Betweenness Analysis on Structural Protein-Protein Interaction Networks

by

Cemal Yamak

A Thesis Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science
in
Computer Science and Engineering

Koc University

January 2016

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Cemal Yamak

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Prof. Attila Gürsoy (Advisor)

_____

Prof. Özlem Keskin Özkaya

_____

Assoc. Prof. Öznur Özkasap

_____

Assoc. Prof. Engin Erzin

_____

Asst. Prof. Ömer Sinan Saraç

Date: _____

To my family …

ABSTRACT

Protein-protein interaction networks (PPIN) that are maps of physical interactions between proteins are important and essential abstractions to understand biological functions and processes. Nodes represent proteins and edges represent pairwise interactions between proteins in a PPIN.

Analysis of the proteins with high betweenness score is important in terms of their role in biological processes, molecular functions and cancer. PPINs do not indicate structural and chemical features of interactions. Our goal is to construct a structural PPI network (SPPIN) and analyze the proteins with high betweenness score found by shortest path, flow and random walk betweenness algorithms. To construct the SPPIN, we assigned weights on the edges of the network to reflect the effect of the simultaneous and mutually exclusive interactions and adjusted these weights for different betweenness algorithms separately. We also generated random samples of the network by selecting a set of interactions that can co-occur for each sample and calculated betweenness scores. We observed that %37.5 of bottleneck proteins of SPPIN are different from bottleneck proteins of PPIN found by shortest path betweenness centrality algorithm. This percentage is %18.75 for flow betweenness and % 31.25 for random walk betweenness algorithm. Bottleneck proteins of SPPIN which are different than bottleneck nodes of PPIN have roles in biological processes such as DNA repair and stimulus and have molecular functions such as hormone receptor binding and transcription activation. In addition, we observed that betweenness is positively correlated with number of mutations on interfaces.

ÖZET

Proteinler arasındaki fiziksel etileşimlerin haritaları olan protein-protein etkileşim ağları, biyolojik işleyişleri anlamak için önemli ve gerekli soyut yapılardır. Bir protein-protein etkileşim ağında, düğümler proteinleri, kenarlar ise ikili etkileşimleri temsil eder.

Yüksek aradalık skoru olan proteinleri analiz etmek, proteinlern biyolojik işleyişlerdeki ve moleküler fonksiyonlardaki rolü açısından önemlidir. Protein–protein etkileşim ağları, etkileşimlerin yapısal ve kimyasal özellikleirni içermez. Biz, yapısal bir protein-protein etkileşim ağı modeli oluşturmayı ve bu ağ modeli üzerinde, kısa yol, akış ve rastgele yol aradalık algoritmalarıyla hesaplanan, yüksek aradalık scoruna sahip proteinleri incelemeyi amaçladık. Yapısal bir protein-protein etkileşim ağı oluşturmak amacıyla, ağın kenarları üzerinde, eşzamanlı ve dışlayan etkeleşimleri temsil edebilmek için ağırlık değerleri belirledik ve bu ağırlıkları herbir aradalık algoritması için düzenledik. Ayrıca, eşzamanlı gerçekleşebilen etkileşimler kümeleri oluşturarak, ağ örnekleri oluşturduk ve bu örnekler üzerinde proteinlerin aradalık değerlerini hesapladık. Kısa yol aradalık algoritmasıyla yapısal protein-protein etkileşim ağında tespit edilen darboğaz özelliği gösteren proteinlerin yüzde 37,52'inin yapısal olmayan etkileşim ağında tespit edilen proteinlerden farklı olduğunu gözlemledik. Bu fark oranı, akış aradalık algoritması için yüzde 18,75; rastgele yol aradalık algoritması için yüzde 31,25 olarak hesaplanmıştır. Yalnızca yapısal protein-protein etkileşim ağında darboğaz özelliği gösteren proteinlerin, DNA tamiri ve uyarılması gibi biyolojik işleyişlerde ve hormone alıcı bağlanması ve transkripsiyon aktivasyonu gibi moleküler fonksiyonlarda rolleri olduğu belirlenmiştir. Ayrıca, aradalık özelliğinin arayüzler üzerindeki mutasyonlarla pozitif korelasyonu olduğunu saptadık.

ACKNOWLEDGEMENTS

I am grateful to my advisors, Prof. Attila Gursoy and Prof. Özlem Keskin Özkaya for their guidance and contribution to my studies. Being a member of COSBI Lab was a pleasure for me. I also would like to thank Asst. Prof. Ömer Sinan Saraç, Assoc. Prof. Öznur Özkasap, Assoc. Prof. Engin Erzin for their valuable time, critical time and comments.

