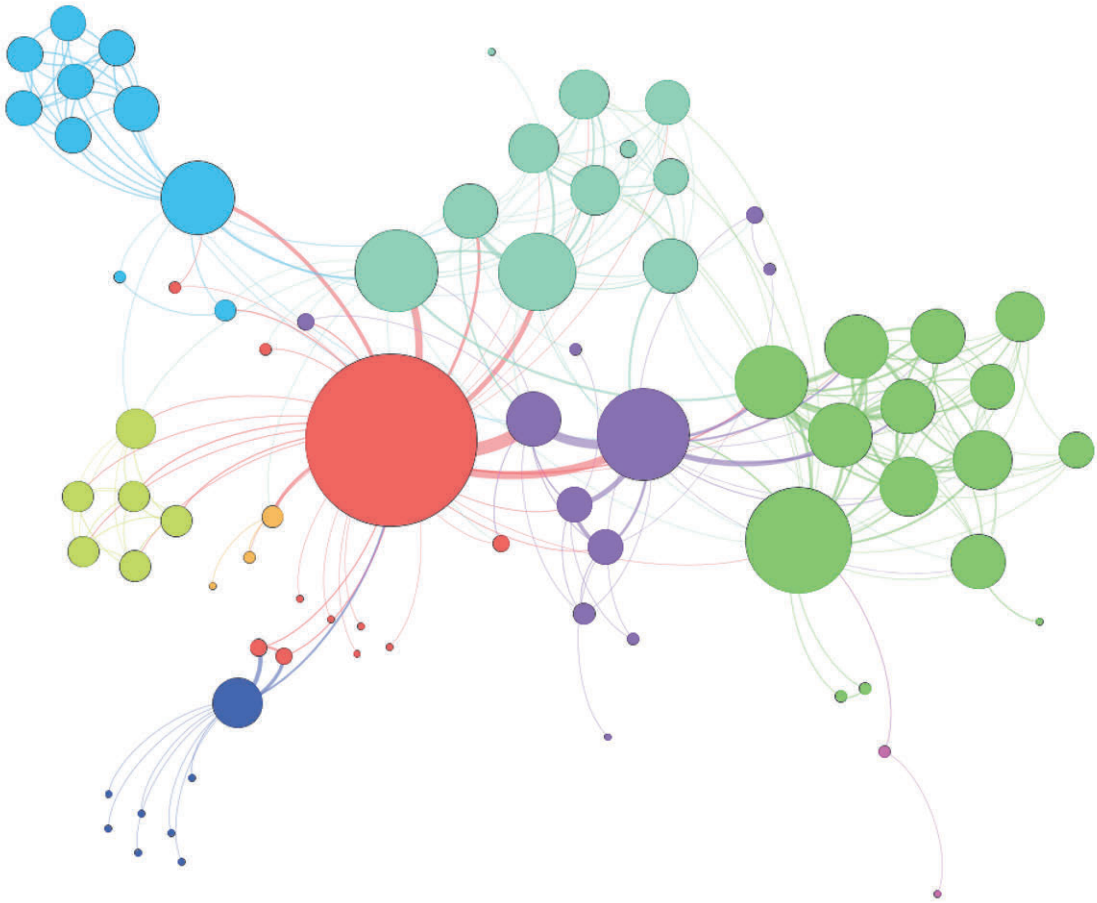
Åbo Akademi University

Krishna Kanhaiya

Target Controllability of Cancer Networks



Krishna Kanhaiya

Born in 1987, Ranchi, India

MSc. in Bioinformatics,
Birla Institute of Technology, Mesra, Ranchi, India, 2011.

PhD. Dissertation defense,
Åbo Akademi University, Turku, Finland, 2020.

Target Controllability of Cancer Networks

Krishna Kanhaiya

Åbo Akademi University
Faculty of Science and Engineering
Agora building, Vattenborgsvägen 3,
20500 Åbo, Finland

2020

Supervisors

Dr. Eugen Czeziler
Faculty of Science and Engineering
Åbo Akademi University
Agora building, Vattenborgsvägen 3, 20500 Åbo,
Finland

Co-supervisors

Prof. Ion Petre
Department of Mathematics and Statistics
University of Turku
20014 Turku,
Finland

Reviewers

Dr. Jing Tang
Faculty of Medicine
University of Helsinki
00014 Helsinki,
Finland

Prof. Mihaela Păun
National Institute for Research and Development of Biological Sciences
060031 Bucharest,
Romania

Opponent

Dr. Jing Tang
Faculty of Medicine
University of Helsinki
00014 Helsinki,
Finland

ISBN 978-952-12-3947-2 (printed)
ISBN 978-952-12-3948-9 (digital)
Painiosalama Oy, Turku, Finland 2020

Abstract

Advances in the field of complex networks theory and network biology pave a new way to define human health through the study of networks of proteins, genes, metabolites, modules across cell signaling pathways, and clinical data. Combinations of large scale biological datasets and concepts from network theory, and systems biology produce new insights into the complex dynamic processes involved in human diseases such as cancer. To develop novel data-driven computational tools for discovering the insights of human diseases and for a new approach to multi-drug therapies for personalized therapeutics, it needs combinations of the high-quality set of human interactome networks, disease-specific expression data, and powerful network controllability algorithmics. Therefore, we address the issue of this thesis with the focus to integrate network biology and network controllability approach, to gain useful insight in the finding of the complex mechanism of cancer networks and open the door for a novel drug target approach called multi-drug therapeutics.

The first part of the thesis presents the network biology approaches to study the interactome of the biological systems and decode the wiring diagram of the cellular information processing systems. It reveals a variety of high-level intramolecular relationships including protein-protein interaction networks (PPI), protein compound interactions, gene regulatory interactions, and metabolic pathways. These interactions play a key role in the development of diseases and various types of cancers. One characteristic of such networks is that a small number of nodes in the networks are highly connected. Another characteristic is that a group of physically and functionally interconnected molecules driving to achieve a common biological process, have a modular structure. Further, through a minimum number of target nodes a full (partial) controllability of these intracellular network can be achieved.

The second part of the thesis presents the network controllability approach and some of the algorithms used in our case studies on different types of cancer PPI signaling networks. Recently, network control theory has been increasingly used in engineering and mathematics which also opens the way to investigate control principals for complex biological interaction

networks through a minimum set of input (driver) nodes. According to control theory, a dynamical system may be steard such that its output is driven towards some desired final states (e.g target cancer essential proteins in PPI networks) via suitably-picked inputs (e.g. manipulating a set of driver proteins). Therefore, it is necessary to understand the dynamics of these complex networks, and their evolution rules (i.e., expressed as a system of linear equations) which govern the systems dynamics over time.

This doctoral thesis provides the target control theory approach fine tuned for the analysis of specific cancer signaling transduction PPI networks. The control approach presented here can be an impressive framework for effective development of multi drug-target therapeutics. We, therefore, expect that our approach can open a new way towards effective and efficient therapeutics target and a key resource towards personalized medicine in cancer.

Sammanfattning

Framsteg inom området teori om komplexa nätverk och nätverksbiologi banar en ny väg när det gäller att definiera mänsklig hälsa genom studiet av nätverk av proteiner, gener, metaboliter, moduler över stigar för cellsignalering, och kliniska data. Kombinationer av storskaliga biologiska datamängder och koncept från nätverksteori samt systembiologi skapar nya insikter i de komplexa dynamiska processer som är inblandade i mänskliga sjukdomar såsom cancer. För att utveckla nya datadrivna beräkningsverktyg för att upptäcka insikter i mänskliga sjukdomar och för en ny anfallsvinkel gällande multiläkemedelsbehandlingar för personlig terapeutika behövs kombinationer av den högkvalitativa mängden av mänskliga interaktoma nätverk, sjukdomsspecifika uttrycksdata, och kraftfulla algoritmer för styrbarhet av nätverk. Därför fokuserar vi problematiken i denna avhandling på att integrera nätverksbiologi och sätt att närma sig styrbarhet av nätverk, på att få användbar insikt i hittandet av de komplexa mekanismerna i cancernätverk, och att öppna dörren för ett nytt sätt att närma sig läkemedelsmål kallat multiläkemedelsterapeutika.

Första delen av avhandlingen introducerar nätverksbiologins sätt att studera interaktomen av de biologiska systemen och dekoda de cellulära informationsprocesseringssystemens kopplingsschema. Den avslöjar en mängd av intramolekylära förhållanden på hög nivå inklusive protein-proteininterak tionsnätverk (PPI), proteinföreningsinteraktioner, genreglerande interaktioner, och metaboliska stigar. Dessa interaktioner spelar nyckelrollen i utvecklingen av av sjukdomar och diverse typer av cancer. Ett kännetecken för sådana nätverk är att ett fåtal noder i nätverket är i hög grad sammanhängande med andra noder. Ett annat kännetecken är att en grupp fysiskt och funktionellt sammankopplade molekyler som försöker uppnå en gemensam biologisk process har en modulär struktur. Dessutom kan full (partiell) styrbarhet av dessa intracellulära nätverk åstadkommas genom ett minimum av målnoder.

I den andra delen av avhandlingen presenteras styrbarhet av nätverk och några av de algoritmer som har använts i våra fallstudier av olika typer av PPI-signalnätverk för cancer. Under den senaste tiden har teori om nätverkskontroll använts allt mer inom ingenjörskonst och matematik, vilket

också banar väg för att undersöka kontrollprinciper för komplexa biologiska interaktionsnätverk via en minimal mängd inmatnings- (förar-)noder. Enligt kontrollteorin kan ett dynamiskt system styras så att dess utmatning drivs mot några önskade slutgiltiga tillstånd (t.ex. i PPI-nätverk inriktar sig på proteiner nödvändiga för cancer). Därför är det nödvändigt att förstå dynamiken i dessa komplexa nätverk och deras utvecklingsregler (uttryckta som ett linjärt ekvationssystem), vilka reglerar systemdynamiken över tid.

Denna doktorsavhandling presenterar målinriktad styrteori, finjusterad för analys av specifika PPI nätverk för transduktion av cancer signalering. Den styrmetod som presenteras här kan vara ett imponerande ramverk för effektiv utveckling av fler-drogterapi. Vi förväntar oss att vårt tillvägagångssätt kan möjliggöra ett nytt sätt för verkningsfull och effektiv målterapi samt en viktig resurs för personlig medicin mot cancer.

Acknowledgements

First of all, I would like to express my profound gratitude to my co-supervisor Prof. Ion Petre, for his continuous support, excellent advice, and his continuous encouragement. I am thankful to him for all the time of research, advice, encouragement, patient guidance, and writing of this thesis. I am extremely fortunate to have a mentor who not only cared about my research work but also a respondent to constant support and frequent insights throughout my period as his PhD student.

Further, immense gratitude goes to my supervisor Dr. Eugen Czeizler for the constant support, many useful and constructive suggestions, and his continuous encouragement for the accomplishment of the work presented in this thesis. His support in this thesis work has been extremely supportive and generous with his considerable amount of time that he spent to give me constructive feedback for the improvement of my thesis. His support related to the target controllability of linear systems was essential for this thesis. I truly appreciate his valuable efforts for finishing this thesis on time and to give me constructive feedback for the improvement of my thesis.

Furthermore, I wish to express my gratitude to Dr. Jing Tang, for agreeing to be a reviewer of my dissertation and act as the opponent at my doctoral defense. I thank him also for providing valuable suggestions and reviews to improve the quality of the thesis. I wish to express my gratitude and thanks to Prof. Mihaela Păun for agreeing to be a reviewer of my thesis, and for providing encouraging comments.

I would like to thank all the co-authors who have contribute in all of my publication and without them, it would not have been possible to prepare this doctoral thesis. I have had the privilege to work with Dr. Vladimir Rogojin, Dr. Dwitiya Tiyagi-Tiwari, Dr. Cristian Gratie, Keivan Kazemi, Wu Kai Chiu. I enjoyed working with them and learn various new things from their research experience and knowledge.

I would also like to express my gratitude to Prof. Rajib Bandopadhyay, University of Burdwan, India and Prof. Seshadev Padhi, Birla Institute of Technology, Mesra, India for their constant encouragement, assistance, and support for an early stage of my research career.

I thank each and every member during my time at Computation Biomod-

eling Laboratory and the Faculty of Science and Engineering at Åbo Akademi University. I have always been received constructive suggestions, scientific discussions, and guidance and support by former and present members. I extend my gratitude to Charmi Panchal, Usman Sanwal, Mikhail Barash, Diana-Elena Gratie, Bogdan Iancu, Utku Kapucu and Victor Bogdan Popescu. I express my special thanks to Petter Sandvik and Joel Sjöberg for providing the Swedish translation of the abstract of this thesis. I also express my gratitude to Christel Engblom, Minna Carla and, Karl Rönnholm for all the administrative and, technical and non-technical matters. I thank TUCS, Åbo Akademi University, the Centre for International Mobility (CIMO), and Business Finland (Tekes) for providing financial assistance of my doctoral studies and research.

Finally, I would like to thank my family members for their cheeriness, all-embracing support, and encouragement throughout these years. I would like to express my special gratitude to my elder brother Vidya Sagar, for his continuous encouragement, support, and mentorship. I wish to express my heartfelt thanks to my wife Neha Kumari for her love and support to me through all these years. I would like to express my eternally grateful to my parents for their love and supporting me throughout my life. I will be forever indebted to your support.

Krishna Kanhaiya
Turku, June 2020

List of original publications

1. **Krishna Kanhaiya**, Eugen Czeizler, Cristian Gratie, and Ion Petre. Controlling directed protein interaction networks in cancer. Scientific Reports 7, no. 1 (2017):10327.

2. **Krishna Kanhaiya**, Vladimir Rogojin, Keivan Kazemi, Eugen Czeizler, and Ion Petre. NetControl4BioMed: a pipeline for biomedical data acquisition and analysis of network controllability. BMC Bioinformatics 19, no. 7 (2018): 185.

3. Eugen Czeizler, Cristian Gratie, Wu Kai Chiu, **Krishna Kanhaiya**, and Ion Petre. Structural target controllability of linear networks. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 15, no. 4, 2018. (A short version of this article received the NVIDIA best paper award at The 14th conference on Computational Methods in Systems Biology (CMSB2016) on the 21st to 23rd September 2016 in Cambridge.)

4. **Krishna Kanhaiya**, Dwitiya Tiyagi-Tiwari. Identification of Drug Targets in Breast Cancer Metabolic Network. Journal of Computational Biology, Volume 26, Number 0, 2019.

5. **Krishna Kanhaiya**. Bioinformatics for Diseases Management: A Personalized Therapeutics Prospective. In Advances in Personalized Nanotherapeutics, pp. 187-199. Springer, Cham, 2017.

