e.g., synchronous, asynchronous) and, despite their apparent simplicity, Boolean networks have been successfully used in a wide variety of biological areas, from cell aging [74] to identifying regulators within cancer networks [75]. Within this context and similar to structural controllability, controllability refers to the ability and steps required to bring a Boolean network from a starting state to a defined desired state with a known value, if possible. However, unlike structural controllability and minimum dominating

sets, where the focus was on the structure of the network and the actual states of the nodes were not taken into consideration, at any moment of time the nodes of the Boolean network have a defined state, which altogether give the state of the entire network and allow for only a finite number of potential states. Thus, over time, without external perturbation, a network will always reach a periodic sequence of states, called attractor. This makes it is possible that some states can never be reached when starting from a specific state. The controllability of Boolean networks has also been shown to be NP-hard and algorithms with an exponential complexity are commonly used to solve it [76], which usually limits their applicability to much smaller networks when compared to the other approaches [77].

Another interesting method, closely related to (node) structural controllability, is represented by the edge structural controllability. Within this setup, denoted as switchboard dynamics, an edge can be directly influenced by its source node, while a node is considered as simply mapping its incoming edges to its outgoing edges, according to an internal switching logic. The goal becomes twofold, in gaining control over the edges of the network using a set of minimal control edges, which can then be subsequently influenced through a set of minimal control input nodes. This type of control can be investigated using the node structural controllability algorithms applied on a slightly modified network, whose nodes represent the edges in the initial network, and whose edges represent the paths of length two in the initial network [78]. Thus, all considerations regarding the complexity and efficiency of the previous algorithms apply for edge structural controllability as well, as the network transformation can be performed in linear time. This method could be better suited for the modelling of natural systems where dynamical processes take place on edges, such as data communication networks or power line networks [79]. At the same time, a target-oriented approach has also been proposed with the similar aim of gaining control over only specific edges in the network, as opposed to all of the edges [80], and efficient algorithms have been suggested [81].

A common criticism of the more traditional controllability methods refers to their assumption of linear dynamics, while many real-life systems are thought to be mainly non-linear [82]. Consequently, several approaches have been proposed for the modelling and controllability analysis of non-linear complex networks, through establishing the accessibility of their desired states [83]. These methods would generally require additional knowledge regarding the network dynamics or the state space. For example, using the network structure for the identification and measurement of specific feedback vertex sets allows for the discovery of any recurrent dynamical behavior in such a network [84]. Another similar approach considers both the structural and the functional network information to determine select components

whose activity can be influenced in attempting to drive it towards a desired state or away from an undesired state [85]. Alternatively, individual nodes and valid corresponding compensatory perturbations can be identified in order to bring the entire network to a different desired state [56].

Several of these network modelling methods have been more thoroughly presented and compared in Publication 1. We surveyed multiple centrality methods, i.e., degree centrality, closeness centrality, eccentricity centrality, betweenness centrality, and eigenvector-based prestige, and provided an overview of the theoretical background for each analyzed modelling method and for two controllability methods, i.e., minimum dominating sets, and structural controllability.

Chapter 3

Applications in biomedicine

In this chapter, we discuss the potential applicability of the network controllability methods described in Chapter 2 on protein-protein interaction networks. We consider the background, structure, and significance of protein-protein interaction network, and briefly mention several available related databases and data repositories. We then shift towards a more personalized approach to the generation and analysis of these networks, by considering additional disease- or patient-specific proteins of interest. Lastly, we discuss the significance and applicability of the results for the structural controllability analysis of these networks, and we provide an overview of our contribution to the field.

3.1 Protein-protein interaction networks

Protein-protein interactions represent physical contacts between two or more protein molecules as a result of biochemical events and serve specific biological functions. The study of these interactions usually occurs in the context of a pathway or a cell, rather than individually [86]. Additionally, the interactome represents the entire set of interactions in a specific cell [31], and it can be modelled by a protein-protein interaction network, as a mathematical representation of the physical contacts between the proteins of the cell. The analysis of these networks helps with the understanding of the biological mechanisms behind, including the potential causes of specific diseases. This knowledge can lead to prospective efficient diagnostics and therapeutics [87].

Many methods, both experimental and computational, exist for the identification and analysis of protein-protein interactions, to different degrees of accuracy [29]. Furthermore, the interactome might also vary, for example with specific diseases or with different stages of development, which impacts the thoroughness and completeness of the obtained data [88]. It is estimated

that its actual size correlates with the biological complexity of the corresponding organism, with the human interactome being placed at around 650,000 interactions [39].

The study of protein-protein interaction networks is closely linked to the understanding of diseases. For example, network properties can be used to separate specific disease and non-disease proteins [89], or to infer the function of proteins within pathways [90]. Such differences between healthy and diseased cells can be brought forth by the reconstruction of the networks around the proteins corresponding to the mutated genes involved in the disease [91], taking into account the perturbation of the network (e.g., removal of proteins or interactions) [86] possibly caused by mutations of the proteins and their encoding genes [92].

In this thesis and in the related original publications, we only considered directed interactions between exactly two proteins (e.g., inhibition or stimulation). Additionally, for a more thorough analysis, it is also possible to include several typically undirected interactions by trying to map a direction [93], or by taking into account both directions. As mentioned before, the structural target controllability approach can circumvent these limitations, as the missing or incomplete data does not affect the validity of results, but merely their optimality.

3.2 Proteins encoded by disease-specific essential genes

Survivability-essential genes are genes indispensable for a cell to grow and reproduce. Knocking out an essential gene can lead to the death of the cell or a block in its division. However, many genes can be considered as essential only under specific circumstances [94]. Consequently, this leads to the identification of several disease-specific genes that have been found to be survivability-essential for diseased cells, but not for healthy cells, for a particular disease. Such disease-specific essential genes have been of great interest in the study and identification of novel therapeutic targets. For example, the suppression of a protein encoded by a cancer-specific survivability-essential gene could lead to the death of only the cancer cells [41]. The proteins encoded by the essential genes are well connected within the protein-protein interaction networks [95], so they can be integrated in their study (Figure 3.1).

In addition, many diseases are caused by gene mutations, or irregularities in the functionality of the healthy cells [92]. The genes suffering mutations can cause the disease onset but can also be altered by the disease itself. For example, besides the gene mutations that are considered as high risk for breast cancer, studies have shown that a wide number of other genes are

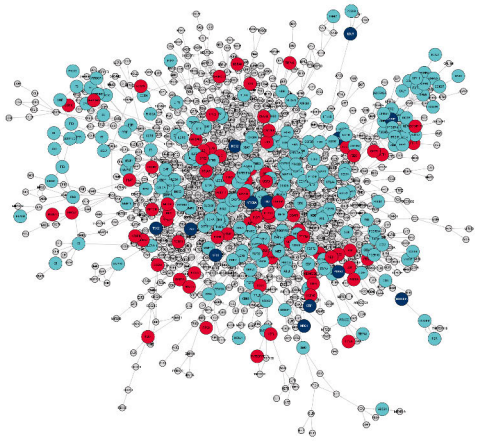


Figure 3.1: The influence of the breast-cancer cell-line HCC1187 in a protein-protein interaction network with 1,179 proteins and 2,098 interactions. The network contains 1 connected component, has a diameter of 17 and an average degree of 3.45. There are 298 disease- and drug-specific proteins of interest highlighted, consisting of 75 proteins encoded by disease-specific survivability-essential genes, 13 proteins encoded by commonly mutated genes within the disease, and 210 drug-targetable proteins. **in red**: proteins encoded by disease-specific survivability-essential genes; **in dark blue**: proteins encoded by commonly mutated genes within the disease; **in light blue**: drug-targetable proteins; **in gray**: other proteins.

also often mutated in the general population [96]. Thus, the inclusion of the proteins encoded by such mutated genes can also help shift the focus towards a more personalized and patient-oriented approach, in addition to the more generic disease-specific one (Figure 3.2).