I acknowledge the financial support of the Scientific and Technological Research Council of Turkey (TÜBİTAK) during my M.Sc. study. Additionally, I am grateful for the facilities that Koç University provided.

Special thanks to Deniz Demircioğlu, Nilay Karahan, Serana Muratcıoğlu, Emine Güven, Bilgesu Erdoğan for their contribution to my studies.

I am also greateful to my other officemates Engin Çukuroğlu, Güray Kuzu, Alper Başpınar, Emel Şen, Ayşe Derya Cavga, Farideh Halakou, Sıla Özdemir, Yusuf Doğuş Doğru, Efe Elbeyli, Emre Büyüktuna and Mahan Rajaei Golsefidi for all the time we spent together at COSBI Lab.

Finally, I want to thank my family for their support, encouragement and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

PPIN	protein-protein interaction network
SPPIN	structural protein-protein interaction network
PRISM	protein interactions by structural matching

Chapter 1
INTRODUCTION

Proteins are biological macromolecules which have important functions in living organisms such as DNA replication, catalyzing metabolic reactions etc. Proteins rarely act alone. They interact with other proteins or other molecules such as peptides, DNA, RNA, lipid etc. and form complex structures to function in biological processes.

Protein-protein interaction networks (PPIN) that are maps of physical interactions between proteins are important and essential abstractions to understand biological functions and processes. Nodes represent proteins and edges represent pairwise interactions between proteins in a PPIN. Betweeness centrality analysis on PPIN is necessary to determine important proteins. However, PPINs do not indicate structural and chemical features of interactions.

Analysis of topological features of PPINs such as shortest path calculations, centrality, clustering coefficients helps us to understand dynamics of them. Topological analysis of PPINs showed that PPINs are Scale-Free networks which are resistant to random node failures but weak to targeted removal of hub nodes [1, 2]. Several studies focus on high betweenness scored nodes on PPINs. These studies revealed that high betweenness proteins are more likely to be essential and play important role in the coordinated functionality of yeast interactome [3]. However, structural features of interactions are not considered in all these studies. Some of the interactions in PPIN can be mutually exclusive.

In 2006, Kim et al. integrated structural information with PPINs [4]. Tuncbag et al. used PRISM prediction algorithm to predict structure of interactions which are extracted from experimental interaction data from the Molecular Interaction Map (MIM) of Kohn and his colleagues and analyzed mutually exclusive and simultanesouly occurable interactions [5, 6]. Demircioglu et al. modelled p53 SPPIN by assigning weights on

edges to represent mutually exclusive interaction feature and analyzed betweenness centrality results of edges [7].

In this study, to construct SPPIN , we assigned weights on the edges of the network to reflect the effect of the simultaneous and mutually exclusive interactions as it is done in the study of Demircoglu et. al and introduced random sampling method. On SPPIN, we analyzed shortest path, flow and random walk betweenness algorithms and compared the results. After defining bottleneck proteins using betweenness scores, biological analysis is done in terms of biological processes, molecular functions, gene essentiality analysis and correlation of betweenness with mutations on interface structures.

The outline of this study is as follows:

In Chapter 2, protein-protein interaction networks (PPINs) and structural protein-protein interaction networks (SPPINs) are described. First, centrality analysis and its importance on networks are given and flow and random walk betweenness centrality measures are explained. Then, protein-protein interaction networks and structural protein-protein interaction networks are presented and studies related to analysis of them are given.

In Chapter 3, the materials and methods are given. It covers the data set and the methodology to analyze betweenness centrality on structural protein-protein interaction networks (SPPINs). After the network data and its modelling are provided, definitions related to biological and topological analysis are given.

Chapter 4 includes results and analysis of betweenness centrality algorithms. First correlation of flow and random-walk betweenness with other measures is discussed. In addition, correlation of betweenness with mutations on interfaces is analyzed. Next, comparison of bottlenecks according to network and algorithm type is given. Finally, biological analysis of high betweenness scored proteins is discussed.

In the last chapter, this thesis is ended with discussions and conclusions of the study.

Chapter 2
LITERATURE REVIEW

In this chapter, centrality on networks, protein-protein interaction networks (PPIN) and structural protein-protein interaction networks (SPPIN) are reviewed. First, centrality analysis and its importance on networks are given and flow and random walk betweenness centralities are explained. Then, protein-protein interaction networks and structural protein-protein interaction networks are presented and studies related to analysis of them are given.