Contents

Chapter 1

Introduction

Cancer is a complex disease and it often occurs due to genetic and epigenetic alterations [82]. These alterations further allow other cells to adapt and over-proliferate as tumor cells, and to develop tumor micro-environments [96]. In molecular pathways, tumor cells perform various cancer-related dysregulations which control essential cell functions such as differentiation, survival and cell's growth factors [82]. Also, these cancer cells have the innate capacity to establish and proliferate in adverse conditions. They can succeed even after anticancer therapeutics and get into an immunosuppressive state [19]. These complex processes which develop tumor cells as malignant cells [103, 82], are transmitted mainly through protein-protein interactions (PPI) [55, 39] and metabolic networks [80]. Proteins and metabolites work as vehicles in the signaling pathways while information is transmitted through interactions among them. For instance, PPI play a key role in the regulation of phosphorylation of serine/threonine residues and initiate the tumor necrosis factor to transmit the signals from the receptors to downstream signals [76]. Also, by using PPI, the receptor tyrosine kinases (RTKs) mediate various intra-molecular interactions [76] which causes downstream signaling of RTKs and rewire the signaling pathways [75]. Usually, overexpression rate of RTK modules is very high in cancer which mostly leads to signaling processes to escalate the tumor progression [2]. Similarly, mutations of tumor cells in metabolic pathways intermittently increase the signaling of the PI3K-AKT-mTOR pathway and allows further activities which cause tumor suppression and oncogene activation [18]. Furthermore, the overexpression level of MYC pathways in cancer increases anabolic growth, mitochondrial metabolism, serine metabolism and promote tumorigenicity [90]. These studies illustrate that to comprehensively understand the complex dynamics of signaling pathways networks, we have to understand how different molecular pathways communicate with each other, and the role of proteins and metabolites which intermediate those signaling components. Therefore,

network approaches give us a valuable tool to define and provide a better understanding of multiple information processing abilities during molecular alteration in cancer cell lines.

Computational modeling of biological networks has revolutionized the human diseasome research and has opened a new way towards the development of novel therapeutic targets and personalized medicine [22, 39]. The network-based analysis not only describes the pattern of molecular signaling interactions but also reveals the transcriptional circuits, enrichment patterns, and system-wide properties[80, 4, 82, 74]. Also, network-based approaches focused on biological research help us to understand the dynamics and control characteristics of multiple complex biochemical networks in cooperation with matching experimental findings.

The research presented in this doctoral dissertation concentrates on network control frameworks, target controllability of linear networks, on model construction to find effective drug targets in cancer, and on its contribution to personalized medicine.

In the second chapter, we describe the network biology approach which studies the interactome of the biological systems and decodes the wiring diagram of the cellular information processing systems. It reveals a variety of high-level intramolecular relationships including PPI, protein compound interactions, gene regulatory interactions, and metabolic pathways essential to these cooperative activities [34, 106, 5]. One characteristic of such networks is that a small number of nodes in the networks are highly connected [4, 5]. Another characteristics is that a group of physically and functionally interconnected molecules driving to achieve a common biological process, have a modular structure, and that through a minimum number of target nodes a full(partial) controllability of these intracellular network can be achieved [5, 30].

The network biology approach leads us to map various disease networks, and to identify the genotypic and phenotypic relationship of essential genes in the diseases to other genes and disease modules [25, 4]. Further, disease networks suggest that essential genes and disease related genes encode the properties of the hub in the networks [35, 11]. In almost all types of diseases, PPI networks play a pivotal role in spreading and maintenance of that particular disease. On the other hand these PPI networks are also a powerful tool to analyze the biomolecular basis of diseases and give clues to the function of the disease proteins [4, 29]. Also, these PPI have the highest number of associations with other diseases. Furthermore, these PPI can predict disease-specific patterns, which can further lead to the discovery of therapeutic targets and diagnostic biomarkers.

Metabolic networks encompass the biochemical reaction pathways and their correlated molecules, which initiates the interactions among cellular molecules and metabolism [10]. Dysregulation or alteration of the metabolic

process causes a number of human diseases [23]. Moreover, metabolic disorders are associated with different types of genetic mutations, which cause enzymes to non-expression and inactivation in certain cellular functions [51]. In disease states, the cascading effect of metabolic pathways are crucial for comprehending disease-specific biochemical pathways [8]. More, alteration in metabolic-related activities maintain malignant properties and their survival in cancer [87].

All of these biomolecular interactions make cancer a complex disease and very robust in nature [19]. Through these interaction maps cancer cells drive tumor growth, energy production and biomass production for cancer [19, 29, 18]. These features of cancer can be crucial to understanding the pattern of cancer networks [91]. Also in cancer, cancer-associated metabolic alterations present in all stages of cellular metabolic interactions, and this makes it difficult to distinguish from the normal cellular proliferation [27]. This in turn, can be used for developing advanced biomarkers to uncover the disease mechanism, predicting cancer genes, improve its classification, and influencing drug development.

Therefore, to develop the next generation of cancer therapeutics, computational based approaches with molecular experimental techniques can significantly define the optimal combination to clinical oncology [59] and be the realm of intuitive therapeutics to personalized medicine in cancer. In these steps, network pharmacology can also improve the traditional approaches by identifying the drug targets and understanding their action on the disease-causing networks [59, 33]. Network-based finding of drugs and respective targets can help in quantification of drug-disease combinations. This combination based approach can offer more efficient clinical targets, able to provide the answer of toxicological related query and play a key role in the treatment of multiple cancer. Further, network pharmacology based drug combination strategy enhances the systems-level understanding, such as how multi signaling pathways are involved in cancerogenesis is inhibited by based on drug combination. This approach can potentially improve the efficacy in the identification of more efficient and effective cancer therapeutics and can improve clinical efficacy [48, 33].

In the third chapter, we discuss the network controllability approach and some of the algorithms used in our case studies on different types of cancer PPI signaling networks. Recently, network control theory has been increasingly used in engineering and mathematics which also opens the way to investigate control principals for complex biological interaction networks through a minimum set of input (driver) nodes [15, 14]. According to control theory, a dynamical system may manage the output of system framework to any desired final states (e.g target cancer essential protein in PPI network) via the direct manipulation pf some suitably picked inputs (e.g. driving a set of driver proteins). Therefore, it is necessary to understand the dynamics of

these complex network, and their evolution rule (i.e., expressed as a system of linear equations) which govern the systems dynamics over time.

We say that a *linear, time invariant dynamical system* is target controllable, if there exist a number of input (driver) nodes which can control a given set of target nodes in finite time. That is, given any desired final configuration of the target nodes and any initial configuration of the systems, we can drive the target nodes to the desired final configuration (in finite time) only by acting upon the designated input (driven) nodes. The target controllability problem for linear networks can be specified as an instance of the output controllability problem [78], and correlated to the full controllability problem, which asks for the control of the entire system. Previously, Liu et al. [58] have presented a polynomial time algorithm for an optimal solution of full controllability. Later, Gao et al. [26] proposed a greedy algorithm for finding the minimum number of driver nodes for target controllability.

In our research, we build on the approach of Poljak and Murota [78, 67] and derive the computational complexity of the target controllability of directed graph structures. We start by presenting the approximation algorithm for the target control problem mentioned by Gao et al. [26], and analyze this algorithm. Finally, we describe our algorithms based on heuristic optimization strategies for more effective investigation used in cancer networks (cancer signaling PPI networks) and aiming for faster calculations and effective optimization.

As the biological networks data are assembling, network based approaches are getting more valid to modeling and understanding of control principles of complex biological systems. Controlling a complex system translates to identifing a set of driver(input) nodes which are essential for its control. As a conclusion of our work, we introduce a target control approach to partial controllability of cancer networks and showed how a set of cancer essential proteins (target nodes) can be controlled from a defined set of drug target proteins (driver nodes) using the directed PPI networks.

In the fourth chapter, we briefly discuss the summary of each research articles included in this thesis. Finally, in the fifth chapter, we conclude our research with a perspective for future research directions.

Chapter 2

The Network Biology Approach

Networks in biological systems are always represented as a complex set of interactions between different entities, such as genes, proteins, metabolites, etc. Therefore, networks are central for our understanding of complex intracellular systems of interactions and provide a conceptual and intuitive framework of structure and function for the different entities within biological systems [34, 105]. The network-based analysis not only describes the pattern of molecular signaling interactions but also reveals the transcriptional factors, binding sites, and system-wide properties [5, 64]. One of the main emphasis of the network biology approach is to unravel the cellular interaction pattern between normal and disease states and describe the role of individual entities in different biological processes [105]. In this chapter, we discuss briefly the basics of the network biology approach, interaction patterns in biological systems and describe the intracellular map of disease and particularly of cancer. We also discuss the network pharmacology approach for the development of the next generation of drugs.

2.1 Background

The high-throughput technologies advanced biological research and generated a massive amount of data. This in turn to provide new opportunities to map the cellular networks. Understanding these complex cellular interactions allows researchers to detect and model the interactome of cells, organs and organism. The network biology approach hence studies the interactome of the biological systems and decodes the wiring diagram of the cellular information processing sytems [5]. It provides both the interaction types and useful explanations to visualize and understand the functions, and interaction patterns of biological systems. Biological networks reveal a

variety of high-level interamolecular relationships including PPI networks, protein compound interactions, gene regulatory interactions, and metabolic pathways essential to these cooperative activities [34, 106, 5]. Further, physical and functional interactions of all these entities are critical and define the working properties of the biomolecules inside these complex biological networks [11]. Physical interactions define the PPI and they are essential for transcription, translation and detection of interaction pattern of groups of proteins [11]. Physical interactions have significantly transform our understanding about relationships between two biomolecules [11]. Functional interactions define how a set of proteins work together to act for a certain function, and aim to connect genes and proteins with similar or related functions.

Biological networks are very complex in nature, and their dynamic characteristics can be expressed in terms of biochemical kinetics, various linear and non-linear relationships, stochasticity and feedback loops [5, 53]. These intracellular interactions can be conveniently defined as networks with nodes which denote biomolecules, and links which denote the interactions between them [5, 10]. Depending on the type of interactions they represent, these corresponding edges can be considered directed or indirected. Along with this, these networks have various topological properties [30]. One such properties is that a small number of nodes in the networks are highly connected while most nodes have very few interactions [30, 5]. These highly connected nodes are defined as hubs and have been often found to play pivotal roles in the biological systems [106, 11]. Similarly, a group of physically and functionally interconnected molecules (proteins, genes, or metabolites) driving to achieve a common biological process, are defined as a module [5, 66]. Modules can be seen in various biological processes; for example, synthesis of proteins and in various signalling pathways [5, 11]. Another property is that of full/partial network controllability: the ability to change a network's state (or that of a target subset of it) [39]. The target control approach can give us efficient ways to control a biological network by using drug-combination based strategy through multi-target perturbations. This method is best fit in disease associated networks and provides a better understanding of the disease associated cellular networks. Further it opens a new way toward network pharmacology and application of multi drug-target based control mechanisms which in turn could pave the way for next paradigm of drug discovery.

2.2 Human disease networks

Complex diseases are mostly a consequence of genetic mutations, which dysregulate multiple molecular processes and create perturbations in the

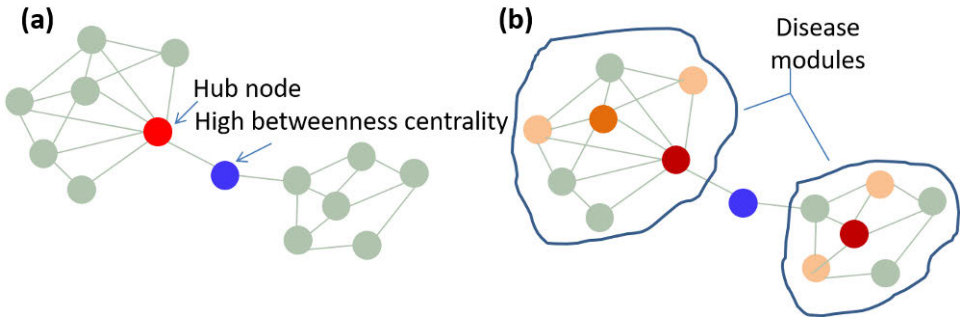


Figure 2.1: **Disease characteristics and modules (a)** Network characteristics such as hub proteins and betweenness centrality have key roles in disease progression. **(b)** A disease module indicates that perturbation of a group of nodes are linked to the occurrence of a particular disease phenotype, shown as red nodes.

expression pattern of a large number of genes [11]. This further disrupts the interconnection mechanism of the cell, the regulatory mechanism of protein/genes and of the metabolic pathways [35, 11]. Therefore, a simple way to provide an insight of the genes underlying human diseases is through network biology, namely to reveal the clues about the genes which are related/interconnected with the disease genes (i.e., the genes which cause the evolution of a particular disease) [4, 63]; see Figure 2.1. (a). This network-based approach of disease study leads to map these networks, and identify the molecular and phenotypic relationship between the essential genes in the disease and the other genes [25]. The integration of physical and functional relationships of a disease networks can reveal genes which are involved in the disease and candidate disease genes. Properties of disease networks suggest that genes/proteins causing a disease encode the properties of the hub and have a key role in occurnace of that particular disease [35, 11]. In humans, hub genes are related to diseases: it has been found that in various cell carcinomas highly upregulated genes have a high degree [4].

The local hypothesis of disease networks exhibit that some proteins/genes involved in the disease show a high tendency to interact with each other. Other studies found that genes linked to diseases show similar phenotypes and have high propensity to interact directly with the other molecules [11, 28]. These observations suggested that, by identifying few disease components, other disease-related components can be easily identified in their network-based vicinity [25, 28]. These well-connected molecules formed a well-defined neighborhood of the interactome, forming a disease module [4]. Therefore, a disease module represents a group of network components which plays a key role in cellular malfunctions and its dysregulations cause disease

phenotype [4, 25]; see Figure 2.1. (b). Disease module characteristics are not always identical. Every disease module interacts with the others modules in a unique way, although some disease modules overlap [25]. Thus, the emergence of a disease is a combinatorial problem, where many different perturbations and dysfunctions result in an identical disease phenotype and modify the activity of the disease modules [25, 4]. These disease mechanisms are found to be well documented and responsible for various epigenetic, transcription and post-translational modification [4].

The highly interconnected nature of disease modules means that it is difficult to consider one disease independent of others; since these modules overlap each other and one disease may trigger another. Therefore, system-based mapping by using network modules can be crucial in finding the mechanism of diseases [4]. Moreover, uncovering these disease links can define how different phenotypes affect and help us to comprehend why a group of diseases arises together. Therefore, network-based approaches identify disease pathogenesis based on their phenotypes, which in turn, can be used for identifying advanced biomarkers to uncover the disease mechanism, predicting disease genes, improving disease classification, and influencing drug development [4].

2.2.1 The human protein-protein interaction (PPI) networks

PPI are an important framework for the study of biological processes. Therefore, a complete map of PPI networks is important to provide detailed insights into the protein mechanism. In human, over 22000 genes encode these proteins, and through alternative splicing mechanism they give rise to other proteins. Mostly, these proteins don't work in isolation and form interactions within each other and with other types of macromolecules [17]. The interactions among all the known proteins are called the interactome [17], and has the key for the study of protein function, and of cellular biochemistry and physiology [4].