Within our structural network controllability framework, we use the proteins encoded by the disease-specific survivability-essential genes as control targets within the network. In this context, control over a protein indicates the ability to influence it in a desired way. Thus, by considering the proteins in the network that control a high number of target proteins, we can pave the way towards the identification of potential therapeutic suggestions. In addition, proteins encoded by disease-specific mutated genes can also be included in the target set, as well as by the patient-specific ones, for a more personalized approach.

3.3 Proteins targetable by drugs

Almost all pharmaceutical drugs available on the market are targeting proteins in the body, as they can have lower toxicity and higher specificity than drugs targeting other macromolecules [97]. Proteins that can be suitable

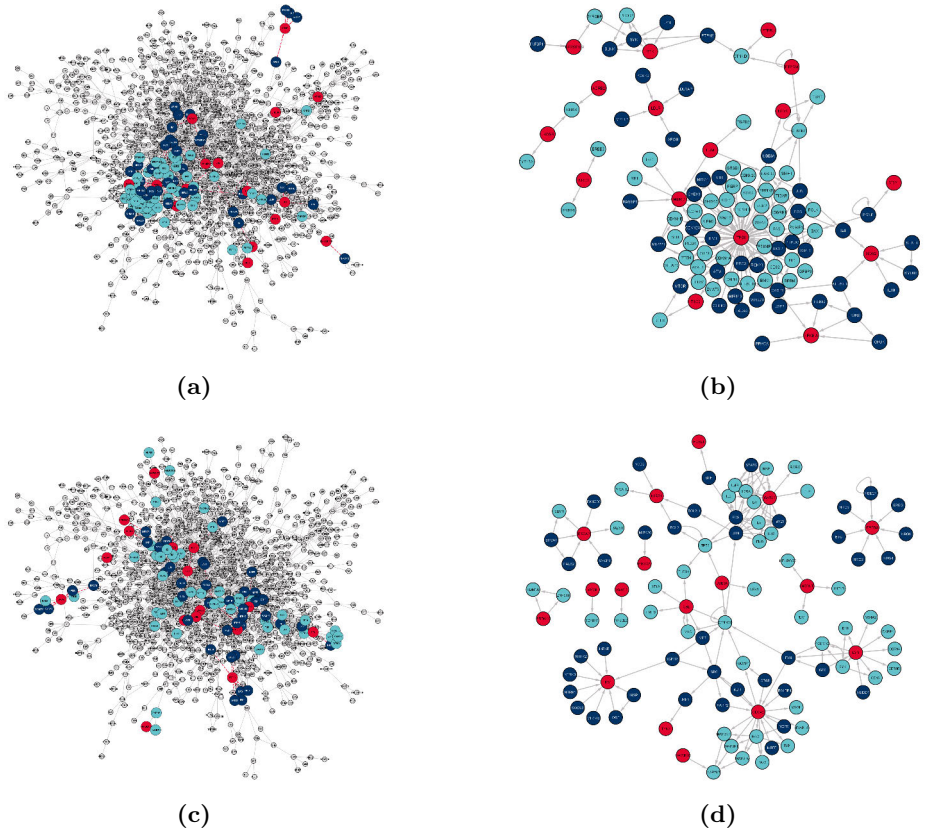


Figure 3.2: The impact of the patient-oriented approach in the generation and analysis of the breast-cancer cell-line HCC1187 protein-protein interaction network with 1,179 proteins and 2,098 interactions. **a**: the complete network, with the patient-specific proteins corresponding to the first patient highlighted. **b**: the patient-specific subnetwork corresponding to the first patient, with 110 proteins and 160 interactions; the network contains 16 proteins encoded by patient-specific mutated genes, 42 proteins that can directly affect them, and 52 proteins that can be directly affected by them. **c**: the complete network, with the patient-specific proteins corresponding to the second patient highlighted. **d**: The patient-specific subnetwork corresponding to the second patient, with 103 proteins and 161 interactions; the network contains 17 proteins encoded by patient-specific mutated genes, 41 proteins that can directly affect them, and 45 proteins that can be directly affected by them. **in red**: proteins encoded by patient-specific mutated genes; **in dark blue**: proteins that can directly affect the proteins encoded by patient-specific mutated genes; **in light blue**: proteins that can be directly affected by the proteins encoded by patient-specific mutated genes; **in gray**: other proteins.

drug-targets would ideally have a critical role in the disease proliferation, while being less involved in other processes of the cell, to limit the potential side-effects [98]. However, relatively few drug candidates complete successfully the required clinical trials, mainly because of lack of efficiency or increased toxicity [99]. Consequently, the network pharmacology approach has been developed to help with the identification of potentially more suitable drug-targets, through the understanding of their precise activity within the diseased interactome [100].

Within our structural network controllability framework, we use the proteins encoded by the drug-targetable genes as preferred control inputs within the network. Such proteins can be considered as prime candidates for controlling the target proteins, thus for providing suggestions for more focused potential therapeutics, as they can already be influenced through external interventions by already existing drugs. This can include, also depending on the used target data, drug repurposing, in the case of drugs unrelated to the studied disease, or personalized therapeutics, in the case of multiple available lines of treatment. It is worth mentioning that, due to the nature of the framework, we can only take into consideration targeted drugs with known drug-targets.

3.4 Available data and databases

Multiple database resources exist including data on proteins, protein-protein interactions, survivability-essential genes, mutated genes, or drugs and drug-targets. Several such databases will be briefly presented next. The list does not aim to be exhaustive, and only mentions the databases which have been used throughout the original publications. Excluding the one indicated exception, all of the other databases are freely and publicly available. However, generally speaking, each database is considered to be a separate entity with specific fields and unique identifiers, and parsing the data is required in order to match their content.

We have used the protein public data from the HGNC [101], Ensembl [102], UniProt [103], NCBI [104], and InnateDB [37] databases, together with the proprietary data from the KEGG [105] database. Each of these databases contains a primary unique identifier and one or more foreign identifiers of other databases, which we have used to match their data. We have compiled a final set of 42,152 proteins to which we have assigned an own additional unique identifier.

We have used the interaction public data from the OmniPath [106], InnateDB [37], and SIGNOR [107] databases, together with the proprietary data from the KEGG [105] database. Each of these databases contains specific protein identifiers that we have matched with the previously obtained

set of proteins. Furthermore, we have additionally filtered the data to keep only the interactions that appear as directed and experimentally validated in at least one database. We have compiled a final set of 48,189 interactions to which we have assigned an own additional unique identifier.

We used cancer cell-line-specific survivability-essential gene data from the COLT [65] database, together with cancer cell-line-specific mutated gene data from the DepMap database [108]. Each of these databases contains lists of specific protein identifiers that we matched with the previously obtained set of proteins. We have compiled 52 sets of proteins encoded by survivability-essential genes and 1,526 sets of proteins encoded by mutated genes for multiple cancer cell-lines.

We used drug and drug-target data from the DrugBank [44] database. The database contains lists of protein identifiers that we matched with the previously obtained set of proteins. We have compiled 9 sets of drugs and their corresponding drug-targetable proteins.

All of the compiled data is integrated and readily available in the application described in Publication 4. The data can be easily accessed, inspected, and downloaded, or it can be directly used for custom protein-protein network generation and structural controllability analysis.

3.5 Generation of personalized interaction networks

The human interactome is estimated to contain around 650,000 interactions [39]. Many diseases, however, such as Alzheimer or Parkinson, are considered to be caused by the aberrant behavior of only specific genes [109]. Switching the focus on only their encoded proteins and their immediate surrounding area, and thus on a smaller part of the complete interactome, allows for a more efficient applicability of available methods and algorithms, and to more targeted and specific results.