2.1 Centrality Measures in Networks

Centrality measures are needed to identify important vertices in networks. Finding key vertices in a computer network, important and influential actors in a social network, bottleneck vertices in a transportation network and key genes, proteins in a biological network are the common applications of centrality measures. Researchers benefited from centrality measures in various types of networks and in a wide range of studies such as analysis of control, risk, influence, exposure, belongingness, independence, power, competence, cooperation.

Centrality measures are varied in terms of what decides the centrality of vertices. When central or key nodes or edges are considered based on how traffic flows in networks, centrality depends on types of flow and transfer paths [8]. Alternatively, when centrality is considered as involvement in cohesiveness of network, it is calculated based on cohesiveness metrics [9].

Centrality measures can be reviewed as the evaluation of participation of the nodes or edges in the walks of a network. That is, the volume or length of walks of some type that originate, terminate or pass through a node is calculated then summed and averaged in the scope of centrality measures. In that concept, centrality measures can be distinguished by four dimensions: types of walks (such as geodesic, random), properties

of walks (volume or length), type of nodal involvement (radial or medial), type of summarization (sum or average) [9].

If we classify centrality measures in terms of type of nodal involvement, centralities are either radial or medial. Radial centralities calculate the number of walks originated from or terminated in the given vertex. Degree and Eigenvalue centralities are in the type of radial centrality. Walk length in the degree centrality is one and in the eigenvalue is infinity. Medial centralities calculate the number of walks passing through the given vertex. Freeman's betweenness centrality is an example of medial centrality [9, 10].

Betweenness centrality was introduced By Linton Freeman as a measure of expressing the control of an individual on the communication between other individuals. It is important and beneficial to measure betweenness in many network studies. And it has beneficial interpretations. For instance, in a network representing transportation, nodes with higher betweenness centrality reveal the bottleneck parts or in a network representing the trade routes between medieval cities, cities with high betweenness can be interpreted as the cities which are central in terms of wealth [11].

Betweenness centrality of a given node is defined as number of geodesic paths from all nodes to all other nodes that pass through that given node [12]. Betweenness centrality of a node can be formulated as follows:

$$b_i = \frac{\sum_{s<t} g_i^{(st)}/n_{st}}{\frac{1}{2}n(n-1)}$$

where n_{st} is the total number of shortest paths from s to t, $g_i^{(st)}$ is the total number of shortest paths passing through i and n is the number of vertices in the network.

Basic calculation of betweenness centrality requires $O(n^3)$ and $O(n^2)$ space where n is the number of nodes in network. Brandes proposed an algorithm whose running time is $O(nm)$ for unweighted networks and $O(nm + n^2 \log n)$ for weighted networks [13, 14].

2.2 Flow Betweenness Centrality

In generic betweenness centrality algorithms, only shortest paths are considered. However, information, rumors or messages do not flow only along the shortest paths, they travels more randomly till they reach the target node [15]. For example in a social network, a message or information in general spreads along the network without a knowledge of geodesics. Therefore, we can say that non-geodesic paths also have importance on betweenness measure of nodes in networks.

Moreover considering only shortest paths can lead some problems. Below figure from Newman's study reveals one of these problems.

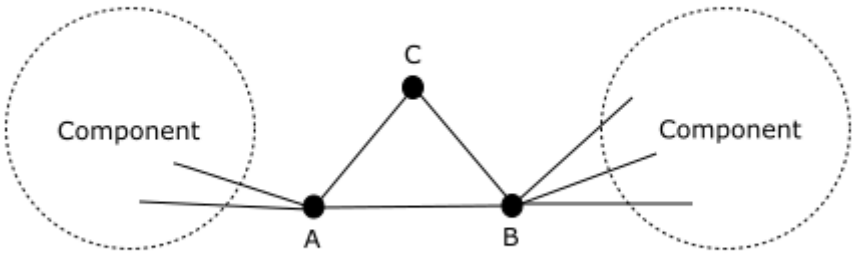

Figure 2-1Problem of shortest path betweenness centrality

In the figure above, the vertices A and B get higher betweenness score while vertex C gets a very low score since generic algorithm for betweenness prefers the path which A and B is on and the path C is on is missed [15].

Freeman suggested a new betweenness measure called flow betweenness that includes some of the non-geodesic paths. Flow betweenness is based on maximum flows. Flow betweenness of vertex i is defined as the amount of flow passing through vertex i when the maximum flow is transmitted from source vertex to target vertex, averaged over all possible sources and targets [16].