Therefore, PPI networks are serving as an important tool to analyze the biomolecular basis of diseases, and give clues to the function of the disease proteins [4, 29]. The PPI networks support in identification of new disease-associated proteins, the properties of networks and its relation to disease's proteins, the identification of disease-associated subnetworks, and classification of disease based on the networks. Disease states not only impact the side effects but also cause the central impact and root origin for initiation or progression of pathology; see Figure 2.2. For example, the central importance of the PPI in tumorigenesis is clearly defined by the p53 tumor suppressor protein, which causes mutations and disrupts p53-HDM2 interactions [37]. The disruption of PPI networks during a disease not only affects single genes/proteins but has implications on a variety of diseases, where PPI

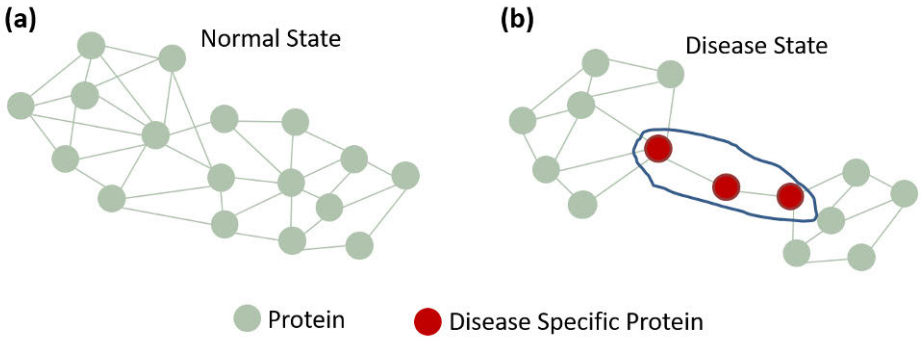


Figure 2.2: **Schematic representation of normal and disease PPI during alterations and it's effect on pathological conditions.** **(a)** A topological view of locally connected neighborhoods of normal PPI. **(b)** A network view of how group of proteins are different after perturbation (mutations, deletions, variations or expression change etc.) and are linked to a specific disease phenotypes, shown as red color.

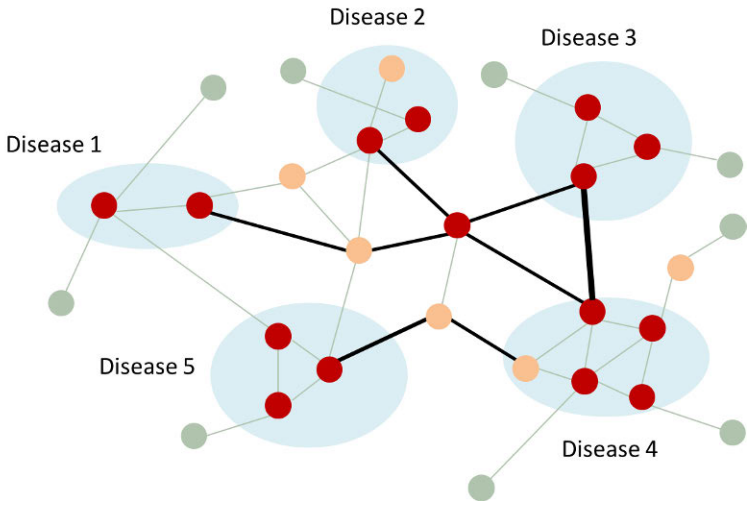


Figure 2.3: **Disease-disease relationship.** Shown in interaction pattern of proteins among various diseases. Disease 3 and Disease 4 are directly connected, whereas Disesae 1, 2, 3 are linked through some other proteins which play a role as a bridge in the network. The color of each protein scales with the change in expression of the corresponding proteins for disease specific proteins (red) versus non-disease ones (gray).

acquire the disease characteristics and establish a disease interactome [29]. Additionally, perturbations in PPI significantly cause expression changes in the various diseases [83], and further affect the composition of protein complexes and influences the disease mechanisms, as well as the mechanism of the diseases also affects other disease mechanisms [83, 102]; see Figure 2.2, 2.3. Therefore, computational based interaction studies of PPI in different human diseases can be used to provide interesting and significant options for further experimental screening for both diagnostic [29] and therapeutic targets and even provide information about interaction details that could have potential for drug combination based therapeutics [4, 39]. Moreover, inhibiting PPI can be envisioned as a disease-specific corrective intervention, which can further lead us to the discovery of new therapeutic targets and of diagnostic biomarkers.

In addition, targeting individual proteins in disease network is not effective, as it has been discovered that single-target drugs are not very efficient to achieve the therapeutic targets. Therefore, the focus is on developing multi-target drugs, and here subnetworks in PPI are significant because these subnetworks address the complexity of dynamic pathological conditions, and lead to the identification of genetic factors that offer mechanistic support in understanding of these diseases. These subnetworks contain functional information of diseases and reveal the information about the interaction patterns of drug and respective targets [29]. Subnetworks are used in the identification of those proteins which have key roles in spreading of that particular disease as well as association with the other diseases [29, 71].

The dysfunctioning of these disease-associated proteins cause many diseases, including cancer. These cancer-associated proteins have significantly different topological properties. Specifically, these proteins show high connectivity and betweenness centrality, have shorter path distance to connect with other cancer proteins, and have even more robust network characteristic than other proteins in the networks [36]. Also, in cancer, these proteins bound with multiple proteins and have hub like characteristics [28]. These hub proteins in cancer have an average high degree in compare to other proteins in the networks [39]. This is often referred to as the scale-free property of disease networks [4]. Because of the hub-like properties of these cancer proteins, they use the interaction of several disease proteins and tend to interact with the partner proteins through specific interfaces with larger interaction sites [41]. On the other hand, these cancer-related hub proteins are strongly interacting with the other hub proteins in the network and therefore play a central role in the progression and initiation of the disease in the network. These PPI based features of cancer proteins can be crucial to understanding the patterns of cancer network [91]. Thus, these properties of PPI cancer networks might be important in finding new and effective targets in cancer as well as in other complex diseases.

2.2.2 Human metabolic interaction networks

A metabolic network (MN) represents the metabolic and biophysical processes needed to understand the mechanism of the physiological and biochemical properties of cellular entities [74]. A MN is connected through its corresponding metabolites and enzymes, which provide the link between them. It consist of a bipartite graphs, where nodes are metabolites and enzymes; see Figure 2.4. Here, edges are biochemical reactions that are catalyzed by definite gene products. Also, the nodes of this network includes information regarding of specific variables in the form of mass or energy flow [89]. Therefore, to understand the functional properties of the MN study of these interaction patterns is of central interest. MN are highly non-random networks and very few of the metabolites act as a hub or are involved in multiple reactions [10]. In MN, the strongly connected modules create self independent clustering, which result in a large size-independent cluster [10]. Feedback loops and multiple-input motifs influence the regulation and dynamics of MN [85]. It has been discovered that enzymatic steps in the metabolic network are often catalyze through protein interactions [21]. Certain proteins are connected with the dominant components of the metabolism and thus essential in several biochemical reactions [21]. Also, protein interactions contributes to metabolic pathways and their physical organization such as increasing the adaptability of the metabolic processes by allowing higher metabolic fluxes [21].

Dysregulation or alteration of the metabolic process causes a number of human diseases [23]. Therefore, an important challenge is to define the relationship among various disease phenotypes which cause the disruptions inside metabolic pathways [60, 81]. Because of these factors, the metabolic-related diseases are of prime interest for defining the interaction map of the whole human cell metabolism [20]. Also they are important in order to determine the accuracy of flux-based balance dependencies in diseases, as well as in finding certain genes role in the grouping of metabolic diseases [20]. Metabolic disorders are associated with different types of mutation, which cause enzymes to be underexpressed, and leading to an inactivation of certain cellular functionality [51]. The consequence of disease phenotype examined as a cell's inability to impart metabolic substrate, is to produce toxicity levels above of the threshold than the normal functionality of molecules [73]. Metabolic disorders also affect several building blocks of cellular function, such as amino acids, carbohydrates, fatty acids, etc [51, 87]. Cellular metabolism is conducted through enzyme-catalyzed biochemical reactions, such that a deficiency of enzyme leads to a cascade of effects affecting the flux of multi subsequent reactions [20, 1]. In disease states, cascading effects of multiple metabolic pathways are directly associated with disease-specific biochemical pathways [8]. Therefore, systematic mapping of metabolic as-

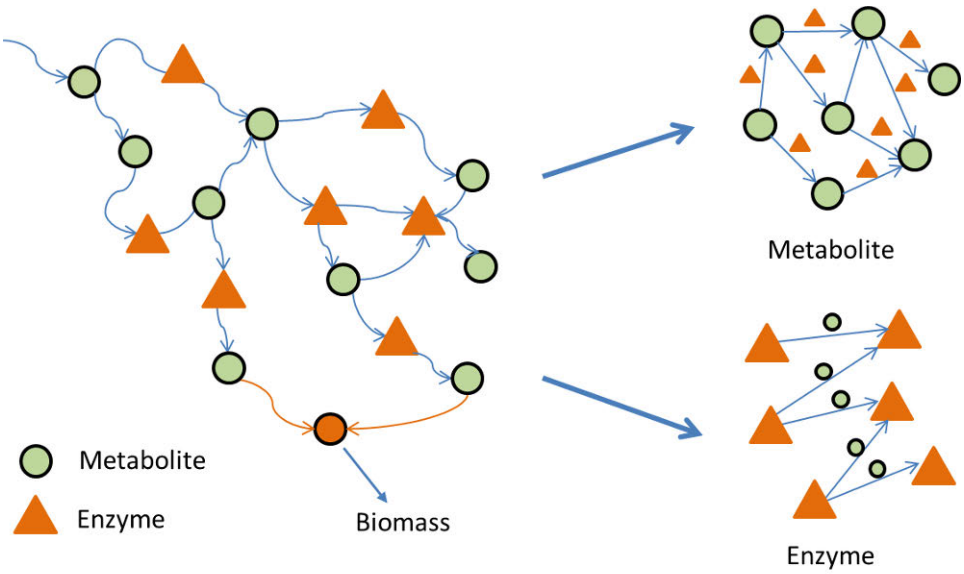


Figure 2.4: **A schematic representations of metabolic networks.** Considering both metabolites and enzymes centric network and relationships called metabolite-centric (top-right), and considering enzymes relationships called enzyme-centric (bottom-right) .

sociated links can help us uncover various critical mechanisms, and finding their pathological and metabolic origins. These findings also highlight that network-based methodologies can be an important tool for exploring and unraveling the interplay between human diseases and molecular networks [28].

Alteration in metabolic-related activities support maintainence of malignant properties and their survival. In addition, cancer cells reprogramme their metabolism to maintain the demand of unproliferated cells and the inhabitation of these cells in their changing micro-environments [31]. Cancer-associated metabolic alterations are present in all stages of cellular metabolic interactions, making it difficult to distinguish from the normal cellular proliferations [27]. However, various metabolic components target specific oncogenic signaling pathways, therefore it is important to determine the complex interaction patterns between oncogenic signaling pathways and metabolic interaction network [27]. Further, these cooperating interactions allow the cancerous cells to build their own micro-environments and fuel the nutrients for cancer cells [27, 18]. Such robust coordination inside the metabolic pathways supports the oncogenesis by simultaneous dysregulation of the PI3K-ATK-mTOR signaling pathways, damage of tumor suppressor genes, and activation of uncontrolled proliferation of oncogenes, survival, and alleviation

of cancer cells [18]. Also, oncogenesis alters the metabolite level in signaling pathways, affecting epigenetics and gene expression levels, and disregulation of metabolic enzymes per se. Thus, a new challenge of network-based computational studies is to elucidate the metabolic interactions between tumors and its host, and to decode how metabolic pathways support cancer cells' survival. Identification of critical metabolites can be used to optimize novel therapeutic interventions and to optimize the control principles of metabolic factors associated with oncogenesis.

2.3 Dynamics in cancer networks

Cancer is a complex disease, defined by its complexity and heterogeneity [19]. Cancer is very robust in nature, and as such, it has an innate capacity to adapt and proliferate in adverse conditions such as after anticancer treatments. Also it has the ability to develop cancer micro-environments with an immunosuppressive state. It is widely accepted that cancer is a result of somatic mutations [62], although the dynamics and the evolution of the cancerous cells are not well characterized, largely because of its innate complex systems [19, 82]. An alteration in a single gene doesn't activate full-blown cancer. For the oncogenesis and malignancy, a subsequent round of mutations is necessary [65, 44], which initiates mutation and cancer driver prognosis. So, the dynamics of cancer is defined by mutation, selection, and malignant tissue organization [65]. Also, cancer is a highly heterogeneous system which is key to enhance its robust interaction network through a specific subpopulation. The dynamics of this subpopulation enables the tumor to maintain its survival and prognosis even after anticancer drug successfully targeted some of this subpopulation [43]. The dynamics of cancer vastly depends on its genomic instability, which plays a key role in the evolution of genetic mutations and the maintenance [52]. The result of genetic instability and heterogeneity give rise to a distinct pattern of mutations and results in various cancer subpopulations within the sites [16]. The dysregulation flow in cancer is driven by various information channels altering the cell signaling pathways. These signaling pathways and modified intra cellular interaction networks (PPIs, metabolic interactions) drive tumor growth, energy production, and biomass production for cancer [19, 29, 18]. The dysregulation of signaling pathways and alteration of MN promotes the synergetic intracellular interactions and development of a tumor micro-environment, which constantly helps in the sustainability of tumor growth [19].

A number of studies have illustrated that in cancer, so called driver genes are responsible of initiating the cancer. These driver genes are caused by driver mutations which affect a number of intra cellular signaling and regulatory pathways [16, 94]. Cancer driver genes are mainly responsible for

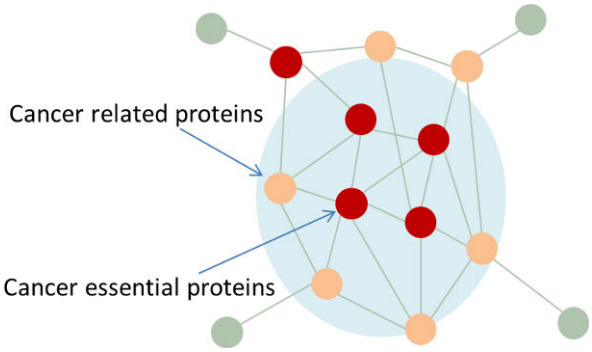


Figure 2.5: **Cancer essential proteins.** Role of cancer essential proteins and cancer related proteins in the networks. Cancer essential proteins (red colour) are a central part of the nework and have a functional role (hub and highly expressed in cancer) for driving proliferation in cancer.

the mutations which increase the malignant cell growth in the cancer microenvironment [92]. A driver gene has a high coverage in disease states and its mutations are well enough to disturb a/multi pathways. Also, a limited number of driver genes act as cancer drivers, and play as interconnection hubs in signaling circuits. The genetic screening of cancer reveals that some of the cancer driver genes govern critical processes for the establishment of mutations in cancer types. [3]

In cancer, some of the genes are identified as essential (for that cancer), if their functioning is essential for the multiplication or survival of those tumor cells. The concept of cancer essential genes impose that inhibiting these genes won't cause loss of functionality of normal cells. Certain genes become essential in cancer because of the presence of a mutation in the driver genes. This means that these genes are essential for pathogenesis (i.e., driver) [61], Figure 2.5, and the tumour becomes fully dependent on the development of oncogenes through these essential genes [61, 104]. Previous studies shows that cancer essential genes have lethal phenotype effects. Cancer essential genes control the cell cycle regulators and protein translation machinery [104], which directly damages the normal cell functions. Cancer essential genes are enriched with housekeeping functions involved in protein metabolism, DNA replication, mRNA processing [61]. Some of the cancer essential genes are included in the cyclin-dependent kinase production [70], promoting downstream consequence of another oncogene in cell lines [99], overexpression in cell lines [61]. Therefore, it is necessary to identify the genes which become essential for more effective therapeutics. Also, because the potential inhibition of only cancer essential genes don't affect the other genes and essential genes in healthy cells, these genes are suitable as new potential targets for antitumor therapy. Other findings show that

cancer essential genes can be an effective and direct target in cancer therapy [39].