Therefore, the generation of a personalized protein-protein interaction network around a set of proteins of interest presents an interesting challenge, consisting of identifying a corresponding subnetwork within the complete interaction network. Consequently, access to complete and accurate interaction data represents a prerequisite for the network generation, which represents one of the intended uses of the compiled data that was presented in Section 3.4. It is again worth noting that while the availability of more thorough data would theoretically improve the fidelity of the generated networks and the accuracy of their analysis, the structural target controllability approach provides valid (although potentially less optimal) results even when presented with incomplete data.

The networks can be generated around any set of proteins of interests,

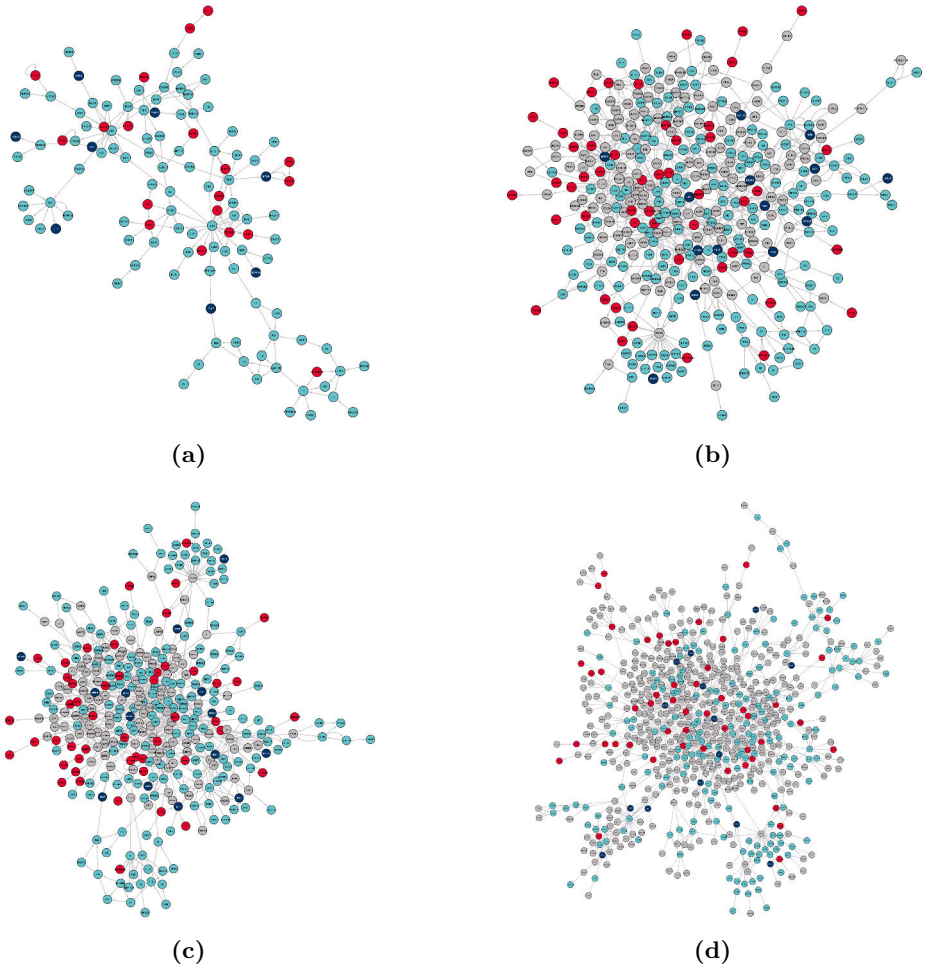


Figure 3.3: The influence of the generation algorithms on the networks generated starting from the same breast-cancer cell-line HCC1187 protein-protein interaction network with 1,179 proteins and 2,098 interactions and the same set of 259 seed proteins. **a**: the network with 121 proteins and 165 interactions generated using the gap algorithm with no intermediary proteins. **b**: the network with 392 proteins and 780 interactions generated using the gap algorithm with 2 intermediary proteins. **c**: the network with 356 proteins and 732 interactions generated using the upstream/downstream algorithm with 2 intermediary proteins. **d**: the network with 651 proteins and 968 interactions generated using the neighbors algorithm. **in red**: survivability-essential seed proteins; **in dark blue**: survivability-essential drug-targetable seed proteins; **in light blue**: drug-targetable seed proteins; **in gray**: other proteins.

depending on the disease, patient, or the goal of the analysis. Typically, this can include any combination of proteins encoded by disease-specific survivability-essential genes, proteins encoded by disease- or patient-specific mutated genes, drug-targetable proteins, or any other protein deemed important to the current study. We denote such proteins as seed proteins for the network generation.

These seed proteins can be then looked up in one or more interaction databases, which can be further filtered or combined depending on the problem at hand. The subset of actual interactions which will appear in the new network vary with the generation algorithm (Figure 3.3). Throughout the original publications mentioned in Section 2.4, we have implemented and used several generation algorithms:

- *neighbors*, which considers all of the interactions containing at least one seed protein;
- *gap*, which considers all of the interactions between seed proteins with a specific number of intermediary proteins;
- *upstream/downstream*, which uses two different sets of seed proteins, one denoted "upstream" and one "downstream" and considers all of the interactions from the upstream proteins to the downstream proteins, with a specific number of intermediary proteins.

The generation of custom protein-protein interaction networks has been a central part of the studies (Table 3.1). In Publication 1, we used the gap algorithm for the generation of three patient-personalized multiple myeloma interaction networks, and, in Publication 5, we used the upstream/downstream algorithm for the generation of several COVID-19-related interaction networks. Furthermore, the application described in Publication 4 allows for the generation of personalized protein-protein interaction networks using the neighbors and the gap algorithms; the generated networks can then be downloaded and imported in external software.

3.6 Structural controllability analysis of interaction networks

The structural controllability framework presented in Section 2.2 can be used for the analysis of protein-protein interaction networks towards potential novel drug information discovery, based on the type of the data provided and the method of application. The immediate result of such an analysis would consist of a list of proteins able to control the entire set of target proteins of interest, as well as the control paths between them (Figure 3.4). The

Disease		#	ST	N	AD	T
		A	SDT	E		DT
COVID-19	[110]	36	175.00	2,235.88	8.49	56.61
		U/D 2	1,529.00	19,642.58		974.83
Breast Cancer	[65]	3	527.00	1,462.66	1.75	120.66
		Gap 1	1,506.00	2,558.33		131.33
Ovarian Cancer	[65]	3	535.33	1,119.66	1.54	153.33
		Gap 1	1,506.00	1,727.66		103.00
Pancreatic Cancer	[65]	3	545.00	1,049.00	1.51	153.33
		Gap 1	1,506.00	1,591.66		90.00
Glioblastoma	[111]	137	881.37	2,779.48	23.03	319.56
		U/D 1	375.00	64,579.12		429.76
Multiple Myeloma	[112]	3	70.00	440.33	3.93	23.00
		Gap 2	27.00	1,724.33		43.33