Flow Betweenness can be formulated as in the following.

$$C_F(i) = \frac{\sum_j^n \sum_k^n m_{jk}(x_i)}{\sum_j^n \sum_k^n m_{jk}}, \quad (j < k)$$

where m_{jk} is the amount of flow between vertex j and vertex k ; and $m_{jk}(x_i)$ is the amount of flow passing through vertex i when maximum flow transmitted from j to k.

Applying maximum flow algorithms directly is not a good idea to calculate flow betweenness since maximum flow path between two nodes is not unique. Instead of this idea, the following approach is being used: calculating maximum flows between all possible pairs of nodes and removing the node whose centrality is being measured and then recalculating the maximum flows. Definition is in the below.

$$c_k = \sum_{i,j} \frac{w_{i,j}^k - w_{i,j}^{k*}}{w_{i,j}^k}$$

where w is the matrix of maximum flows between pairs of nodes; w^k is the matrix deleting the column and row k from w ; and $w^{k*}$is the matrix constructed by deleting node k from network and recalculating the maximum flow matrix [9]. Running time for calculating flow betweenness on a network with n vertices and m edges is $O(n^3m^2)$.

2.3 Random-Walk Betweenness Centrality

Flow betweenness solves the problem emphasized in Freeman's study however in a network as below it faces with the similar problem with shortest-path betweenness measure.

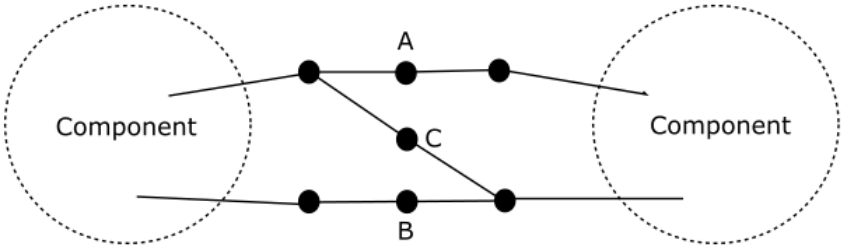

Figure 2-2Problem of flow betweenness centrality

In the figure from Freeman's study above, the vertices A and B get higher betweenness score while vertex C gets a very low score since maximum flow is transmitted on the routes that A and B is on and the path C is on is missed [15]..

Shortest-path and flow betweenness measures estimate that information flows through ideal routes. A realistic betweenness measure should weigh all possible paths. Therefore a new betweenness measure called random walk betweenness is proposed. Random walk betweenness of a vertex i is defined as the number of times that a random walk from s to t passes through i, averaged over all s and t [15].

$$b_i = \frac{\sum_{s<t} I_i^{(st)}}{\frac{1}{2}n(n-1)}$$

where $I_i^{(st)}$ is the net flow through vertex i. $I_i^{(st)}$ is calculated as follows:

$I_i^{(st)} = \frac{1}{2}\sum_j A_{ij}\left|T_{is} - T_{it} - T_{js} + T_{jt}\right|$, for $i \neq s,t$ where $T = (D_t - A_t)^{-1}$. Running time of random walk betweenness calculation on a network with n vertices is $O(n^3)$.

2.4 Protein-Protein Interaction Networks

Proteins are biological macromolecules which have important functions in living organisms such as DNA replication, catalyzing metabolic reactions etc. They are consisting of one or more amino acid residue chain. Sequence of amino acids coded by genes differentiates proteins in terms of folding into a 3D structure and structure of a protein determines its function. Structures of proteins are often analyzed in four hierarchical categories: primary structure that is amino acid sequence, secondary structure that is helices and strands, tertiary structure that is arrangements of motifs or domains in 3D space and quaternary structure that is protein complexes formed by arrangement of monomeric proteins.

Proteins interact with other proteins or other molecules such as peptides, DNA, RNA, lipid etc. and form complex structures to function in biological processes. Proteins interact through their surface by means of shape and biochemical complementarity [17, 18]. Region that proteins interact through by non-covalent bonds is called interface. Interfaces are the interacting residues involved in two different chains.