Finding an effective target for cancer is always a challenge for current available therapeutic drugs because of the adaptive complexity of cancer. Although the dynamics and mechanism underlying in cancer are better understood than ever [19, 31], these therapeutics couldn't succeed to provide a cure in all patients. The side effects of chemotherapy and radiation therapy on healthy cells is well known, as well as targeting of kinase inhibitor is causing serious limitation and a problem of drug resistance [6]. However, cancer heterogeneity is a critical challenge for finding suitable targets, as well as hindrance in the discovery and validation of effective therapeutics biomarkers [19, 24]. Therefore, for a better identification of drug candidates that can be less toxic and maximize effectiveness in the adaptive nature of cancer, combination therapies can leverage a significant benefit in cancer. A combinatorial drug-target approach can explore the drug resistance problem by targeting multiple genes/ proteins simultaneously, and exploring the possibility of effective drug targets [39, 95]. Combination therapy showed promising results in a mouse model, where the use of PD0325901, rapamycin with MEK inhibitors, reduced the growth of prostate cancer [42]. This result highlighted that combination therapy *in silico* together with *in-vitro* validation might provide an effective treatment strategy in various cancers. In some cases, this approach didn't deliver significant benefits [95]. The challange is, how to effectively counter the complex dynamics of cancer, where still its genetic diversity is posing a formidable challenge to effective drug target in most cancer. Therefore, to develop the next generation of cancer therapeutics, computational-based approaches paired with molecular experimental techniques can significantly define the optimal combination in clinical oncology and be the realm of a new therapeutics to personalized medicine in cancer.

2.4 Network pharmacology for the next generation of drugs

Despite continuous progress in different therapeutic approaches, which allow us to better define the cellular dysfunction and mechanism inside disease signaling pathways, the success rate of new drug candidate approvals for clinical therapies is almost stagnant [59]. In particular, a very small percentage of drug candidates survive the late-stage attrition of drug development. These failures are mainly caused due to lack of efficacy in clinical testing, wrong selection of drug target, and drug toxicity [59, 33]. Although, the existing approaches work well in some cases, the continuous failure of this reductionistic approach in complex diseases indicates alternative approaches for drug

discovery. Focusing on only a single target knockout gene exhibits little or no effect on disease phenotypes and ignores its pathogenesis, see Figure 2.6. Network pharmacology improves this traditional approach by identifying the drug target and understanding its precise action on the disease-causing network [4], see Figure 2.7.

Cellular networks and interactions have illustrated that most of the cellular system dynamics are derived from the structure of their molecular networks [4, 11]. Dysfunction of disease is inherited in its module or subnetwork and therefore these disease proteins perturb the robust cellular phenotype [4, 11]. Increasing understanding of the structure of the disease networks provide valuable information to identify proteins whose perturbation can be a desired outcome for therapeutics [39], and discover drug target agents leads to perturb those proteins. Single target drugs may be able to modify some dysfunctional nodes of the disease module, yet, they could likewise modify the dynamics of the entities (i.e., proteins, metabolites, transcription factors etc.) that are arranged in the vicinity of the disease module, prompting significant side effects [4, 48]. The network-based perspective of drug target approach illustrates that most of the disease phenotypes are hard to invert using a single target protein, that is a part of an intervention by using a single target node in the network. However, to effectively identify targets through network analysis, we have to recognize interaction patterns of a particular drug as it has been revealed that some of the multiple patterns with a role in genetic deletion lead to unwanted side effect and toxicity [48, 33].

By using PPIs-based analysis, Wang and Loscalzo [97] explored the relationship between drugs, drug targets, drug interactors, and disease proteins to identify novel drug candidates in acute myocardial infarction, which has less side effect and toxicity. For this, they used the bipartite network of drug-targets, disease proteins and 12 drug-target disease modules for identifying novel insight into clinical therapeutics for this disease. In an another example, Li and colleagues [56] use flux balance analysis of metabolic networks in hyperuricemia-related purine metabolic pathways to identify potential drug targets. By using the steady flux balance reactions in the pathologic state, they determined the fluxes that are very effective in medication with the minimal side effects. The key fluxes have been identified by comparing and examining the fluxes of reactions during a change in the systems. In another approach [39], we identified the minimum number of drug target proteins in cancer PPI networks, needed to control the maximum number of cancer essential proteins in the network. We showed how to employ the use of well established drug-target proteins in order to achieve a structural control over essential target proteins within specific cancer protein-protein interaction networks, and apply this to breast, pancreatic, and ovarian cancer signaling transductions PPI networks. We demonstrated that instead of

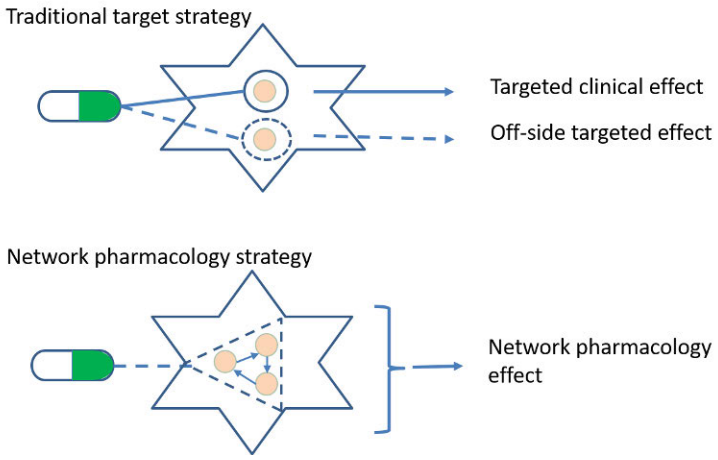


Figure 2.6: **Network pharmacology strategies.** Difference between traditional target strategy and network pharmacology strategy. In traditional approaches drugs target a single disease protein and leave other disease protein unharmed. Network pharmacology applies holistic approach and targets many of the disease specific proteins, candidate of disease specific proteins, disease modules and it's subnetwork.

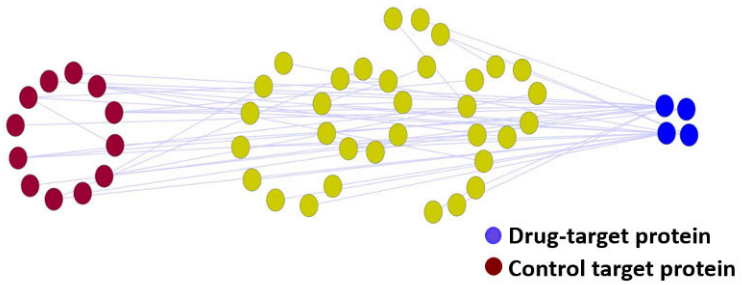


Figure 2.7: In a biomedical interaction network, network pharmacology approach shows how a small set of drug target proteins can be successfully used as multi-target on disease.

aiming for the overall control of the entire network, partial controllability is more effective and efficient in the development of therapies for various cancers. Therefore, it seems that network pharmacology has the potential to transform drug discovery and can improve clinical efficacy.

Chapter 3

Network Controllability

Control theory is a well established approach applicable in dynamical and complex networks. According to this theory, a dynamical systems is controllable if, with a reasonable selection of input sources, it can be driven from any underlying state to any desire final state inside a limited time [14, 58]. Recent advancement in network control approaches have offered a strong mathematical and computational framework to investigate the structural and functional relationship in a wide variety of networks, such as complex biological systems, social and mechanical systems, electric circuits, aircrafts, robotics [14, 26, 58, 12]. In complex biological systems, the network control framework provides meaningful insights and understandings by using different data sources to observe its structural and functional relationships [39, 101, 46]. By using this approach authors showed in [39] how this framework has been successfully used to identifiy a minimum number of drug targets needed to control a cancer PPI network, whereas in [101] authors have successfully predicted an important nervous function for a neuronal systems within C elegans. In this chapter, we discuss the network control framework, target controllability of linear networks and model constructions employed by us in order to find effective drug targets in cancer PPI networks, as well as the contribution of this appraoch in personalized medicine.

3.1 The network control framework

Network control theory is increasingly becoming a powerful tool in engineering and mathematics. Regulation and control are a central part of systems biology to understand its systems behavior. Therefore, an essential and ambitious query is how to successfully control system's behavior [58]. According to control theory, a dynamical system may steer the output of the system framework to a desired final state (e.g. target cancer essential protein in PPI network) via suitably picked inputs. Therefore, it is necessary

to understand the systems's dynamics expressed as equations which govern the system evolution over time. In PPI networks, this corresponds to a set of equations which describes the expression patterns at time $t+dt$ given the protein's inputs and current state at time t.

A *linear, time invariant dynamical system* having n states, m inputs and l outputs can be described by the linear equations:

$$\begin{cases} \frac{d(x)}{dt} = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases}$$

where A, B, C are matrices of size $n \times n$, $n \times m$ and $l \times m$ respectively, $x_t \in R^n$ is the state vector, $u_t \in R^m$ is the input vector and $y_t \in R^n$ is a output vector $t \in Z_{\geq 0}$. The matrices describe the complete interaction. Here, *states*, *input* and *output* dynamics are described by A, B and C respectively, while n, m and l are described as total number of variables of *states*, *input* and *output* systems respectively. Here, (A, B, C) imply the system with matrices A, B and C; if C is the identity matrix I_n, then C is omitted from (A, B).

An output state $y \in R^l$ is driven from an initial state $x = x_0 \in R^n$, denoted $x \longrightarrow y$ if there exists an input function $u_t \in R^m$ and some finite value t such that $y_t = y$. In a system (A, B, C), the output controllable subspace $y \in R^l | 0 \longrightarrow y$ is the vector subspace where all the values are driven from the initial state $0 \in R^n$. The dimension of the output controllable systems's subspace is denotes as $d(A, B, C)$.

The system (A, B, C) is called output controllable if its dimension *(A, B, C) = l = rank(C)*, and just controllable when $C = I_n$ [14].

Here, the output controllability matrix is denoted as

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

The following result is know as Kalman's rank criteria of output controllability.

Theorem 3.1.1: $(Kalman)$.

Given a linear time invariant dynamical system (A, B, C), we have

$$d(A, B, C) = rank\, OC(A, B, C)$$

that is, the system (A, B, C) is output controlable *iff rank OC (A, B, C) = rank(C)*.

An analogus concept can be extended to a group of systems, where systems within a group share a similar set of non-zero relationship. From this

perspective, we can consider structural equivalance of two or more systems by omitting the strength of its relationship and only consider their existances.

Matrices A and B are called *structurally equivalent* and denoted as $A \sim B$ if they share the zero values in the same postions, i.e.,

$$A_{ij} = 0 \textit{ iff } B_{ij} = 0 \text{ for all the entries of } A_{ij} \text{ and } B_{ij},$$

System (A, B, C) is *structurally equivalent* to $(A^{'}, B^{'}, C^{'})$ if $A \sim A^{'}$, $B \sim B^{'}$, $C \sim C^{'}$.

It is necessary to verify the equivalence relationship of structural equivalance systems. For this, the conventional properties shared among the equivalence class [A, B, C] of a given systems (A, B, C) can be examined. Traditionally, at whatever point the word generic or structural is used, it is comprehended that the corresponding property is considered among the equivalent class of a given framework. The maximum dimension among all output controllable subspaces of a system (A, B, C) is characterized as *generic dimension* $gd(A, B, C)$, i.e.,

$$gd(A, B, C) = max\{d(A^{'}, B^{'}, C^{'})|(A^{'}, B^{'}, C^{'}) \in [A, B, C]\}.$$

The results in *Structural (output) controllability* are similar to those in (output) control with $d(A, B, C)$ replaced by *gd(A, B, C).*

Note. We assume the initial condition of a system is always zero. Any general system can be reduced to this: one can consider a linear transformation $x \mapsto x - x_0$.

3.2 Graph theoretic approaches to control theory

Graph theory has advanced the algorithmic development for identifying the minimum size input controller u (and input matrix B) for a particular network and target. A first step in this direction was to discard the specific numerical setup inside the linear system, and characterize the intrinsic wiring diagram of the system's variables. According to control theory a linear time-invariant dynamical system (A, B, C) is *structurally target controllable* (related to a given size-k target set T) if there exists a time-dependent input vector $u(t) = (u_1(t), ..., u_m(t))^T$ and a numerical setup for the non-zero values inside the matrices A, B and C, that can drive the set T of target nodes to any desire final numerical setup in limited time. Also, it is well known, see e.g., [57, 86] that if a dynamical system is structurally (target) controllable, then it is (target) controllable in nearly all of numerical setups of the non-zero entries in A, B and C.

Linear systems can be defined as directed weighted graphs. The number n of variables defines the nodes of the graphs, and the non-zero values in the state transition matrix correspond to directed edges. Therefore, a directed edge is possible from node x_i to node x_j, with weight v, *iff* $A(x_j, x_i) = v \neq 0$. Likewise, a size-m controller vector can be defined as m *input nodes* $u_1, ...u_m$, also called *driver nodes*, while the edges in between the driver nodes and the network are determined by the input matrix B. Therefore, the directed edges are possible from u_i to x_j, with weight w, *iff* $B(x_j, u_i) = w \neq 0$. The nodes x_j are called the *driven nodes* in the network, if there exist i with $B(x_j, u_i) \neq 0$. These driven nodes are directly capable in the network to drive the whole system to the desire final state.

In [57], it was shown that structural controllability approach has a counterpart formulation in terms of network graphs. The system (A, B) is structurally controllable from the input controller m and control matrix B *iff* we can choose a set of n directed path from the driver nodes (here, this set is denoted as $\mathcal{U}$) to every nodes in the network (i.e., as ending points), such that no two paths would cross at a point at the same distance d from their end points. This formulation about structural controllability is related to the idea of *linking* and *complex graph* investigated in [78, 67]. For the target controllability problem an target set $T = \{t_1, t_2, \ldots, t_k\} \subseteq X$, the graph formulation can be naturally adjusted and described as follows. We introduce the k new *output nodes* $\mathcal{C}_\mathcal{T} = \{c_1, c_2, \ldots, c_k\}$ (also introduce as $\mathcal{C}$) and edges (t_i, c_i), for all $1 \leq i \leq k$. Important is that the above wiring diagram described as output matrix C_T. Here, objective is to find a family of paths contains k directed paths, and able to connect all the available driver nodes (as input points) to the output nodes (called the final-point), so that no two paths can intersect at the same distance d from their final points. Nevertheless, in contrast to full controllability, the above graph condition is only essential for target control, but not sufficient [67].