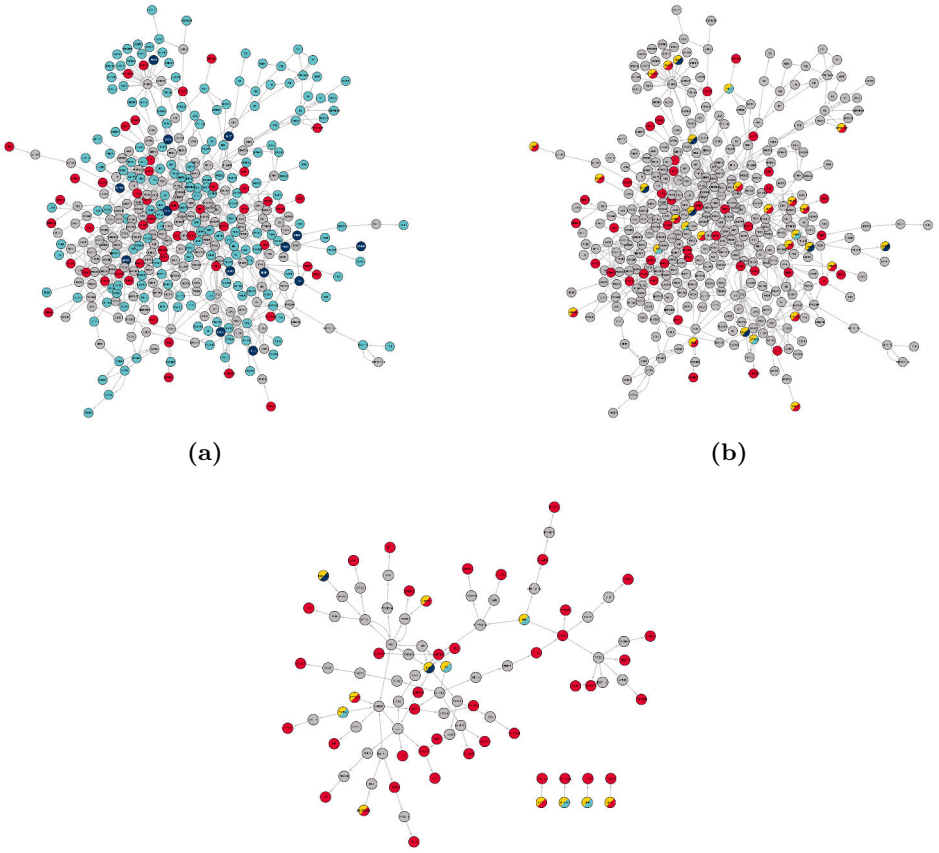
Table 3.1: An overview of disease- and patient-specific networks generated and analyzed throughout the studies and the corresponding publications. The values are averaged over all the generated networks for each disease. **#**: the number of generated networks; **A**: the main algorithm used for generating the networks; **ST**: the number of seed target proteins used for generating the networks; **SDT**: the number of seed drug-targetable proteins used for generating the networks; **N**: the number of proteins in the generated networks; **E**: the number of interactions in the generated networks; **AD**: the average degree of the generated networks; **T**: the number of target proteins in the generated networks; **DT**: the number of drug-targetable proteins in the generated networks.

interpretation of these results can vary depending on the current study, such as for the discovery of novel drug targets or drug repurposing for a specific disease, or the identification of personalized lines of treatment tailored to a specific patient.

As briefly mentioned, the first result returned by the structural controllability analysis of protein-protein interaction networks consists of one or more sets of proteins able to control the entire set of target proteins. Each such controlling protein is associated with one or more target proteins that it can control, and can be ranked based on this number. Depending on the input data and on the end goal, these controlling proteins can lead to different promising potential outputs, such as:

- novel drug-target suggestions for the studied disease, in the case of the non- and input-constrained approaches, corresponding to the top controlling proteins that are not known to be drug-targetable;

- drug-repurposing suggestions for the studied disease, in the case of the input-constrained approach, corresponding to the top controlling proteins that are known to be drug-targetable;



(a) **(b)** **(c)**

Figure 3.4: The general setup for the structural controllability analysis of the breast-cancer cell-line HCC1187-based protein-protein interaction network with 392 proteins and 780 interactions generated using the gap algorithm with 2 intermediary proteins. **a**: The analyzed network. **b**: The results of the analysis. **c**: The returned control paths of non-zero length. **in red**: target proteins; **in light blue**: drug-targetable proteins; **in dark blue**: target drug-targetable proteins; **in yellow**: controlling proteins; **in gray**: other proteins.

- drug combination suggestions for the studied disease, in the case of the input-constrained approach, corresponding to combinations of two or more top controlling proteins that together control a large part of the interaction network;

- personalized therapeutics suggestions for the studied disease and patient, in the case of the non- and input-constrained approaches, corresponding to the top controlling protein that are known to be drug-targetable by a specific line of treatment.

An interesting aspect of the controllability analysis, however, is that it can also return the potential interaction pathways from the controlling or drug-targetable proteins to the target proteins, which allows for a better understanding of the disease and the potential drug mechanisms. As an additional step towards a greater applicability and due to the inherent dissipation of a drug's input and to limit its side effects, the analysis can also be restricted to considering only shorter pathways.

The application of the structural controllability framework on protein-protein interaction networks and the analysis of its results has been a central part of the studies. We analyzed three multiple myeloma interaction networks generated in Publication 1 and we compared the treatment lines returned by the structural controllability analysis with those of the topological and minimum dominating sets analyses. Additionally, we applied the novel genetic algorithm developed in Publication 3 on several real-world and random networks and we performed a brief literature validation of the drug suggestion results obtained for breast, ovarian, and pancreatic cancer interaction networks. Lastly, in Publication 5, we applied the structural controllability analysis on two COVID-19 interaction networks, suggesting several novel drug and drug combination treatment approaches and performing a thorough literature validation of the results. Furthermore, the application described in Publication 4 allows for the structural controllability analysis of any network, with a focus on protein-protein interaction networks; the analysis results can then be downloaded and imported in external software.

Chapter 4

Summaries of the included publications

In this chapter, we present the summaries of the publications included in this thesis.

4.1 Publication 1: Network modeling methods for precision medicine

Elio Nushi, Victor-Bogdan Popescu, Jose Angel Sanchez-Martin, Sergiu Ivanov, Eugen Czeizler, and Ion Petre. In Systems Biology Modelling and Analysis: Formal Bioinformatics Methods and Tools, ed. Elisabetta De Maria, Wiley (2022).

In this publication, we survey multiple network modelling methods, and we study their applicability to personalized medicine. We start by assessing several network centrality measures, such as degree-, closeness-, eccentricity-, and betweenness-centrality, together with eigenvector-based prestige. For each measure, we provide an intuitive description, the formal definition and theoretical background, and a brief presentation of corresponding algorithms and their time complexities. We also investigate two system controllability methods, minimum dominating sets and structural controllability. We provide the formal definitions and an overview of the corresponding algorithms, and we detail several variants and generalizations for each of them. Further, we briefly introduce several software tools that can be used for the generation, visualization, and analysis of networks, and for the reproducibility of the publication's results. We then demonstrate the applicability of the presented methods on the protein-protein interaction networks of three multiple myeloma patients, built around the genes of interest specific to each

patient. For each network, we show how the methods can be used for the identification of personalized combinatorial drug treatment.

4.2 Publication 2: Fixed parameter algorithms and hardness of approximation results for the structural target controllability problem

Eugen Czeizler, Alexandru Popa, Victor-Bogdan Popescu. In Algorithms for Computational Biology, ed. Jesper Jansson, Carlos Martin-Vide Miguel and A. Vega-Rodriguez, Springer International Publishing (2018): 103-114.

In this publication, we investigate the target structural controllability problem, and we introduce several constraints that would improve its relevance for the exhaustive analysis of real-life networks. We begin by showing that the target structural controllability problem is fixed-parameter tractable by the number of target nodes. Then, motivated by the applicability in the control of protein interaction networks in cancer, we introduce an even lower-complexity fixed-parameter algorithm depending on an additional parameter generally bounded by much lower limits, the maximum allowed length of a control path. Lastly, we prove that the problem is hard to approximate at a factor better than $O(\log n)$.