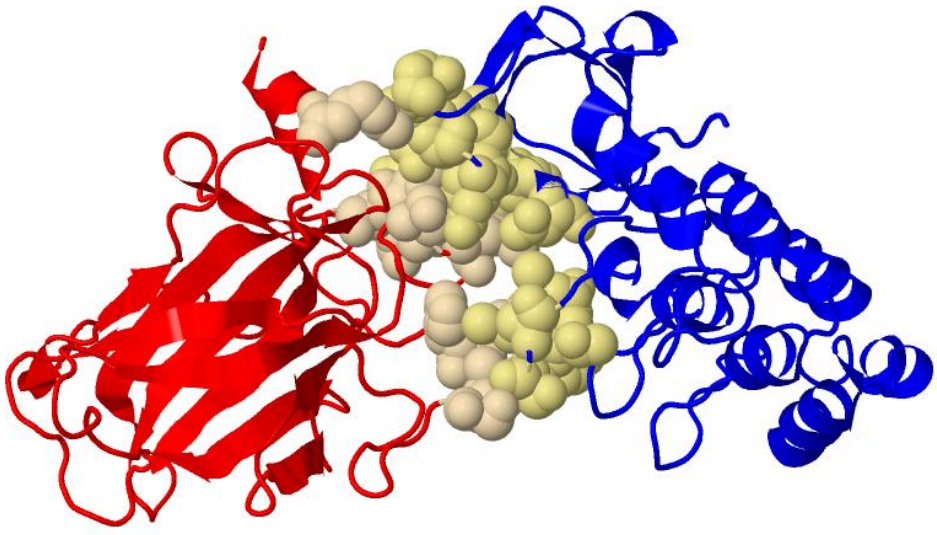

Figure 2-3 Structure and Interface region of p53 Tumor Suppressor.

In the Figure 2.3, interface region of p53 suppressor generated using HotRegion Web Server is seen as an example [19]. The red part on the left corresponds to ‘A’ chain and the blue part on the right corresponds to ‘B’ chain of the structure. Ball shaped atoms in the middle represent the atoms of interface that is. Light yellow atoms are belong to chain A and dark yellow ones are belong to chain B.

Properties of protein-protein interfaces such as flexibility, hydrophobicity, solvent accessibility, evolutionary conservation are analyzed using crystal structure of protein-protein complex. Interface residues can be determined by the means of the distance or accessible surface area [20-24].

Biological processes are generally driven as a result of proteins interact with each other rather than act alone. The amount of interaction data is getting bigger as a result of related studies such as the mass spectrometry [25, 26], yeast two-hybrid assay [27, 28], data mining techniques [29] etc. Thus, analysis of protein interactions is necessary to determine important proteins and understand functional organization of cell. Protein-protein interaction networks are constructed for that purpose. In a protein-protein interaction network, nodes represent proteins and edges represent physical interactions between proteins. Formally PPIN is modelled as a graph G= (V, E) where V is set of proteins and E is set of edges which identify pairwise interaction between two proteins.

Analysis of topological structure of biological networks such as shortest path calculations, centrality, clustering coefficients gives researchers chance to understand dynamics of these networks. Topological analysis of biological networks showed that the nodes with high degree called hub nodes are associated with genes that cause disease and bottleneck nodes may be correlated with essentiality [25, 28, 30, 31].

Topological analysis of PPINs showed that PPINs are Scale-Free networks which are resistant to random node failures but weak to targeted removal of hub nodes [1, 2]. Hence, specifying hub nodes with degree centrality measure in a PPIN gives information about on which nodes network is vulnerable to attacks. Betweenness centrality is another important measure for analysis of PPINs. Several studies focus on high betweenness scored nodes on PPINs. These studies revealed that high betweenness

proteins are more likely to be essential and play important role in the coordinated functionality of yeast interactome. Betweenness centrality was also used to analyze topology of yeast protein interactome for coordinated functionality [3, 31, 32].

2.5 Structural Protein-Protein Interaction Networks

Protein-protein interaction networks that are maps of physical interactions between proteins are important and essential abstractions to understand biological functions and processes. Nodes represent proteins and edges represent pairwise interactions between proteins in a PPIN. However, PPINs do not indicate structural and chemical features of interactions.

In 2006, Kim et al. integrated structural information with PPINs [4]. They used yeast protein interaction network generated by various sources and refined it with statistical methods. They further map edges in this network structurally to known complexes and distinguished the interfaces of each interaction. If two or more proteins interacting with a common protein use the same interface on the common protein, the interactions were defined as mutually exclusive. In the case they use different interfaces, the interaction were categorized as simultaneously possible interactions. As a result they generated structural yeast interaction network. They also categorized hubs with one or two interfaces as singlish-interface hubs and hubs with more interfaces as multi-interface hubs. They concluded that multi-hubs are more likely to be essential, more conserved and members of large and stable complexes and they are more likely to be coexpressed with their neighbors than singlish-interface hubs.