From an algebraic point of view, the driver nodes represent the nonzero columns in the control matrix B, while the nonzero rows of B correspond to the driven nodes. Following the above mentioned criteria, we analyze the structural controllability of *linear, time invariant system (LTIS)*/dynamical-networks by trying to minimize the total number of driver and driven nodes.

In reality, an input controller (driver node) is able to directly influence several nodes at the same time. This type of influence in a system leads to the direct interaction of the input controller to its elements.

Definition 1. We say that in a given *LTIS* (A, B, C_T) the input controller is $N - bounded$ *iff* control matrix B contains the maximum N non-zero value on every column.

3.2.1 Full controllability

For full controllability, a network requires the minimum number of input (driver) nodes, which could be determined through the maximum matching. In a directed graph, a node is defined as matched if it is the endpoint of an edge in the matching set, else it is unmatched. Moreover, full control over a directed network is possible if and only if we managed to control the unmatched nodes in the network and there is a direct link from the input nodes to all the matched nodes. The maximum matching of any directed network can be identified in the maximum $O(N^{1/2}L)$ steps [32], where L denotes the available links (i.e., edges) in the networks.

3.2.2 Target controllability

In this section, we describe the various approximation algorithms used for the target control problem, trying in to reach the optimal solution in a time-efficient manner. The schematic representations of target controllability in linear networks is defined in Figure 3.1.

The first greedy algorithm used for the target control approach has been derived by Gao et al. [26]. The approach to solve the problem is by trying to create a linking inside the associated network, which is called the *dynamic graph*; this method is based on studies of Poljak and Murota [78, 67].

Here, we briefly describe the approach mentioned in Poljak and Murota [78, 67]. Further, we present the approximation algorithm for target control problem mentioned in [26], and analyze this algorithm. Finally, we describe our algorithms based on heuristic optimization used in several case studies (cancer signaling PPI networks).

Let (A, B, C_T) be defined as an LTIS over n variables, m inputs, and l targets (i.e., $|T| = l$), and $G = (V, E)$ is the associated graph. Then, the dynamical graph $\overline{G}$ is described as a time-disjoint representation of the graph G. $\overline{G}$ is called time-disjoint, if in each state (from $t = 1$ to $t = n$) and each input variable (from $t = 0$ to $t = n - 1$) is defined as a different node at distinct time points, whereas time-point $t = n + 1$ is only associated with target states.

Generally, a graph $\overline{G} = (\overline{V}, \overline{E})$ is defined with $\overline{V} = \overline{V_A} \cup \overline{V_B} \cup \overline{V_C}$, where

- $\overline{V_A} = \{v_{i,t} \mid i = 1..n,\ t = 1..n\}$,
- $\overline{V_B} = \{v_{n+j,t} \mid j = 1..m,\ t = 0..n-1\}$, and
- $\overline{V_C} = \{v_{n+m+k} \mid k = 1..l\}$.

Here, the node in $\overline{V_C}$ correspond in an one-to-one relationship with the nodes V_C and target T. The set of edges $\overline{E}$ in grpah $\overline{G}$ is defined as follows:

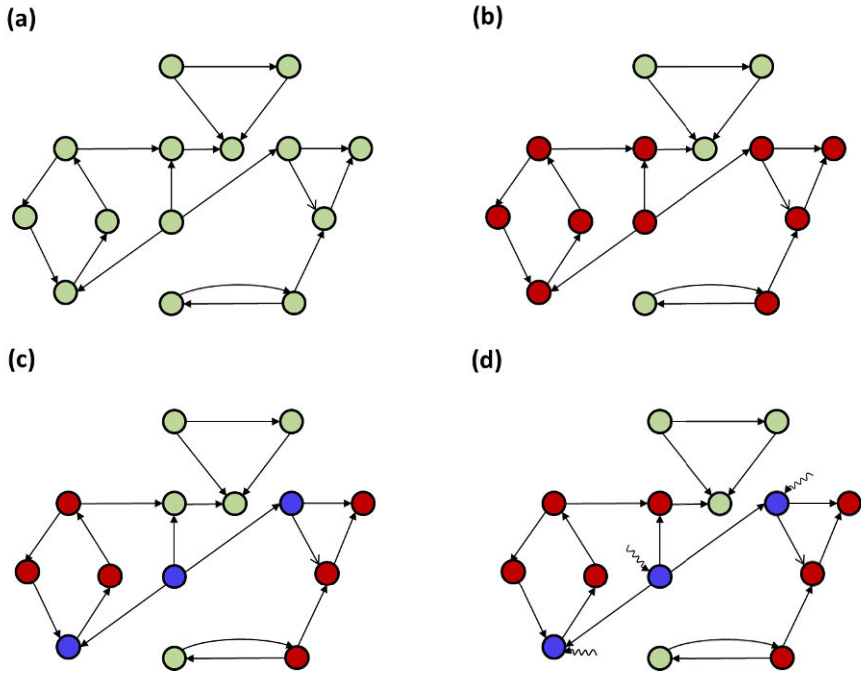


Figure 3.1: **Target controllability of a linear network** The main steps in the target controllability approach: (a) Construct a linear network model for the system. (b) Identify the part of the network that should be controlled (red nodes), e.g. targeted nodes. (c) Compute the set of actionable control nodes (blue). (d) Engineer the control nodes to drive the network into a more favorable dynamics and internal state (indicated with jiggled lines) .

- $\{(v_{j,t}v_{i,t+1}) \mid \text{for all } i \text{ and } j \text{ such that } A_{i,j} \neq 0,\ t = 1..n\} \cup$
- $\{(v_{n+j,t}v_{i,t+1}) \mid \text{for all } i \text{ and } j \text{ such that } B_{i,j} \neq 0,\ t = 0..n-1\} \cup$
- $\{(v_{j,n}v_{n+m+i}) \mid \text{for all } i \text{ and } j \text{ such that } C_{i,j} \neq 0\}$.

In a dynamical graph $\overline{G}$, a group $L = (p_1, p_2, ..., p_k)$ of k edge-disjoint paths is defined as a linking of size k. We say L is a linking for (S, T), if the set of initial and terminal nodes in the path L is defined as $S, T \subseteq \overline{V}$. In [67], it has been derived that if (A, B) is an LTIS with m driver nodes and size-l of target set T which is controllable from available driver nodes, then there is an $(\overline{V_B}, \overline{V_C})$-linking of size l. It has been an inquiry for a long time whether the opposite of the above outcome additionally holds. To be specific, if for a LTIS (A, B, C_T) there is a $(\overline{V_B}, \overline{V_C})$-linking of size l, does it suggest at that point that the size-m driver set related to B is controlling the objective T, i.e., $\text{rank}OC(A, B, C_T) = l$? In spite of the fact that the response to this inquiry was turned out to be negative [67], it turned out

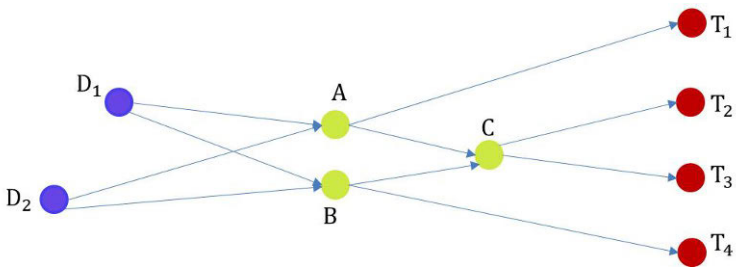


Figure 3.2: Target controllability of directed graph. The targeted structural controllability problem for the directed graph $G = (V_A, E_A)$ with n nodes and a subset $T \subseteq V_A$ with I target nodes, is equivalent with deciding if there exists a set of l directed paths in G such that each node in T is an end point of one such path and no two paths intersect at the same distance from their end points, [57]. In this example, the paths from the driver nodes D_1, D_2 to the target nodes $T_1 - T_4$ intersect in the internal nodes A, B, and C. The controllability theorem of [57] implies that the lengths of the paths CT_2 and CT_3 is different, and that either the length of the path $A \Rightarrow T_1$, $A \Rightarrow T_2$, and $A \Rightarrow T_3$ are pairwise different, or the length of the path $B \Rightarrow T_2$, $B \Rightarrow T_3$, and $B \Rightarrow T_4$ are pair-wise different (or both).

to be evident that any counter-claim for this case must comply with some exceptionally strict structure conditions in regards to the controlling way from the driver nodes to the target.

In [26], author employed the above approach and presented a greedy algorithms for the structural control issue. To be specific, given an LTIS A and a target T, their algorithm scans for a set V_B for which there exists a $(\overline{V_B}, \overline{V_C})$-linking. In turns, such a set V_B would have a high likelihood for characterizing a set of driver nodes for the objective T. However, after applying this algorithms, one needs to perform a validation step which checks whether the set of driver/driven nodes chosen by the algorithms are to be sure controlling the target. This can be achieved by examining that the rank of the controllability matrix $OC(A, B, C_T)$ is indeed equal to $|T|$.

In the next section, we present the detailed description of the algorithms in [58] and introduce our algorithms based on heuristically improved variants of it. We have also includeded some of the algorithm presented in our published articles.

Basic target control algorithm

Let A be is an LTIS of n variables and $G = (V_A, E_A)$ be the directed graph associated to it. Let's assume $T \subseteq V_A$ is a set of target nodes (the schematic representations of target controllability of directed graph is mentioned in

Figure 3.2). In the following algorithm the output is denoted as a set of driven nodes D which has direct correspondence to the examined set V_B and there exists a $(\overline{V_B}, \overline{V_C})$-linking.

Step 1: Let $i = 0$, $C^i = T$, and $D = D^i = \emptyset$.

Step 2: Define a bipartite graph G_{bi} whose nodes are $L \cup R$, where $L = V_A$, $R = C^i$, and any node appearing in both V_A and C^i is considered separately in L and R. For $l \in L$ and $r \in R$ there exists an edge (l, r) in G_{bi} *if and only if* $(l, r) \in E_A$ is edge in the initial directed graph G.

Step 3: Find a maximum matching set of (M_L, M_R) in G_{bi}, $M_L \subseteq L$ and $M_R \subseteq R$, and derive that $C^{i+1} = M_L$ is the set of the left sided matched nodes and $D_i = R \setminus M_R$ is the set of right sided un-matched nodes. Let $D = D \cup D_i$.

Step 4: Consider C^{i+1} as a new set of target nodes. If $C^{i+1} = \emptyset$ then algorithm complet and output is D. If not, then proceed for Step 5.

Step 5: If $i < n$ then $i = i + 1$ and continue to Step 2 with the updated target C^i and driven set D. Otherwise, proceed to Step 6.

Step 6: Output D defined as the set of driven nodes.

Note: If the algorithm mentioned above completes in step 6 then it implies that the target set C^n is non-empty. Since n is the total number of nodes count in G, it indicates that rest of the nodes in C^n can be separated into distinct cycles. Nodes remain in this cycles and includes in C^n, can be controled from any driver (input) node. However, in some cases a number of driven nodes require to be added. Another possibility to add the remaining nodes in C^n is as driven nodes in D.

3.2.3 Heuristic target control algorithm

In the basic target control algorithms if a node x is chosen as a driven node, i.e., included in D in Step 3, we don't check whether node x showed up before in some past control path. In this case, since we realize that node x is chosen for being a driven node, we can shorten that control path subsequent to achieving node x. In step 3 of the above mentioned algorithm, at every iteration of the search procedure we locate a maximum matching in between the nodes of G and the present target C^i. Although, such maximum matching probably won't be unique, in which case a few of these maximum matching may be more reasonable to be chosen.

Let's assume that i is some iteration process for the search procedure in the algorithm, and $C^1, ..., C^i, D^1, ..D^{i-1}$ and D is the computed sets of

targets and driven nodes. Let bipartite graph G_{bi} built in iteration i, with nodes denoted as $L \cup R$, where $L = V_A$, $R = C^i$, where any new node appearing in V_A and C^i is treated differntly in L and R. While, searching of maximum matching (M_L, M_R) in G_{bi} by using $M_L \subseteq L$ and $M_R \subseteq R$, we set the following heuristic criteria for directing the procedure towards a minimum number of driven nodes. Note, not all criteria underneath can be followed in the same time.

Algorithm: Within Step 3.1 of the basic target control algorithm, select a maximum matching (M_L, M_R) following the Criteria 1, 2, 3, and 4, in this exact order of significance.

- Criterium 1: When calculating the maximum matching (M_L, M_R), maximize the use of the earlier defined driven nodes in M_L.
- Criterium 2: During calculation of the maximum matching (M_L, M_R) try to ignore the formation of cyclic controlling path. Especially, ignore selecting nodes $x \in M_L$ such as there exists $j \leq i$ and a sequence $u_{i+1}, ..., u_j$ such that $u_k \in C^k$ for all $j \leq k \leq i$, $u_{i+1} = u_j = x$, and for all $j \leq k \leq i$, u_k is connected to u_{k+1} in the associated bipartite graph.
- Criterium 3: Further, compute the maximum matching (M_L, M_R), and maximize the use of nodes in M_L which have appeared previously in $C^j, j < i$, on a path that is controlled earlier (ends with a driven node).
- Criterium 4: Compute the maximum matching (M_L, M_R), and maximize the use of nodes in M_L which have appeared in previous $C^j, j < i$, on a path which is not controlled earlier.

Similarly to basic target control algorithm, heuristic target control algorithm follows certain steps to fulfil the above mentioned criteria;

Step 1: (Similar to basic target control algorithm): Let's define $i = 0$, $C^i = T$, and $D = D^i = \emptyset$.

Step 2: (Similar to basic target control algorithm): Define a bipartite graph G_{bi} with nodes $L \cup R$, where $L = V_A$, $R = C^i$, and any node appearing both in V_A and in C^i is treated differently in L and R. For $l \in L$ and $r \in R$ there exists an edge (l, r) in G_{bi} if and only if $(l, r) \in E_A$ is an edge defined in the initial directed graph G.

Step 3.1: Following the criteria above, compute a maximum matching (M_L, M_R) in G_{bi}, $M_L \subseteq L$ and $M_R \subseteq R$, and denote it $C^{i+1} = M_L$ to be the set of the left side of matched nodes and $D_i = R \setminus M_R$ be the set of right side of un-matched nodes.