4.3 Publication 3: Network controllability solutions for computational drug repurposing using genetic algorithms

Victor-Bogdan Popescu, Krishna Kanhaiya, Iulian Năstac, Eugen Czeizler, and Ion Petre. Scientific Reports, vol. 12, no. 1437 (2022).

In this publication, we propose a novel method for solving the target structural controllability problem and its input-constrained variant using a genetic algorithm. The algorithm is based on the algebraic theory behind the problem, rather than the graph theory that the currently existing methods employ. We also introduce multiple approaches for each of the genetic operators (e.g., crossover or mutation), for better applicability in different domains. We apply the algorithm on multiple random networks of varying size generated according to the Erdős-Rényi, the scale-free, and the small world models, and we compare the results against those of similar algorithms.

The comparison considers the number of returned solutions, the number of control nodes in a solution, the average length of the control paths, and the elapsed time. Additionally, we investigate the use in computational drug repurposing by applying the algorithms on several cancer protein interaction networks and performing a literature review of the resulting drug-targets and drugs. Overall, the proposed genetic algorithm is proven to be consistently better for the identification of relevant drug-targets.

4.4 Publication 4: NetControl4BioMed: A web-based platform for controllability analysis of protein-protein interaction networks

Victor-Bogdan Popescu, Jose Angel Sanchez-Martinez, Daniela Schacherer, Sadra Safadoust, Negin Majidi, Andrei Andronescu, Alexandru Nedea, Diana Ion, Eduard Mititelu, Eugen Czeizler, and Ion Petre. Bioinformatics, vol. 37, no. 21 (2021): 3976-3978.

In this publication, we introduce a novel free open-source web-based application for the generation and structural controllability analysis of protein-protein interaction networks. Specifically, the application allows users to upload, import, or generate personalized protein-protein interaction networks, and then to analyze them using different algorithms and in various controllability setups. These analyses can also be tailored to focus on existing available drugs and their drug-targets, providing potential therapeutic suggestions. The software displays a modern interface and includes the possibility to store the created networks and analyses under individual user accounts, enabling data sharing and easier collaboration. Additionally, the application also provides considerable already-compiled and ready-to-use biological data on proteins from the HGNC, Ensembl, UniProt, NCBI, and InnateDB databases, on protein-protein interactions from InnateDB, OmniPath, and SIGNOR databases, on cancer cell-lines from COLT and DepMap, and on drug-targets and drugs from DrugBank.

4.5 Publication 5: Network controllability analysis for drug repurposing in COVID-19

Nicoleta Siminea, Victor-Bogdan Popescu, Jose Angel Sanchez-Martin, Ana-Maria Dobre, Daniela Florea, Georgiana Gavril, Corina Iţcuş, Krishna Kanhaiya, Octavian Pacioglu, Laura Ioana Popa, Romica Trandafir, Maria

Iris Tușa, Manuela Sidoroff, Mihaela Păun, Eugen Czeizler, Andrei Păun, and Ion Petre. Briefings in Bioinformatics, vol. 23, no. 1 (2022).

In this publication, we apply the structural controllability methods for the identification of possible drugs and combination of drugs for the potential treatment of COVID-19. We begin by generating extensive protein-protein interaction networks that include drug targets and host factors for the SARS-CoV-2 infection at both low and high multiplicity of infection. We then perform structural controllability analysis of these networks and determine sets of drug targets that are repeatedly reported as exerting control over the host factors. Finally, we validate these results against existing literature and ongoing clinical studies and show that our approach can provide novel insights into the mechanisms and potential therapeutics for COVID-19.

Chapter 5

Conclusions and future work

The focus of this thesis is the development and real-life application of a structural network controllability framework. The framework encompasses the generation of customized networks and their structural controllability analysis towards the identification of nodes of interest with the potential to influence large parts of the networks. The presented research includes the theory and algorithms, their software implementations, and the available data required for the immediate applicability in biomedicine, with the aim of discovering novel drugs or drug-repurposing suggestions, and suitable personalized therapeutics.

In the first part of the research, we focused on a few theoretical aspects of structural controllability. We proved that the structural controllability problem is fixed-parameter tractable and hard to approximate at a factor better than $O(\log n)$, and we identified several parameters constrained by lower bounds within general real-life usage. Building on these results, we proposed an exhaustive-search approximation algorithm that has an exponential complexity in terms of the bounded parameters, and polynomial in all of the rest. Moving on, we designed a novel genetic algorithm for solving the input-constrained structural target controllability problem, and we implemented a multi-platform standalone graphical application using it for an easier adoption. Additionally, we extended the previously existing greedy algorithm for the same problem, and we performed a thorough comparison between the two when applied on several random and biological networks. With the same goal of facilitating the application of these structural controllability-based methods in biomedicine, we implemented a web application that integrates multiple biological data sources and that allows for the easy generation and analysis of personalized protein-protein interaction network, providing drug suggestions and interaction pathway results. In order to provide a context for the developed algorithms, we also compared them to other network modelling methods in identifying

targeted patient-specific therapeutic suggestions for three multiple myeloma patients. Lastly, to demonstrate the applicability and usefulness of our proposed approaches, we performed structural controllability analysis on two COVID-19-based networks, identifying drugs and drug-combinations that can efficiently control the corresponding networks, and we validated the obtained results against literature and clinical studies.

5.1 Benefits

The structural controllability method is a very powerful tool for the analysis of complex networks. Its versatility makes it ideal for applications in a large variety of domains, and it allows for a wide array of setups, where the same input data can be differently considered in order to lead to multiple valid outputs and interpretations. An additional advantage consists of its ability to manage large data sets, and to provide theoretically accurate control results even when facing missing or incomplete data.

This approach is very well suited for applications in biomedicine, on protein-protein interaction networks, for the identification of proteins of interests in these networks. The existence of as yet undiscovered interactions within the interactome does not affect the validity of its current control results, as the introduction of additional edges in a network can at most affect the optimality of the previously obtained solutions. Furthermore, it allows for the analyzed data to be as general or as detailed as available or as required, and it can easily integrate with drug-, disease-, or patient-specific additional data. Moreover, the proposed structural control framework has been specifically designed with the main goal of identifying potential therapeutics, be they custom to a disease or a patient, which ensures an immediate applicability for drug discovery and repurposing. Similarly, the method can be easily extended to cover more longitudinal studies, where the development of a patient during the progression of a disease or treatment can also be integrated for a truly personalized line of treatment.

5.2 Limitations

While potentially very useful, the structural controllability method does come with several caveats. One of the main limitations concerns its entirely qualitative and non-quantitative approach, as the control results offer no indication on how the control can actually be exerted, but merely that it is theoretically possible. Consequently, the numerical setup may hold additional information of interest that a purely structural approach will fail to consider. This brings to light multiple possible issues for real life scenarios. For example, the method can identify suitable inputs for control, but cannot

provide an actual combination of suitable input functions that could bring the system to the desired final state. Secondly, even in the more favorable cases where both the control inputs and their associated functions would be known, there is no indication whether control would actually be feasible, for example due to possible practical limitations such as the cost of the energy required, or the lack of technology advanced enough for precise control over the inputs.

Another limitation draws from the linear modelling required by network controllability. Many complex networks in general, and biological networks in particular, are often regarded to be non-linear by nature, which may raise questions about the framework's ability to capture the innate underlying characteristics of real-life networks. Furthermore, in addition to these general constraints, the applicability in biomedicine brings forth supplementary concerns. Firstly, the method is highly dependent on the provided data. While the control results are valid even when dealing with incomplete data, their optimality can be affected and, therefore, the potential returned control pathways might not accurately reflect the actual effects of the external drug interventions in the network. In addition, the approach is purely computational. Even though accurate experimental protein, interaction, and drug data can be integrated in the analysis, the results would mainly provide only a starting point towards novel therapeutics and would require thorough subsequent experimental and clinical validation.