In 2012, a tool called SAPIN was developed by Yang et al. to identify simultaneously possible and mutually exclusive interactions [33].

In 2009, Kar et al. analyzed the interfaces and topological properties of structural interaction network of cancer-related proteins and concluded that cancer-related proteins have smaller, more planar, more charged and less hydrophobic binding sites than non-cancer proteins [34].

Tuncbag et al. extracted experimental interaction data from the Molecular Interaction Map (MIM) of Kohn and his colleagues and used PRISM prediction algorithm to predict structure of interactions [5, 6]. PRISM (Protein Interactions by Structural Matching) is an algorithm predicting protein-protein interactions. Interfaces are structurally more conserved than the other regions through the evolution. PRISM predicts interactions and binding sites by using structural and evolutionary similarity to known protein binding sites. Below figure shows the simple mechanism of Prism [5].

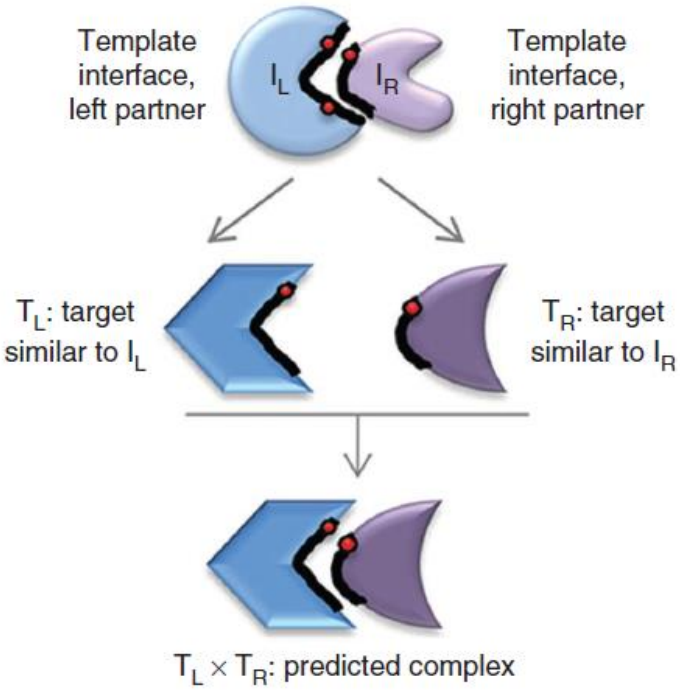

Figure 2-4 Schematic representation of principle of the PRISM algorithm.

Using structural information generated by PRISM, Tuncbag et al. showed how integrating structural information with PPINs can help to understand time-dimensionality and to predict mutually exclusive and simultaneously occurrence properties of interactions [5].

Demircioglu et al. gathered and refined the p53 network data generated by Tuncbag et al with improved version of PRISM. They modelled a structural p53 pathway PPIN by assigning weights on edges to represent mutually exclusive interaction feature and analyzed betweenness centrality results of edges [7]. They classified hubs and bottlenecks on the network they generated and compared them with the ones coming from unweighted version of same PPIN. According to their results, the bottleneck edges of weighted PPIN differ from the ones of unweighted version by %68. They also concluded that the proteins that are at the ends of the bottleneck edges have regulatory roles, play roles in cell cycle and molecular functions such as protein binding, transcription factor binding and kinase activity.

Chapter 3
MATERIALS AND METHODS

This chapter covers the data set and the methodology to analyze betweenness centrality on structural protein-protein interaction networks (SPPINs). First, the network data and its modelling are provided. Then, definitions related to biological and topological analysis are explained.

3.1 p53 Protein-Protein Interaction Network

Demircioglu et al. used the p53 network data collected from literature by Tuncbag et al. In this study, , the same network in their work is used consisting of 81 nodes and 240 edges [7]. The visualization of it using Cytoscape is in the Figure 3.1 [35].