Step 3.2: For each $x \in D_i \setminus D$, do:

- If node x is already present in any previous C^j, $j < i$, then it is good to discard the entire control path from that exist (in C^j) onward, and update all the sets C^k, D^k with $j \leq k \leq i+1$ accordingly. Then update D as $D = \bigcup_{p=0,..,i} D^p$.
- End For (from Step 3.2)

Step 4: Then consider $D = D \cup D^i$ as a new set of driven nodes, and $C^{i+1} \setminus D$ as a new set of targets. If $C^{i+1} = \emptyset$ then we finish the algorithm and output D. Otherwise, proceed to Step 5.

Step 5: (Similar to basic target control algorithm): If $i < n$ then $i = i + 1$ and proceed to Step 2 with the upgraded target node C^i and driver set D. Otherwise, proceed to Step 6.

Step 6: Add all the remaining nodes in C^n one by one to the driven set D and, for each new addition to D, carry out the check from Step 3.2, i.e., shorten the already controlling path for each new addition in D.

Step 7: Output D as the set of driven nodes.

3.2.4 Target control with preferred operators

Building on the approximation algorithm from [14] for structural target controllability we introduced a new algorithm for structural target control with preferential operators. Namely, given a directed network (e.g., a protein/gene signaling network), a set of target nodes (e.g., a set of disease-specific essential genes), and a set of preferential operators (e.g. a set of genes/proteins known to be directly targeted by specific drugs) all within the network, find a close to minimal set of nodes that maximizes the use of the available operators, in order to control the targets. We detail bellow this algorithm.
Note: Given two sets A and B, we denote by $A \cup B$ and $A \sqcup B$ the union and disjoint union, resp., of these sets.

Let $G = (V, E)$ be a directed graph, let $T \subseteq V$ be the set of target nodes, and let O be a set of preferential operator nodes. We construct a sequence of sets $C^i, D^i, i \geq 0$, (and $i \leq |V|$) with $C^0 = T$, $D^0 = \emptyset$, and $|C^i| \geq |C^{i+1}|$, such that the union set $\bigcup_{1 \leq k \leq i} D^k$ is a set of nodes controlling the target T, where the use of the nodes in O is maximized in the generation of D; we refer to [14] for the explanation on why this claim holds.

The target control algorithm for target structural control with preferred operators:

Step 1 i:=0, $C^0 := T$, $D := D^0 := \emptyset$

Step 2 Define the bipartite graph $G^i = (L^i \sqcup R^i, E^i)$, where $L^i = V, R^i = C^i$, and E^i contains edges $(l, r) \in (L^i, R^i)$ such that $(l, r) \in E$ is an edge also in the initial graph.

Step 3 Find a maximum (cardinality) matching (M_L^i, M_R^i) (following the 6 heuristic criteria below) in G^i, where $M_L^i \subseteq L^i$ and $M_R^i \subseteq R^i$, and let $C^{i+1} = M_L^i$ be the set of the left sided matched nodes and $D_i = R^i \backslash M_R^i$ be the set of right sided un-matched nodes.

Step 4 For each $x \in D_i \setminus D$

Step 5 If $x \in \bigcup_{j<i} C^j$ (i.e., x appears in any previously computed C^j, $j < i$)

Step 5.1 remove the entire control path from that occurrence (in C^j) onward, and update all the sets C^k, D^k with $j \leq k \leq i+1$ accordingly.

Step 5.2 Update D as $D := \bigcup_{0 \leq p < i} D^p$.

End If (from Step 5)

End For (from Step 4)

Step 6 Update D as $D := D \cup D^i$ and C^{i+1} as $C^{i+1} := C^{i+1} \setminus D$.

Step 7 If $C^{i+1} = \emptyset$ then output D as a set of control nodes for T and stop. Else proceed to Step 8.

Step 8 If $i < n$ then $i := i + 1$ and proceed to Step 2. Else, proceed to Step 9.

Step 9 For all the remaining nodes in C^n, add them one by one to the driven set D and, at each new addition to D, perform the check from Step 5, i.e., pruning the existing controlling path for each new addition in D.

Step 10 Output D as a set of control nodes for T.

On Step 3 above we mention 6 heuristic criteria for implementing a maximum (cardinality) matching in between the left, $L^i = V$, and right, $R^i = C^i$, disjoint sets of the bipartite graph G^i. This is due to the fact that the maximum matching might not be unique, and, depending on which maximum matching we chose, the size of the final set controlling the target nodes can differ significantly. Also at this point in the algorithm we can intervene so that a maximal amount of preferred operators is chosen as actual driven nodes. In the following we are introducing this set of 6 heuristic criteria.

- Criteria 1: All preferred nodes from O appearing in a control path are directly controlled, i.e., the maximum matching is performed between sets $L^i := V$ and $R^i := C^i \setminus O$) while $D^i := D^i \cup \big(C^i \cap O\big)$,

- Criteria 2: Maximize the use of already driven nodes in M_L^i.

- Criteria 3: Maximize the use of preferred operators in M_L^i. This is done by initiating the maximum matching algorithm by a first (maximal) matching which maximizes the number of pairs $(x, y) \in (M_L^i, M_R^i)$ where x is a preferred operator,i.e., $x \in O$.

- Criteria 4: Try to avoid the creation of cyclic controlling path. That is, avoid selecting nodes $x \in M_L^i$ such that there exists $j \leq i$ and a sequence $u_{i+1}, ..., u_j$ such that $u_k \in C^k$ for all $j \leq k \leq i$, $u_{i+1} = u_j = x$, and for all $j \leq k \leq i$, u_k is matched to u_{k+1} in the corresponding bipartite graph.

- Criteria 5: Maximize the use of nodes in M_L^i which have appeared in some previous $C^j, j < i$, on a path that is already controlled (ends with a driven node).

- Criteria 6: Maximize the use of nodes in M_L^i which have appeared in some previous $C^j, j < i$, on a path that is not controlled yet.

3.3 Target controllability of cancer PPI network

In [39] we applied the target controllability algorithm (with preffered operator) to PPI signaling interaction networks on breast, pancreatic and ovarian cancer, and identified respective sets of driver (input) proteins for controlling the particular networks. We identified a set of proteins in the network called "cancer essential proteins"[39], whose functioning is essential for the proliferation or survival of those tumor cells. The concept of cancer essential proteins impose that inhibiting these proteinss won't cause loss of functionality of normal cells. Therefore, rather than trying to accomplish full control of the whole network, which in itself is very unpredictable in complex biological networks, our approach is based on target control approach to control those cancer essential proteins. We identified that in order to control the entire set of available essential proteins inside the networks, through our approach it needs to put the direct intervention on only 6.6 to 13 % of whole networks node, while for full control it needs around 70 % of networks' nodes [39]. Furthermore, we analyzed the topological features of the available driver drug-target proteins inside the networks. We concluded that driver drug-target proteins have respectively high average degree. This confirms that drug-target proteins have multiple interactions

inside the networks and can be feasible in the application for control over essential(target) nodes. We observed that some of the driver drug-target proteins are oncogenes and expressed in multiple cancer and could have a high impact in case of therapeutic effects in these cancers.

3.4 Data resources

Here, we describe the data which was used during our analysis and in our case studies. We show how our network frameworks enable to integrate multiple types of biomedical data for deeper mechanistic and molecular insight.

3.4.1 Cancer data

The protein data used in our study for various cancers are mainly collected from UniprotKB [13] and various previous published articles. These data are a list of protein IDs. For the authenticity of our studies, we collect only those proteins which are reviewed and used for previous studies. We have collected the genome-scale metabolic model (GEMs) for breast cancer from Human Metabolic Atlas [79]. Further GEMs have been used to build the metabolic network for the analysis. We performed our study over breast, pancreatic and ovarian cancer for directed PPI signaling network and breast cancer for the metabolic network.

3.4.2 Essential protein data

In cancer, some of the genes are identified as essential (for that cancer), if their functioning is essential for the multiplication or survival of those tumor cells [61]. The concept of cancer essential genes impose that inhibiting these genes won't cause loss of functionality of normal cells. Therefore cancer-specific essential genes could used for effective drug targeting. Certain genes become essential in cancer be used because of the presence of a mutation in the driver genes. This means that these genes are essential for pathogenesis. We collected these cancer essential data for breast, pancreatic and ovarian cancer from the COLT-Cancer database [45]. This database has cancer essential proteins for a total of 72 cell lines for breast, pancreatic and ovarian cancer. For our studies, we have considered 29, 23, and 15 cell lines data respectively for breast, pancreatic and ovarian cancer.

In particular, we considered the MDA-MB-231, HPAF-II and OV-90 cell lines respectively for breast, pancreatic and ovarian cancer, and follow the GARP (Gene Activity Rank Profile) and GARP-P value of corresponding proteins mentioned in the database. Since previous studies [61] showed that proteins with lower GARP scores tend to be classified as essential and directly associated with oncogenesis. We selected only those essential proteins

whose GARP value is in the negative range, and whose GARP-P value is less than 0.05 ($p \leq 0.05$).

3.4.3 Drug target data

We collected drug-target protein data from the open-source DrugBank [98] database. The DrugBank database provides comprehensive and freely accessible information on drug and drug targets. For our analysis, we have collected the FDA-approved drug-target proteins.

Our network controllability heuristics algorithms for target controllability of cancer networks [39] use drug target proteins as input nodes. After this selection, the algorithm gives first preference to initiate the control pathways starting from such nodes (for details of the algorithm, see subsections 3.2.2 and 3.2.4). It later maximizes the number of drug target proteins offered by our network controllability algorithm for better treatment of cancer by using our networks.

3.4.4 Interaction data and network construction

We created the directed PPI signaling network from SIGNOR (SIGnaling Network Open Resources) databases [77], which generates binary matrix representations for the user-provides protein list and allows to create directed graphs between signaling entities. Interactions provided here are based on the *directed* influence, such as activation and inhibition of protein. For building the directed PPI signaling network for each cancer type, we individually uploaded the data as a list for each study which includes cancer proteins, cancer essential proteins and drug target proteins in the SIGNOR database. Next, SIGNOR generated PPI signaling networks for each cancer.

Further, to build the breast cancer metabolic network, we construct the stoichiometric matrix (S). Here S denotes metabolites as rows and reactions as columns. It parses and adds an edge based on the sign change value.

3.4.5 NetControl4BioMed

In our web-based pipeline NetControl4BioMed [40], we build automatically biological networks based on data from KEGG [38], WikiPathways [49], Pathway Commons [9], and SIGNOR [77]. These datasets have their own formats, therefore it makes it challenging to integrate and use all of them as a single network. Firstly we generated the networks from each dataset, and then integrate them into a big single network.

This pipeline is developed based on the Anduril workflow framework [72]. It is an open-source platform for biomedical data analysis. This platform allows integrating a range of software analysis and algorithm, and computational simulation tools into a single data analysis pipeline. Our pipeline used

the Moksiskaan platform [50] to create biomolecular interaction networks. Moksiskaan integrates PPI, genome, pathways and literature mining data into a network, from a given set of input nodes. Therefore, this pipeline is a nice example of the integration of multi-types of biomedical data and powerful network controllability algorithm, and analysis of biomedical data.

The network is generated by combining seed nodes provided by the user. Seed nodes are list of proteins which are used by Moksiskaan to generate the network in the pipeline. Further, seed nodes define the network-based of all known paths within the network, whose length is not surpassing the *gap* value. The *gap* is a parameter that maximizes the number of intermediate nodes inside the networks between the seed nodes. A higher *gap* value allows the network to grow quickly. The $gap = 1$ means that additionally to the edges between the seed nodes, also path going through one extra node (to be added to the network are included). Users may use our given set of cancer essential proteins for breast, pancreatic and ovarian cancer as target nodes or have their own set of target nodes. For input nodes, users may use our given set of drug-target proteins or use their own data.

The network controllability algorithm applied in the NetControl4BioMed pipeline generates as a result a set of control target nodes that are controlled from a set of input nodes. It provides a list of control pathways from input nodes (drug targets) to target nodes (cancer essential proteins). These results may be used to find possible action mechanisms of these drug targets in the specific context of those particular targets. The pipeline generates distinct control pathways for the same input node if it controls multiple target nodes within the network. Further, the pipeline generates *.xml* file which can be uploaded in Cytoscape [84] and used for the visualization as well as for the integration of the network with other data.

3.5 Other approaches

3.5.1 Minimum Dominating Set

In a network $G = (V, E)$, a set $S \subseteq V$ of nodes is defined as a minimum dominating set (MDS), if every node $n \in V$ is either an element of S or adjacent to an element of S. In other words, a set of nodes is called an MDS if it can reach to the remaining node in the network by one interaction. The MDS uses to find the key driver nodes that can control the whole network. MDS model is used in structurally control of complex networks where each node is covered by at least two nodes in the MDS set [69]. Also, the MDS approach can be used for various types of complex dynamic networks, analyze the controllability of networks. MDS is not always uniquely determined, this is because of the presence of multiple MDS. Therefore, to adapt to this non-uniqueness issue, the ideas of basic, irregular and repetitive nodes were

applied to the MDS [68]. Some key nodes are available in all MDS, whereas nodes that belong to few but not all MDS are determined as intermittent, and nodes that are not available in any MDS are called redundant [68].

3.5.2 Feedback control system

A feedback control loop is a well defined and powerful tool for controlling a system. Generally, feedback control is applied as a control system when the output is taken into consideration and it enables the systems to produce the performance to compare the real output with the desired output response. The meaning of "feedback" is that some part of output is returned back to the input. The feedback loop is designed to automatically engineer the systems so that it can achieve and maintain the ambitious output action by comparing it with the real condition. More, a feedback loop system is a fully automated control framework, where the control is being dependent on its output.

Chapter 4

Summaries of the included articles

4.1 Paper 1: Controlling Directed Protein Interaction Networks in Cancer

- **Krishna Kanhaiya, Eugen Czeizler, Cristian Gratie, and Ion Petre. "Controlling directed protein interaction networks in cancer." Scientific reports 7, no. 1 (2017): 10327.**

Advances in systems biology are offering not only insights into complex molecular interactions but are also useful for the discovery of new disease proteins and of new therapeutic targets for disease intervention. Here we employ a control theory approach for the analysis of specific disease networks, allowing us to drive the system dynamics towards favorable traits, as well as helping us to understand better the regulatory mechanisms of these biochemical networks. We show how to employ the use of well established drug-target proteins in order to archive a structural control over essential target proteins within specific cancer protein-protein interaction networks. We apply this to breast, pancreatic, and ovarian cancer signaling transductions PPI networks. We demonstrate that instead of aiming for overall control of entire networks, partial controllability is more effective and efficient in the development of therapies for various cancers.

4.2 Paper 2: NetControl4BioMed: a pipeline for biomedical data acquisition and analysis of network controllability

- **Krishna Kanhaiya, Vladimir Rogojin, Keivan Kazemi, Eugen**

Czeizler, and Ion Petre. "NetControl4BioMed: a pipeline for biomedical data acquisition and analysis of network controllability." BMC bioinformatics 19, no. 7 (2018): 185.