5.3 Future work

Despite its limitations, the structural controllability approach is a very promising field, with several interesting research directions already taking shape towards a greater future applicability on real-world complex networks. A first challenge would aim to see the framework shift towards a more quantitative approach, such as taking into consideration the numerical setup and the values of and in-between the internal components when aiming to bring the system to a desired state. Our own attempts at including the type of the interactions (e.g., activation or inhibition) into the analysis for the identification of complementary drugs have shown some promising early computational results. Indeed, although several recent studies have been focused on this area, such as on the input time functions required for control [113], there is a need for efficient and scalable algorithms and software implementations that would be easily integrated with and applied on real-life data. Furthermore, another possible direction would concern overcoming the practical limitations of exerting the control, such as the required energy [114]. It is worth noting that, while the algorithms behind can be relatively generic, the challenges of control are strongly coupled with

the field of application (e.g., it can be more potentially difficult to influence as desired a node in a protein-protein interaction network compared to a node in a circuit network), which highlights the need for more field-specific perspectives.

On the other hand, when considering the same approach to biomedicine described in this thesis, there are multiple additional research directions that could greatly improve its applicability. For example, as the method is greatly dependent on the available data, the aggregation of higher quality data from multiple sources, together with its consolidation, can potentially lead to better and more accurate results. To this end, we are currently working on integrating additional databases, such as the Human Interactome [115] and the Guide To Pharmacology [116], into our framework. Moreover, only the proteins encoded by disease-specific genes have been considered so far, but the method's versatility would allow it to be adapted to also take into account the proteins encoded by survivability-essential genes for healthy cells, as areas of the network which should not be altered through the control paths. The corresponding data and the control paths can be further and more thoroughly analyzed in order to identify and limit the potential side effects that would be caused by the suggested drugs. The additional integration of patient-specific data, such as mutated genes or previous response to treatment lines, could further emphasize the method's potential for personalized medicine. We have already taken several steps towards this direction through our current study of individual data for over a hundred glioma patients, with the aim of identifying suitable drug combinations tailored to the particularities of each patient. Last, but not least, a great benefit would be brought about in the future from close collaboration and integration with laboratory work, for the experimental validation of the offered predictions and general guidance towards the interactome area and drugs which to potentially prioritize.

The relatively recent field of network controllability that encompasses this thesis has been under continuous development and enjoys an increasing relevance to this day, with promising developments for the future. The presented work has a high potential of bringing about a significant impact in multiple areas of science and establishes a basis for the future successful application of network modelling methods in personalized medicine.

Bibliography

[1] Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68, December 2010.

[2] Mika Gustafsson, Colm E Nestor, Huan Zhang, et al. Modules, networks and systems medicine for understanding disease and aiding diagnosis. *Genome Medicine*, 6(10), October 2014.

[3] Giulia Fiscon, Federica Conte, Lorenzo Farina, and Paola Paci. Network-based approaches to explore complex biological systems towards network medicine. *Genes*, 9(9):437, August 2018.

[4] Tong Yang, Xiangyu Chen, Zhigang Mei, et al. An integrated analysis of network pharmacology and experimental validation to reveal the mechanism of chinese medicine formula naotaifang in treating cerebral ischemia-reperfusion injury. *Drug Design, Development and Therapy*, Volume 15:3783–3808, September 2021.

[5] Naiqiang Zhu, Jingyi Hou, and Ning Yang. Network pharmacology integrated with experimental validation revealed the anti-inflammatory effects of andrographis paniculata. *Scientific Reports*, 11(1), May 2021.

[6] Jiansong Fang, Qihui Wu, Fei Ye, et al. Network-based identification and experimental validation of drug candidates toward SARS-CoV-2 via targeting virus–host interactome. *Frontiers in Genetics*, 12, September 2021.

[7] National Research Council. *Network Science*. National Academies Press, December 2005.

[8] Jongkwang Kim and Thomas Wilhelm. What is a complex graph? *Physica A: Statistical Mechanics and its Applications*, 387(11):2637–2652, April 2008.

[9] Shuangming Zhao, Pengxiang Zhao, and Yunfan Cui. A network centrality measure framework for analyzing urban traffic flow: A

case study of wuhan, china. *Physica A: Statistical Mechanics and its Applications*, 478:143–157, July 2017.

[10] Ying Ying Keng, Kiam Heong Kwa, and Kurunathan Ratnavelu. Centrality analysis in a drug network and its application to drug repositioning. *Applied Mathematics and Computation*, 395:125870, April 2021.

[11] Rex Ying, Ruining He, Kaifeng Chen, et al. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, July 2018.

[12] Shashank Pandit, Duen Horng Chau, Samuel Wang, and Christos Faloutsos. Netprobe. In *Proceedings of the 16th international conference on World Wide Web - WWW '07*. ACM Press, 2007.

[13] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In I. Guyon, U. Von Luxburg, S. Bengio, et al., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[14] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466, June 2018.

[15] R. E. Kalman. Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics Series A Control*, 1(2):152–192, January 1963.

[16] David G. Luenberger. *Introduction to dynamic systems: Theory, models, and applications*. Wiley, 1979.

[17] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, May 2011.

[18] Linying Xiang, Fei Chen, Wei Ren, and Guanrong Chen. Advances in network controllability. *IEEE Circuits and Systems Magazine*, 19(2):8–32, 2019.

[19] Lin Wu, Min Li, Jian-Xin Wang, and Fang-Xiang Wu. Controllability and its applications to biological networks. *Journal of Computer Science and Technology*, 34(1):16–34, January 2019.

[20] S. Wuchty. Controllability in protein interaction networks. *Proceedings of the National Academy of Sciences*, 111(19):7156–7160, April 2014.

[21] Xueming Liu and Linqiang Pan. Detection of driver metabolites in the human liver metabolic network using structural controllability analysis. *BMC Systems Biology*, 8(1):51, 2014.

[22] Shi Gu, Fabio Pasqualetti, Matthew Cieslak, et al. Controllability of structural brain networks. *Nature Communications*, 6(1), October 2015.

[23] Gang Yan, Petra E. Vértes, Emma K. Towlson, et al. Network control principles predict neuron function in the caenorhabditis elegans connectome. *Nature*, 550(7677):519–523, October 2017.

[24] Yang-Yang Ding, Hannah Kim, Kellyn Madden, et al. Network analysis reveals synergistic genetic dependencies for rational combination therapy in philadelphia chromosome–like acute lymphoblastic leukemia. *Clinical Cancer Research*, 27(18):5109–5122, July 2021.

[25] Wei-Feng Guo, Shao-Wu Zhang, Yue-Hua Feng, et al. Network controllability-based algorithm to target personalized driver genes for discovering combinatorial drugs of individual patients. *Nucleic Acids Research*, 49(7):e37–e37, January 2021.

[26] James Ladyman, James Lambert, and Karoline Wiesner. What is a complex system? *European Journal for Philosophy of Science*, 3(1):33–67, June 2012.

[27] Javier De Las Rivas and Celia Fontanillo. Protein–protein interactions essentials: Key concepts to building and analyzing interactome networks. *PLoS Computational Biology*, 6(6):e1000807, June 2010.

[28] Iman Tavassoly, Joseph Goldfarb, and Ravi Iyengar. Systems biology primer: the basic methods and approaches. *Essays in Biochemistry*, 62(4):487–500, October 2018.

[29] Kevin Titeca, Irma Lemmens, Jan Tavernier, and Sven Eyckerman. Discovering cellular protein-protein interactions: Technological strategies and opportunities. *Mass Spectrometry Reviews*, 38(1):79–111, June 2018.