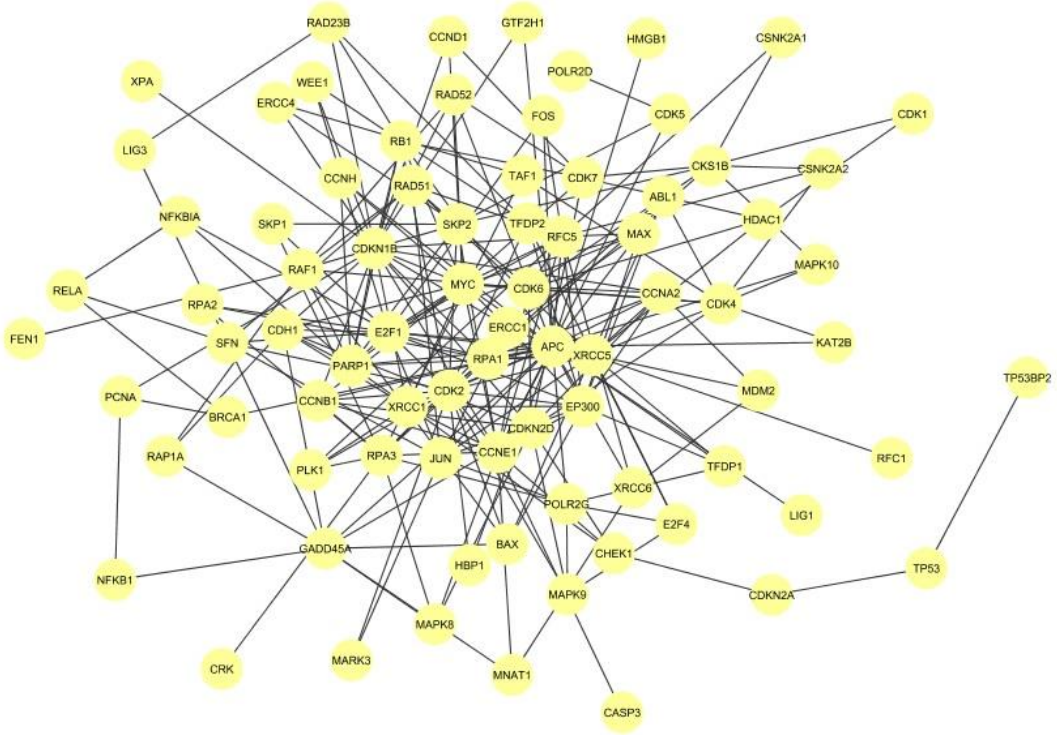

Figure 3-1 p53 network with 81 genes and 240 interactions

3.2 Structural p53 Protein-Protein Interaction Network

Demircioglu et al. run PRISM algorithm on p53 PPIN to predict interfaces of interactions [7]. The structure of 131 out of 240 interactions is predicted by PRISM. Therefore, the binding residues on the surface are determined for 131 interactions. With those predictions, structural p53 is formed. In this study, we used their PRISM results to model structural network for betweenness computation.

3.3 Definition of Overlapping Binding Sites

Proteins interact through their interfaces. In other words, interfaces are the specific binding residues on which a protein establishes interactions with other proteins. With the help of the structure prediction of interactions by PRISM, some residues on the interfaces are defined as overlapping in case that two or more interaction partners use them in the interactions [7].

In the Figure 3.2, there is an example case for overlapping binding site which is similar to the case given by Demircioglu et al. APC protein interacts with EP300 and MYC protein [7]. For APC and EP300 interaction, PRISM predicted 3g33BD interface (complementary binding sites of two interacting proteins) for APC and EP300 interaction and 1pcEF interface for APC and MYC interaction.

The common residues in corresponding residues of 3g33BD interface and 1pcEF interface on APC are: ARG.232, GLN.236, LEU.145, LEU.229, ALA.239, GLU.152, SER.142, ARG.141, ILE.228, ASP.149, ASP.156, GLU.225. These residues are used in binding sites of both APC-MYC and APC-EP300 interactions. Therefore, APC-MYC and APC-EP300 interactions are mutually exclusive.

Figure 3.2 is drawn in UCSF Chimera for visual explanation of the case of APC-MYC and APC-EP300 interactions [36]. In Figure 3.2.a, interactions are seen in the classical network representation and binding sites and overlapping regions are shown in b and c in details.