Network controllability for biomedical networks focuses on finding of combinatorial approachs for which external interventions within a biological system can drive it to a desired final configuration. In practice, this approach converts into finding of muti drug-target therapeutics for discovering and development of novel and rational therapeutics approaches for complex and dynamics diseases like cancer. We develop a novel biomedical data analysis pipeline called *NetControl4BioMed* based on the network control approach using linear networks. Our pipeline produces novel biomolecular interaction networks by joining pathway information from different open databases by using user's query. The pipeline further distinguishes a set of nodes that is sufficient to control a given, users defined set of essential proteins related to disease in the networks i.e., it can induce a transition of the network's configuration from initial states to desire final states. Also, we provide new insights into the efficient control of dynamical disease networks which can assist in the discovery of novel cancer-associated proteins and biomarkers. Users can use our pipeline online as well as install the source code to run locally. This pipeline can be useful in defining the controllability and understanding the complex biomolecular interactions and for combinatorial multi-drug therapies for more effective therapeutic strategies and personalized medicine.

4.3 Paper 3: Structural Target Controllability of Linear Networks

- **Eugen Czeizler, Kai-Chiu Wu, Cristian Gratie, Krishna Kanhaiya, and Ion Petre. "Structural target controllability of linear networks." IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 15, no. 4 (2018): 1217-1228.**

Computational modelling of the structure of cellular interaction networks reveals many important novel therapeutics targets for complex disease like cancer. Recent research on network analysis shows that network control theory is increasingly becoming a powerful tool apply in engineering and mathematics. Also, regulation and control are the central part of the biological system to understanding its systems behavior. Previously, by Liu et al. [58] has presented a polynomial time algorithm for optimal solution of full controllability. Later, Gao et al. [26] proposed a greedy algorithm for finding the minimum number of input nodes needed to impose a certain target

controllability. However, the full complexity of target control optimization problem hasn't been successfully handled. We found that in all the practical setup target controllability problem perform as NP-hard, i.e. when the controllability of individual input in a system is bounded by some constant. Further, we show that greedy algorithm provided in [26] fails to generate correct solutions in some cases, and needs extra validation steps. We show that our algorithms which are based on heuristic optimization strategies are more effective in several case studies(cancer signaling PPI networks) and for faster calculations and effective optimization.

4.4 Paper 4: Identification of drug targets in breast cancer metabolic network

- **Krishna Kanhaiya and Dwitiya Tiwari. "Identification of drug targets in breast cancer metabolic network." JOURNAL OF COMPUTATIONAL BIOLOGY, Volume 26, Number 0, 2019.**

Genome-scale metabolic models have been proven to be valuable for defining cancer or to indicate the severity of cancer. However, identifying effective metabolic drug-target (DT) of the active small-molecule compound is difficult to unravel and need to be investigated. In this study, we identify effective DT for breast cancer using proposed network analysis of enzymes-centric networks in the metabolic model. Our network-based analysis revealed that high degree nodes(HDN) of enzymes are key to progression/development of cancer. These HDN show highly interconnections inside the network. It has been found that these HDN are crucial driver nodes for effectively targeting in breast cancer metabolic network. Furthermore, based on the correlation and principal component analysis (PCA) we have shown that certain proteins play a significant role in the network and can be used as an effective DT in cancer therapeutics. More, these proteins stimulate the active site of enzymes to activate the target metabolites. Overall, we have shown that a better understanding of the metabolic networks using statistical model could be valuable in DT identification for developing effective therapeutic approaches and personalized medicine.

4.5 Paper 5: Bioinformatics for Diseases Management: A Personalized Therapeutics Prospective

- **Krishna Kanhaiya. "Bioinformatics for Diseases Management: A Personalized Therapeutics Prospective." In Ad-**

vances in Personalized Nanotherapeutics, pp. 187-199. Springer, Cham, 2017.

Advances in multi-omics technology and development of modern bioinformatics and integrated tools change the way of analyzing and understanding of complex disease mechanism and transform the healthcare sector towards smarter disease diagnostic and management. These advancements bring a high amount of data for the doctors and researchers in the form of genomics and proteomics which can further deliver for computer-aided therapeutic solutions for screening and early diagnostic of the patients. Therefore, a proper method for integration and management of biomedical data sets could be applied into cost-viability, high-value and rational drug therapeutics for effective personalized treatment. Also, it could decrease diagnostic expenses, improve individual patient care and help the doctor to create individual based patient care. This review describes an overview of integrated bioinformatics approaches to define effective disease management techniques in the role to define next generation disease management facilities, which can further establish accurate, reliable, safer healthcare for each and every patient.

Chapter 5

Conclusion and future work

Complex dynamics of cancer remains at the forefront of the quest to understand the structure and function of this disease and also offer an exceptional model system in perspective of controlling of other disease networks. We applied the target control theory approach for the specific cancer signaling transduction PPI networks. We find that this approach can be an impressive framework to discover effective drug-target proteins as driver nodes to target cancer-specific essential proteins, defines as target nodes. The concept of cancer essential proteins impose that inhibiting these proteins won't cause loss of functionality of normal cells. Further, these proteins are very likely to affect the real behavior of cancer networks and the tumour becomes fully dependent on the development of oncogenes through these essential genes [61, 104]. In [39], we identify the minimum number of drug target protein in cancer PPI network to control the maximum number of cancer essential protein in the network. We show that by employing the use of well established drug-target proteins in order to achieve a structural control over essential target proteins within specific cancer PPI networks and apply this to breast, pancreatic, and ovarian cancer signaling transductions PPI networks. We demonstrate that instead of aiming for overall control of entire networks, partial controllability is more effective and efficient in the development of therapies for various cancers.

Importantly, the method applied in our approach over various cancer PPI networks provide a new dimension to apply target controllability that can be unlikely to be successful implemented in other biological networks. Further, finding a set of driver nodes among the entire network in the cancer system can produce results that can be used for experimental validation. Similarly, for other diseases, the emphasis should now fall on formulating appropriate control problems in order to identify the network dynamics in particular on finding disease-specific essential proteins and subsequent modules/subnetworks to successful control. Some of the disease network features

and properties are nonlinear and feedback loops play an important role. It could be an important challenge to add them to the target controllability framework.

Network control theory analyzes the structure of a complex network and can efficiently provide the minimum number of driver nodes through which it can be driven by interconnected nodes in the networks [15]. Recent advancements in the network control approach have shown that full controllability is empirical in the case of medical setups where the control approach can be implemented through a minimum set of FDA based drug-target proteins [39]. Other approaches based on the so-called minimum dominating sets (MDS) have also been applied on control dynamics of PPI networks [100]. This approach is not very feasible for those cases where the network type is represented as a directed graph (such as signaling transduction based disease network or metabolic based disease networks). Another approach based on feedback loops showed that some of these loops play a vital role in the signaling transduction networks by affecting many oscillation and switching the signals [47]. The missing component here is an appropriate mechanism for targeting of cancer essential proteins within the system and its implementation toward the development of combinatorial therapeutic approaches. Therefore, the meaningful strategy here is to find how a minimum set of driver nodes can enhance the designation to control biochemical networks. Hence, these control techniques too should be carefully taken into consideration in the aim of building new and effective control strategies for identifying controllable related input nodes in the biological systems.

For an efficient target control strategy, the key idea is to ignore the nodes which are redundant and focus on the nodes that are necessary to control. For example, in cancer, some of the proteins are identified as essential (for that cancer), if their functioning is essential for the multiplication or survival of those tumor cells. Certain proteins become essential in cancer because of the presence of a mutation in the driver genes. This means that these proteins are essential for pathogenesis (i.e., driver) [61], and the tumour becomes fully dependent on the development of oncogenes through these essential proteins [61, 104]. Cancer essential proteins also control the cell cycle regulators and protein translation machinery [104]. More, it has been also know that the proliferation of cancer states is only associated with a small part of the entire state space. Hence, effective target control of complex biological network is possible, if we are successfully define and able to control that small part of the entire space (i.e., essential proteins in cancer). Moreover, practical constraints and optimal selection of driver nodes are the key strategies for optimal control. Therefore, the development of effective target control approach can enhance the potential to find complex mechanism of human diseases and open the door for novel drug development.

Recently, several control approaches have been applied to the biologi-

cal domain and opened a new way to control complex biological networks [39, 47]. Although, the availability of data is still a hindrance in this process: for example biochemical and signaling pathways databases have very low amount of curated sets of protein interactions [88]. Sometimes, these missing data create constraints for exact experimental validations. Similarly, controllability of nonlinear dynamics can also discover the key componenets of cellular dynamics and help in finding a new set of driver nodes, but development of its control based optimization approach is constrained by the lack of datasets for detailed models. Since these models are hard to extract from biological data, we may focus more on the structure of the networks and by this target controlling of nonlinear biological networks can be more feasible.

The more we learn about biological interactions, the more we can define the different mechanisms of its behavior, and get to map genes, proteins, and other micro molecules. These interaction maps can, in turn, inform the control features in the biological systems. Indeed, controlling these interactions in the biological systems can be a promising avenue to explore through the experimental hypothesis. The benefits of the control approach are that it can be applied to investigate any type of interaction network. An appealing extension in the current controllability framework would be to offer deep insights into the different biochemical reactions in account to time and energy, such as steering a chemical reaction into a metabolic pathway [54, 7, 93]. These reactions might involve different sets of reaction types at different time. These types of problem would be interesting to control and target controllability would be an efficient platform.

Therefore, theoretical and experimental advancement in the current approach will significantly improve the control principle in various diseases and will provide new dataset based approaches and online pipelines. Disease management technologies are continuing to develop at an astonishing rate, and promise of new datasets will complement and supersede the existing interactome model. Functional genomics and proteomics data can provide us a new paradigm for investigating and calculate the control mechanism of biological networks [39], such as how control is achieved and why certain disease behavioral states are favorite over others.

Bibliography

[1] Eivind Almaas. Optimal flux patterns in cellular metabolic networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 17(2):026107, 2007.

[2] Ido Amit, Ron Wides, and Yosef Yarden. Evolvable signaling networks of receptor tyrosine kinases: relevance of robustness to malignancy and to cancer therapy. *Molecular Systems Biology*, 3(1):151, 2007.

[3] Matthew H Bailey, Collin Tokheim, Eduard Porta-Pardo, Sohini Sengupta, Denis Bertrand, Amila Weerasinghe, Antonio Colaprico, Michael C Wendl, Jaegil Kim, Brendan Reardon, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*, 173(2):371–385, 2018.

[4] Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68, jan 2011.

[5] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101, 2004.

[6] Rina Barouch-Bentov and Karsten Sauer. Mechanisms of drug resistance in kinases. *Expert Opinion on Investigational Drugs*, 20(2):153–208, 2011.

[7] Georg Basler, Zoran Nikoloski, Abdelhalim Larhlimi, Albert-László Barabási, and Yang-Yu Liu. Control of fluxes in metabolic networks. *Genome Research*, 26(7):956–968, 2016.

[8] Jorge Burns and Gina Manda. Metabolic pathways of the warburg effect in health and disease: Perspectives of choice, chain or chance. *International Journal of Molecular Sciences*, 18(12):2755, 2017.

[9] Ethan G Cerami, Benjamin E Gross, Emek Demir, Igor Rodchenkov, Özgün Babur, Nadia Anwar, Nikolaus Schultz, Gary D Bader, and Chris Sander. Pathway commons, a web resource for biological pathway data. *Nucleic Acids Research*, 39(suppl_1):D685–D690, 2010.

[10] Guilhem Chalancon, Kai Kruse, and M. Madan Babu. *Metabolic Networks, Structure and Dynamics*, pages 1263–1267. Springer New York, New York, NY, 2013.

[11] Dong-Yeon Cho, Yoo-Ah Kim, and Teresa M Przytycka. Network biology approach to complex diseases. *PLoS Computational Biology*, 8(12):e1002820, 2012.

[12] Reuven Cohen and Shlomo Havlin. *Complex Networks: Structure, Robustness and Function*. Cambridge university press, 2010.

[13] UniProt Consortium. Uniprot: a hub for protein information. *Nucleic Acids Research*, 43(D1):D204–D212, 2015.

[14] Eugen Czeizler, Cristian Gratie, Wu Kai Chiu, Krishna Kanhaiya, and Ion Petre. *Target Controllability of Linear Networks*, pages 67–81. Springer Nature, LNCS, 2016.

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

[16] Ibiayi Dagogo-Jack and Alice T Shaw. Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology*, 15(2):81, 2018.

[17] 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, 2010.

[18] Ralph J DeBerardinis and Navdeep S Chandel. Fundamentals of cancer metabolism. *Science Advances*, 2(5):e1600200, 2016.

[19] Youcef Derbal. Perspective on the dynamics of cancer. *Theoretical Biology and Medical Modelling*, 14(1):18, Oct 2017.

[20] Natalie C Duarte, Scott A Becker, Neema Jamshidi, Ines Thiele, Monica L Mo, Thuy D Vo, Rohith Srivas, and Bernhard Ø Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences*, 104(6):1777–1782, 2007.

[21] Pawel Durek and Dirk Walther. The integrated analysis of metabolic and protein interaction networks reveals novel molecular organizing principles. *BMC Systems Biology*, 2(1):100, 2008.

[22] Fabian V Filipp. Precision medicine driven by cancer systems biology. *Cancer and Metastasis Reviews*, 36(1):91–108, 2017.

[23] Joshua Finkelstein, Noah Gray, Marie Thérèse Heemels, Bar-

bara Marte, and Deepa Nath. Metabolism and disease. *Nature*, 491(7424):347–348, 2012.

[24] R Fisher, L Pusztai, and C Swanton. Cancer heterogeneity: implications for targeted therapeutics. *British Journal of Cancer*, 108(3):479, 2013.

[25] Laura I Furlong. Human diseases through the lens of network biology. *Trends in Genetics*, 29(3):150–159, 2013.

[26] Jianxi Gao, Yang-Yu Liu, Raissa M D'souza, and Albert-László Barabási. Target control of complex networks. *Nature Communications*, 5:5415, 2014.

[27] Pouyan Ghaffari, Adil Mardinoglu, and Jens Nielsen. Cancer metabolism: a modeling perspective. *Frontiers in Physiology*, 6:382, 2015.

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

[29] Mileidy W Gonzalez and Maricel G Kann. Protein interactions and disease. *PLoS Computational Biology*, 8(12):e1002819, 2012.

[30] Attila Gursoy, Ozlem Keskin, and Ruth Nussinov. Topological properties of protein interaction networks from a structural perspective. *Biochemical Society Transaction*, 36(6), 2008.

[31] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674, 2011.

[32] John E Hopcroft and Richard M Karp. An n^5/2 algorithm for maximum matchings in bipartite graphs. *SIAM Journal on Computing*, 2(4):225–231, 1973.

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

[34] Trey Ideker and Ruth Nussinov. Network approaches and applications in biology. *PLoS Computational Biology*, 13(10):e1005771, 2017.

[35] Trey Ideker, Owen Ozier, Benno Schwikowski, and Andrew F Siegel. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, 18(suppl_1):S233–S240, 2002.

[36] Sarika Jalan, Krishna Kanhaiya, Aparna Rai, Obul Reddy Bandapalli, and Alok Yadav. Network topologies decoding cervical cancer.