[30] Pandjassarame Kangueane and Christina Nilofer. Databases for protein-protein interaction. In *Protein-Protein and Domain-Domain Interactions*, pages 113–124. Springer Singapore, 2018.

[31] Laura Bonetta. Interactome under construction. *Nature*, 468(7325):851–852, December 2010.

[32] Lin Gao, Peng-Gang Sun, and Jia Song. Clustering algorithms for detecting functional modules in protein interaction networks. *Journal of Bioinformatics and Computational Biology*, 07(01):217–242, February 2009.

[33] R. Sever and J. S. Brugge. Signal transduction in cancer. *Cold Spring Harbor Perspectives in Medicine*, 5(4):a006098–a006098, April 2015.

[34] Juan A Godoy, Juvenal A Rios, Juan M Zolezzi, et al. Signaling pathway cross talk in alzheimer's disease. *Cell Communication and Signaling*, 12(1):23, 2014.

[35] G. Rickey Welch. The 'fuzzy' interactome. *Trends in Biochemical Sciences*, 34(1):1–2, January 2009.

[36] Einat Sprinzak, Shmuel Sattath, and Hanah Margalit. How reliable are experimental protein–protein interaction data? *Journal of Molecular Biology*, 327(5):919–923, April 2003.

[37] Karin Breuer, Amir K. Foroushani, Matthew R. Laird, et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Research*, 41(D1):D1228–D1233, November 2012.

[38] Damian Szklarczyk, Annika L Gable, David Lyon, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1):D607–D613, November 2018.

[39] M. P. H. Stumpf, T. Thorne, E. de Silva, et al. Estimating the size of the human interactome. *Proceedings of the National Academy of Sciences*, 105(19):6959–6964, May 2008.

[40] István Bartha, Julia di Iulio, J. Craig Venter, and Amalio Telenti. Human gene essentiality. *Nature Reviews Genetics*, 19(1):51–62, October 2017.

[41] T. Wang, K. Birsoy, N. W. Hughes, et al. Identification and characterization of essential genes in the human genome. *Science*, 350(6264):1096–1101, October 2015.

[42] William C. Hahn, Joel S. Bader, Theodore P. Braun, et al. An expanded universe of cancer targets. *Cell*, 184(5):1142–1155, March 2021.

[43] Wook Lee, De-Shuang Huang, and Kyungsook Han. Constructing cancer patient-specific and group-specific gene networks with multi-omics data. *BMC Medical Genomics*, 13(S6), August 2020.

[44] David S Wishart, Yannick D Feunang, An C Guo, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1):D1074–D1082, November 2017.

[45] Yunxia Wang, Song Zhang, Fengcheng Li, et al. Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Research*, November 2019.

[46] Jianxi Gao, Yang-Yu Liu, Raissa M. D'Souza, and Albert-László Barabási. Target control of complex networks. *Nature Communications*, 5(1), November 2014.

[47] Eugen Czeizler, Kai-Chiu Wu, Cristian Gratie, et al. Structural target controllability of linear networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(4):1217–1228, July 2018.

[48] Mahmoud Saleh, Yusef Esa, and Ahmed Mohamed. Applications of complex network analysis in electric power systems. *Energies*, 11(6):1381, May 2018.

[49] Nour Raeef Al-Molhem, Yasser Rahal, and Mustapha Dakkak. Social network analysis in telecom data. *Journal of Big Data*, 6(1), November 2019.

[50] Stephen P. Borgatti, Martin G. Everett, and Jeffrey C. Johnson. *Analyzing social networks*. Sage, 2018.

[51] Matthew D. B. Jackson, Salva Duran-Nebreda, and George W. Bassel. Network-based approaches to quantify multicellular development. *Journal of The Royal Society Interface*, 14(135):20170484, October 2017.

[52] Yu-Shuai Li, Da-Zhong Ma, Hua-Guang Zhang, and Qiu-Ye Sun. Critical nodes identification of power systems based on controllability of complex networks. *Applied Sciences*, 5(3):622–636, September 2015.

[53] Danilo Delpini, Stefano Battiston, Massimo Riccaboni, et al. Evolution of controllability in interbank networks. *Scientific Reports*, 3(1), April 2013.

[54] Alexander J. Gates and Luis M. Rocha. Control of complex networks requires both structure and dynamics. *Scientific Reports*, 6(1), April 2016.

[55] Junjie Jiang and Ying-Cheng Lai. Irrelevance of linear controllability to nonlinear dynamical networks. *Nature Communications*, 10(1), September 2019.

[56] Sean P. Cornelius, William L. Kath, and Adilson E. Motter. Realistic control of network dynamics. *Nature Communications*, 4(1), June 2013.

[57] Ching-Tai Lin. Structural controllability. *IEEE Transactions on Automatic Control*, 19(3):201–208, June 1974.

[58] R. Shields and J. Pearson. Structural controllability of multiinput linear systems. *IEEE Transactions on Automatic Control*, 21(2):203–212, April 1976.

[59] K. Murota and S. Poljak. Note on a graph-theoretic criterion for structural output controllability. *IEEE Transactions on Automatic Control*, 35(8):939–942, 1990.

[60] Wenwu Yu, Guanrong Chen, Ming Cao, and Jurgen Kurths. Second-order consensus for multiagent systems with directed topologies and nonlinear dynamics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(3):881–891, June 2010.

[61] John E. Hopcroft and Richard M. Karp. A n5/2 algorithm for maximum matchings in bipartite. In *12th Annual Symposium on Switching and Automata Theory (swat 1971)*. IEEE, October 1971.

[62] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, et al. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, March 2004.

[63] R. Milo, S. Shen-Orr, S. Itzkovitz, et al. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, October 2002.

[64] Krishna Kanhaiya, Eugen Czeizler, Cristian Gratie, and Ion Petre. Controlling directed protein interaction networks in cancer. *Scientific Reports*, 7(1), September 2017.

[65] J. L. Y. Koh, K. R. Brown, A. Sayad, et al. COLT-cancer: functional genetic screening resource for essential genes in human cancer cell lines. *Nucleic Acids Research*, 40(D1):D957–D963, November 2011.

[66] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11–15, 2008.

[67] Wei-Feng Guo, Shao-Wu Zhang, Ze-Gang Wei, et al. Constrained target controllability of complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(6):063402, June 2017.

[68] James W. Demmel. 7. iterative methods for eigenvalue problems. In *Applied Numerical Linear Algebra*, pages 361–387. Society for Industrial and Applied Mathematics, January 1997.

[69] Kousik Das, Sovan Samanta, and Madhumangal Pal. Study on centrality measures in social networks: a survey. *Social Network Analysis and Mining*, 8(1), February 2018.

[70] Elisabetta Bergamini, Michele Borassi, Pierluigi Crescenzi, et al. Computing top-k closeness centrality faster in unweighted graphs. *ACM Transactions on Knowledge Discovery from Data*, 13(5):1–40, October 2019.

[71] Juris Hartmanis. Computers and intractability: A guide to the theory of NP-completeness (michael r. garey and david s. johnson). *SIAM Review*, 24(1):90–91, January 1982.

[72] Abdel-Rahman Hedar and Rashad Ismail. Simulated annealing with stochastic local search for minimum dominating set problem. *International Journal of Machine Learning and Cybernetics*, 3(2):97–109, August 2011.

[73] Julian D. Schwab, Silke D. Kühlwein, Nensi Ikonomi, et al. Concepts in boolean network modeling: What do they all mean? *Computational and Structural Biotechnology Journal*, 18:571–582, 2020.

[74] Lea Siegle, Julian D. Schwab, Silke D. Kühlwein, et al. A boolean network of the crosstalk between IGF and wnt signaling in aging satellite cells. *PLOS ONE*, 13(3):e0195126, March 2018.

[75] Meike Dahlhaus, Andre Burkovski, Falk Hertwig, et al. Boolean modeling identifies greatwall/MASTL as an important regulator in the AURKA network of neuroblastoma. *Cancer Letters*, 371(1):79–89, February 2016.

[76] Tatsuya Akutsu, Morihiro Hayashida, Wai-Ki Ching, and Michael K. Ng. Control of boolean networks: Hardness results and algorithms for

tree structured networks. *Journal of Theoretical Biology*, 244(4):670–679, February 2007.

[77] Enrico Borriello and Bryan C. Daniels. The basis of easy controllability in boolean networks. *Nature Communications*, 12(1), September 2021.

[78] Tamás Nepusz and Tamás Vicsek. Controlling edge dynamics in complex networks. *Nature Physics*, 8(7):568–573, May 2012.

[79] Furong Lu, Kaikai Yang, and Yuhua Qian. Target control based on edge dynamics in complex networks. *Scientific Reports*, 10(1), June 2020.

[80] Shao-Peng Pang and Fei Hao. Target control of edge dynamics in complex networks. *Physica A: Statistical Mechanics and its Applications*, 512:14–26, December 2018.

[81] Xizhe Zhang, Huaizhen Wang, and Tianyang Lv. Efficient target control of complex networks based on preferential matching. *PLOS ONE*, 12(4):e0175375, April 2017.

[82] Reka Albert, John Baillieul, and Adilson E. Motter. Introduction to the special issue on approaches to control biological and biologically inspired networks. *IEEE Transactions on Control of Network Systems*, 5(2):690–693, June 2018.

[83] Yang-Yu Liu and Albert-László Barabási. Control principles of complex systems. *Reviews of Modern Physics*, 88(3), September 2016.

[84] Atsushi Mochizuki, Bernold Fiedler, Gen Kurosawa, and Daisuke Saito. Dynamics and control at feedback vertex sets. II: A faithful monitor to determine the diversity of molecular activities in regulatory networks. *Journal of Theoretical Biology*, 335:130–146, October 2013.

[85] Jorge G. T. Zañudo and Réka Albert. Cell fate reprogramming by control of intracellular network dynamics. *PLOS Computational Biology*, 11(4):e1004193, April 2015.

[86] Albert-László Barabási and Zoltán N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, February 2004.

[87] Nahid Safari-Alighiarloo, Mohammad Taghizadeh, Mostafa Rezaei-Tavirani, et al. Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterology and hepatology from bed to bench*, 7(1):17–31, 2014.

[88] H. Yu, P. Braun, M. Yildirim, et al. High-quality binary protein interaction map of the yeast interactome network. *Science*, 322(5898):104–110, October 2008.

[89] Jianzhen Xu and Yongjin Li. Discovering disease-genes by topological features in human protein–protein interaction network. *Bioinformatics*, 22(22):2800–2805, September 2006.

[90] Martijn A Huynen, Berend Snel, Christian von Mering, and Peer Bork. Function prediction and protein networks. *Current Opinion in Cell Biology*, 15(2):191–198, April 2003.

[91] M. G. Kann. Protein interactions and disease: computational approaches to uncover the etiology of diseases. *Briefings in Bioinformatics*, 8(5):333–346, June 2007.

[92] K.-I. Goh and I.-G. Choi. Exploring the human diseasome: the human disease network. *Briefings in Functional Genomics*, 11(6):533–542, October 2012.

[93] Dana Silverbush and Roded Sharan. A systematic approach to orient the human protein–protein interaction network. *Nature Communications*, 10(1), July 2019.

[94] Svetlana Gerdes, Robert Edwards, Michael Kubal, et al. Essential genes on metabolic maps. *Current Opinion in Biotechnology*, 17(5):448–456, October 2006.

[95] Jonathan E. Dickerson, Ana Zhu, David L. Robertson, and Kathryn E. Hentges. Defining the role of essential genes in human disease. *PLoS ONE*, 6(11):e27368, November 2011.

[96] Asfandyar Sheikh, Syed Ather Hussain, Quratulain Ghori, et al. The spectrum of genetic mutations in breast cancer. *Asian Pacific Journal of Cancer Prevention*, 16(6):2177–2185, April 2015.

[97] M. Uhlen, L. Fagerberg, B. M. Hallstrom, et al. Tissue-based map of the human proteome. *Science*, 347(6220):1260419–1260419, January 2015.

[98] J M Harrold, M Ramanathan, and D E Mager. Network-based approaches in drug discovery and early development. *Clinical Pharmacology & Therapeutics*, 94(6):651–658, September 2013.

[99] Ismail Kola and John Landis. Can the pharmaceutical industry reduce attrition rates? *Nature Reviews Drug Discovery*, 3(8):711–716, August 2004.

[100] Andrew L Hopkins. Network pharmacology: the next paradigm in drug discovery. *Nature Chemical Biology*, 4(11):682–690, October 2008.

[101] Bryony Braschi, Paul Denny, Kristian Gray, et al. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Research*, 47(D1):D786–D792, October 2018.

[102] Andrew D Yates, Premanand Achuthan, Wasiu Akanni, et al. Ensembl 2020. *Nucleic Acids Research*, November 2019.

[103] UniProt Consortium, Alex Bateman, Maria-Jesus Martin, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1):D480–D489, November 2020.

[104] Garth R. Brown, Vichet Hem, Kenneth S. Katz, et al. Gene: a gene-centered information resource at NCBI. *Nucleic Acids Research*, 43(D1):D36–D42, October 2014.

[105] M. Kanehisa. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, January 2000.

[106] Dénes Türei, Tamás Korcsmáros, and Julio Saez-Rodriguez. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nature Methods*, 13(12):966–967, November 2016.

[107] Luana Licata, Prisca Lo Surdo, Marta Iannuccelli, et al. SIGNOR 2.0, the SIGnaling network open resource 2.0: 2019 update. *Nucleic Acids Research*, October 2019.

[108] Jesse S. Boehm, Mathew J. Garnett, David J. Adams, et al. Cancer research needs a better map. *Nature*, 589(7843):514–516, January 2021.

[109] Fabrizio Chiti and Christopher M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, 75(1):333–366, June 2006.

[110] Zharko Daniloski, Tristan X. Jordan, Hans-Hermann Wessels, et al. Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell*, 184(1):92–105.e16, January 2021.

[111] Lisa Scarpace, Tom Mikkelsen, Soonmee Cha, et al. Radiology data from the cancer genome atlas glioblastoma multiforme [tcga-gbm] collection, 2016.

[112] Jens G. Lohr, Petar Stojanov, Scott L. Carter, et al. Widespread genetic heterogeneity in multiple myeloma: Implications for targeted therapy. *Cancer Cell*, 25(1):91–101, January 2014.

[113] Christian Commault, Jacob van der Woude, and Paolo Frasca. Functional target controllability of networks: Structural properties and efficient algorithms. *IEEE Transactions on Network Science and Engineering*, 7(3):1521–1530, July 2020.

[114] Gustav Lindmark and Claudio Altafini. Minimum energy control for complex networks. *Scientific Reports*, 8(1), February 2018.

[115] Katja Luck, Dae-Kyum Kim, Luke Lambourne, et al. A reference map of the human binary protein interactome. *Nature*, 580(7803):402–408, April 2020.

[116] Simon D Harding, Jane F Armstrong, Elena Faccenda, et al. The IUPHAR/BPS guide to PHARMACOLOGY in 2022: curating pharmacology for COVID-19, malaria and antibacterials. *Nucleic Acids Research*, October 2021.

ISBN 978-952-12-4205-2