PloS One, 10(8):e0135183, 2015.

[37] AC Joerger and AR Fersht. Structure–function–rescue: the diverse nature of common p53 cancer mutants. *Oncogene*, 26(15):2226, 2007.

[38] Minoru Kanehisa. Toward pathway engineering: a new database of genetic and molecular pathways. *Sci. Technol. Jap.*, 59:34–38, 1996.

[39] Krishna Kanhaiya, Eugen Czeizler, Cristian Gratie, and Ion Petre. Controlling directed protein interaction networks in cancer. *Scientific Reports*, 7(1):10327, September 2017.

[40] Krishna Kanhaiya, Vladimir Rogojin, Keivan Kazemi, Eugen Czeizler, and Ion Petre. Netcontrol4biomed: a pipeline for biomedical data acquisition and analysis of network controllability. *BMC Bioinformatics*, 19(7):185, 2018.

[41] Gozde Kar, Attila Gursoy, and Ozlem Keskin. Human cancer protein-protein interaction network: a structural perspective. *PLoS Computational Biology*, 5(12):e1000601, 2009.

[42] Carolyn Waugh Kinkade, Mireia Castillo-Martin, Anna Puzio-Kuter, Jun Yan, Thomas H Foster, Hui Gao, Yvonne Sun, Xuesong Ouyang, William L Gerald, Carlos Cordon-Cardo, et al. Targeting akt/mtor and erk mapk signaling inhibits hormone-refractory prostate cancer in a preclinical mouse model. *The Journal of Clinical Investigation*, 118(9):3051–3064, 2008.

[43] Hiroaki Kitano. Biological robustness. *Nature Reviews Genetics*, 5(11):826, 2004.

[44] Alfred G Knudson. Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*, 1(2):157, 2001.

[45] Judice LY Koh, Kevin R Brown, Azin Sayad, Dahlia Kasimer, Troy Ketela, and Jason Moffat. Colt-cancer: functional genetic screening resource for essential genes in human cancer cell lines. *Nucleic Acids Research*, 40(D1):D957–D963, 2012.

[46] Walter Kolch, Melinda Halasz, Marina Granovskaya, and Boris N. K Holodenko. The dynamic control of signal transduction networks in cancer cells. *Nature Reviews Cancer*, 15(9):515–527, aug 2015.

[47] Walter Kolch, Melinda Halasz, Marina Granovskaya, and Boris N Kholodenko. The dynamic control of signal transduction networks in cancer cells. *Nature Reviews Cancer*, 15(9):515, 2015.

[48] Michael Kuhn, Monica Campillos, Ivica Letunic,

Lars Juhl Jensen, and Peer Bork. A side effect resource to capture phenotypic effects of drugs. *Molecular Systems Biology*, 6(1):343, 2010.

[49] Martina Kutmon, Anders Riutta, Nuno Nunes, Kristina Hanspers, Egon L Willighagen, Anwesha Bohler, Jonathan Mélius, Andra Waagmeester, Sravanthi R Sinha, Ryan Miller, et al. Wikipathways: capturing the full diversity of pathway knowledge. *Nucleic Acids Research*, 44(D1):D488–D494, 2016.

[50] Marko Laakso and Sampsa Hautaniemi. Integrative platform to translate gene sets to networks. *Bioinformatics*, 26(14):1802–1803, 2010.

[51] D-S Lee, Juyong Park, KA Kay, Nicholas A Christakis, Zoltan N Oltvai, and A-L Barabási. The implications of human metabolic network topology for disease comorbidity. *Proceedings of the National Academy of Sciences*, 105(29):9880–9885, 2008.

[52] Christoph Lengauer, Kenneth W Kinzler, and Bert Vogelstein. Genetic instabilities in human cancers. *Nature*, 396(6712):643, 1998.

[53] Erel Levine and Terence Hwa. Stochastic fluctuations in metabolic pathways. *Proceedings of the National Academy of Sciences*, 104(22):9224–9229, 2007.

[54] Min Li, Hao Gao, Jianxin Wang, and Fang-Xiang Wu. Control principles for complex biological networks. *Briefings in Bioinformatics*, 09 2018. bby088.

[55] Zenggang Li, Andrei A Ivanov, Rina Su, Valentina Gonzalez-Pecchi, Qi Qi, Songlin Liu, Philip Webber, Elizabeth McMillan, Lauren Rusnak, Cau Pham, et al. The oncoppi network of cancer-focused protein–protein interactions to inform biological insights and therapeutic strategies. *Nature Communications*, 8:14356, 2017.

[56] Zhenping Li, Rui-Sheng Wang, and Xiang-Sun Zhang. Two-stage flux balance analysis of metabolic networks for drug target identification. *BMC Systems Biology*, 5(1):S11, 2011.

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

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

[59] Joseph Loscalzo, Albert-László Barabási, and Edwin K Silverman. *Network Medicine: Com-*

plex Systems in Human Disease and Therapeutics. Harvard University Press, 2017.

[60] Joseph Loscalzo, Isaac Kohane, and Albert-Laszlo Barabasi. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Molecular Systems Biology*, 3(1):124, 2007.

[61] Richard Marcotte, Kevin R Brown, Fernando Suarez, Azin Sayad, Konstantina Karamboulas, Paul M Krzyzanowski, Fabrice Sircoulomb, Mauricio Medrano, Yaroslav Fedyshyn, Judice LY Koh, et al. Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discovery*, 2(2):172–189, 2012.

[62] Iñigo Martincorena and Peter J Campbell. Somatic mutation in cancer and normal cells. *Science*, 349(6255):1483–1489, 2015.

[63] Jörg Menche, Amitabh Sharma, Maksim Kitsak, Susan Dina Ghiassian, Marc Vidal, Joseph Loscalzo, and Albert-László Barabási. Uncovering disease-disease relationships through the incomplete interactome. *Science*, 347(6224):1257601, 2015.

[64] Daniele Merico, Ruth Isserlin, Oliver Stueker, Andrew Emili, and Gary D Bader. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PloS One*, 5(11):e13984, 2010.

[65] Franziska Michor, Yoh Iwasa, and Martin A Nowak. Dynamics of cancer progression. *Nature Reviews Cancer*, 4(3):197, 2004.

[66] Koyel Mitra, Anne-Ruxandra Carvunis, Sanath Kumar Ramesh, and Trey Ideker. Integrative approaches for finding modular structure in biological networks. *Nature Reviews Genetics*, 14(10):719, 2013.

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

[68] Jose C Nacher and Tatsuya Akutsu. Analysis of critical and redundant nodes in controlling directed and undirected complex networks using dominating sets. *Journal of Complex Networks*, 2(4):394–412, 2014.

[69] Jose C Nacher and Tatsuya Akutsu. Structurally robust control of complex networks. *Physical Review E*, 91(1):012826, 2015.

[70] Cancer Genome Atlas Research Network et al. Integrated genomic analyses of ovarian carcinoma. *Nature*, 474(7353):609, 2011.

[71] Rod K Nibbe, Salim A Chowdhury, Mehmet Koyutürk, Rob Ewing, and Mark R Chance. Protein–protein interaction networks and subnetworks in the biology of disease. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 3(3):357–367, 2011.

[72] Kristian Ovaska, Marko Laakso, Saija Haapa-Paananen, Riku Louhimo, Ping Chen, Viljami Aittomäki, Erkka Valo, Javier Núñez-Fontarnau, Ville Rantanen, Sirkku Karinen, et al. Large-scale data integration framework provides a comprehensive view on glioblastoma multiforme. *Genome Medicine*, 2(9):65, 2010.

[73] Venkat R Pannala, Martha L Wall, Shanea K Estes, Irina Trenary, Tracy P O'Brien, Richard L Printz, Kalyan C Vinnakota, Jaques Reifman, Masakazu Shiota, Jamey D Young, et al. Metabolic network-based predictions of toxicant-induced metabolite changes in the laboratory rat. *Scientific Reports*, 8(1):11678, 2018.

[74] Kiran Raosaheb Patil and Jens Nielsen. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of Sciences*, 102(8):2685–2689, 2005.

[75] T Pawson and N Warner. Oncogenic re-wiring of cellular signaling pathways. *Oncogene*, 26(9):1268, 2007.

[76] Tony Pawson and Piers Nash. Protein–protein interactions define specificity in signal transduction. *Genes & Development*, 14(9):1027–1047, 2000.

[77] Livia Perfetto, Leonardo Briganti, Alberto Calderone, Andrea Cerquone Perpetuini, Marta Iannuccelli, Francesca Langone, Luana Licata, Milica Marinkovic, Anna Mattioni, Theodora Pavlidou, et al. Signor: a database of causal relationships between biological entities. *Nucleic Acids Research*, 44(D1):D548–D554, 2016.

[78] Svatopluk Poljak. On the generic dimension of controllable subspaces. *IEEE Transactions on Automatic Control*, 35(3):367–369, 1990.

[79] Natapol Pornputtapong, Intawat Nookaew, and Jens Nielsen. Human metabolic atlas: an online resource for human metabolism. *Database*, 2015, 2015.

[80] Gianmarco Rinaldi, Matteo Rossi, and Sarah-Maria Fendt.

Metabolic interactions in cancer: cellular metabolism at the interface between the microenvironment, the cancer cell phenotype and the epigenetic landscape. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 10(1):e1397, 2018.

[81] Eric E Schadt, John Lamb, Xia Yang, Jun Zhu, Steve Edwards, Debraj GuhaThakurta, Solveig K Sieberts, Stephanie Monks, Marc Reitman, Chunsheng Zhang, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genetics*, 37(7):710, 2005.

[82] Richard Sever and Joan S Brugge. Signal transduction in cancer. *Cold Spring Harbor Perspectives in Medicine*, 5(4):a006098, 2015.

[83] Jared A Sewell and Juan I Fuxman Bass. Cellular network perturbations by disease-associated variants. *Current Opinion in Systems Biology*, 3:60–66, 2017.

[84] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003.

[85] Erin R Shellman, Charles F Burant, and Santiago Schnell. Network motifs provide signatures that characterize metabolism. *Molecular BioSystems*, 9(3):352–360, 2013.

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

[87] Annapoorna Sreedhar and Yunfeng Zhao. Dysregulated metabolic enzymes and metabolic reprogramming in cancer cells. *Biomedical Reports*, 8(1):3–10, 2018.

[88] R Greg Stacey, Michael A Skinnider, Jenny HL Chik, and Leonard J Foster. Context-specific interactions in literature-curated protein interaction databases. *BMC Genomics*, 19(1):758, 2018.

[89] Arion I Stettner and Daniel Segrè. The cost of efficiency in energy metabolism. *Proceedings of the National Academy of Sciences*, 110(24):9629–9630, 2013.

[90] Zachary E Stine, Zandra E Walton, Brian J Altman, Annie L Hsieh, and Chi V Dang. Myc, metabolism, and cancer. *Cancer Discovery*, 5(10):1024–1039, 2015.

[91] Jingchun Sun and Zhongming Zhao. A comparative study of cancer proteins in the human protein-protein interaction network. *BMC Genomics*, 11(3):S5, 2010.

[92] Collin J Tokheim, Nickolas Papadopoulos, Kenneth W Kinzler, Bert Vogelstein, and Rachel Karchin. Evaluating the evaluation of cancer driver genes. *Proceedings of the National Academy of Sciences*, 113(50):14330–14335, 2016.

[93] Emma K Towlson, Petra E Vértes, Gang Yan, Yee Lian Chew, Denise S Walker, William R Schafer, and Albert-László Barabási. Caenorhabditis elegans and the network control framework—faqs. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1758):20170372, 2018.

[94] Bert Vogelstein and Kenneth W Kinzler. Cancer genes and the pathways they control. *Nature Medicine*, 10(8):789, 2004.

[95] Thuy Vu, Mark X Sliwkowski, and Francois X Claret. Personalized drug combinations to overcome trastuzumab resistance in her2-positive breast cancer. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1846(2):353–365, 2014.

[96] Maonan Wang, Jingzhou Zhao, Lishen Zhang, Fang Wei, Yu Lian, Yingfeng Wu, Zhaojian Gong, Shanshan Zhang, Jianda Zhou, Ke Cao, et al. Role of tumor microenvironment in tumorigenesis. *Journal of Cancer*, 8(5):761, 2017.

[97] Rui-Sheng Wang and Joseph Loscalzo. Illuminating drug action by network integration of disease genes: a case study of myocardial infarction. *Molecular BioSystems*, 12(5):1653–1666, 2016.

[98] David S Wishart, Craig Knox, An Chi Guo, Dean Cheng, Savita Shrivastava, Dan Tzur, Bijaya Gautam, and Murtaza Hassanali. Drugbank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research*, 36(suppl_1):D901–D906, 2008.

[99] Laura D Wood, D Williams Parsons, Siân Jones, Jimmy Lin, Tobias Sjöblom, Rebecca J Leary, Dong Shen, Simina M Boca, Thomas Barber, Janine Ptak, et al. The genomic landscapes of human breast and colorectal cancers. *Science*, 318(5853):1108–1113, 2007.

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

[101] Gang Yan, Petra E Vértes, Emma K Towlson, Yee Lian Chew, Denise S Walker, William R Schafer, and Albert-László Barabási. Network control principles predict neuron function in the caenorhabditis elegans connectome. *Nature*, 550(7677):519, 2017.

[102] Esti Yeger-Lotem and Roded Sharan. Human protein interaction networks across tissues and diseases. *Frontiers in Genetics*, 6:257, 2015.

[103] Akihiko Yoshimura. Signal transduction of inflammatory cytokines and tumor development. *Cancer Science*, 97(6):439–447, 2006.

[104] Tianzuo Zhan and Michael Boutros. Towards a compendium of essential genes–from model organisms to synthetic lethality in cancer cells. *Critical Reviews in Biochemistry and Molecular Biology*, 51(2):74–85, 2016.

[105] Bai Zhang, Ye Tian, and Zhen Zhang. Network biology in medicine and beyond. *Circulation: Genomic and Precision Medicine*, 7(4):536–547, 2014.

[106] Elena Zotenko, Julian Mestre, Dianne P O'Leary, and Teresa M Przytycka. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Computational Biology*, 4(8):e1000140, 2008.

Krishna Kanhaiya

Target Controllability of Cancer Networks

In this thesis, we integrate network biology and network controllability approach, to gain useful insight into the finding of the complex mechanism of cancer networks. We show that how through a minimum number of target nodes a full (partial) controllability of these intracellular networks can be achieved. According to control theory, a dynamical system may be steered such that its output is driven towards some desired final states (e.g. target cancer essential proteins in PPI networks) via suitably-picked inputs (e.g. manipulating a set of driver proteins). Therefore, it is necessary to understand the dynamics of these complex networks, and their evolution rules (i.e., expressed as a system of linear equations) which govern the systems dynamics over time. The control approach presented here can be an impressive framework for effective development of multi drug-target therapeutics. We, therefore, expect that our approach can open a new way towards effective and efficient therapeutics target and a key resource towards personalized medicine in cancer.

ISBN 978-952-12-3948-9