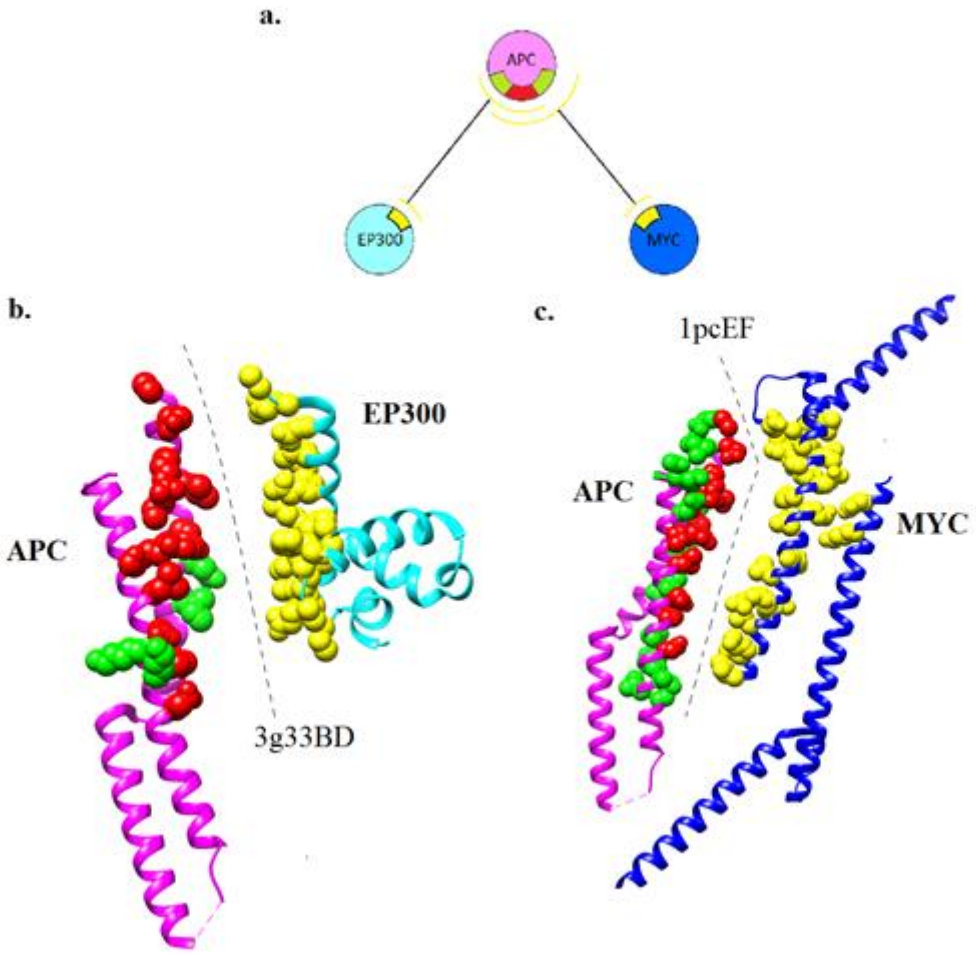

Figure 3-2 Representation of APC-MYC and APC-EP300 interactions with interfaces predicted by PRISM on them.

a) Simple network representation of APC in magenta, MYC in cyan and EP300 in blue and interactions between them. b) 3D representation of the interaction between APC and EP300. Corresponding residues of 3g33BD interface on APC shown with red and green spheres and on EP300 with yellow spheres. Red spheres represent overlapping residues with residues of interface 1pcEF on APC. c) 3D representation of the interaction between APC and MYC. Corresponding residues of 1pcEF interface on APC shown with red and green spheres and on MYC with yellow spheres. Red spheres represent overlapping residues with residues of interface 3g33BD on APC.

3.3 Modeling Structural Protein-Protein Interaction Network for Betweenness computation

We present methods for modeling Structural PPI Network (SPPIN) for shortest path, flow and random walk betweenness centrality algorithms. We used weight assignment method proposed in previous study and adapted it for flow and random walk betweenness centrality algorithms. In addition, we proposed a modelling method for SPPIN by Monte Carlo simulation.

3.3.1 Weighted Graph Representation of p53 Structural Protein-Protein Interaction

With the help of the overlapping binding site information which is generated by using PRISM results, some interactions are categorized as mutually exclusive or cannot be established at the same. This information is reflected on PPIN by using the same approach proposed in the study of Demircioglu and his colleagues. They assigned weights on edges to reflect mutually exclusive interface feature [7].

All edge weights are set to 1 initially. Then, for every interaction i, corresponding edge weight is increased by the number mutually exclusive interactions with i. It is given by

$$w_i = 1 + N_i \ ,$$

where w_i is the weight of the corresponding edge of interaction i , and N_i is the number of interactions that cannot co-occur with interaction i.

For example, if interfaces of P1-P2 and P1-P3 interactions have overlapping binding sites on P1 protein, weights of edge that connects P1 and P2 and edge that connects P1 and P3 are increased by 1 since P1-P2 and P1-P3 interactions are mutually exclusive. Similarly, if interface of P1 and P2 interaction are overlapping with the interface of P2 and P4 interaction on P2 protein, corresponding edge weights of P1-P2 and P2-P4 interactions are increased by 1. In the figure 3.3, more extended version of this sample case is shown.

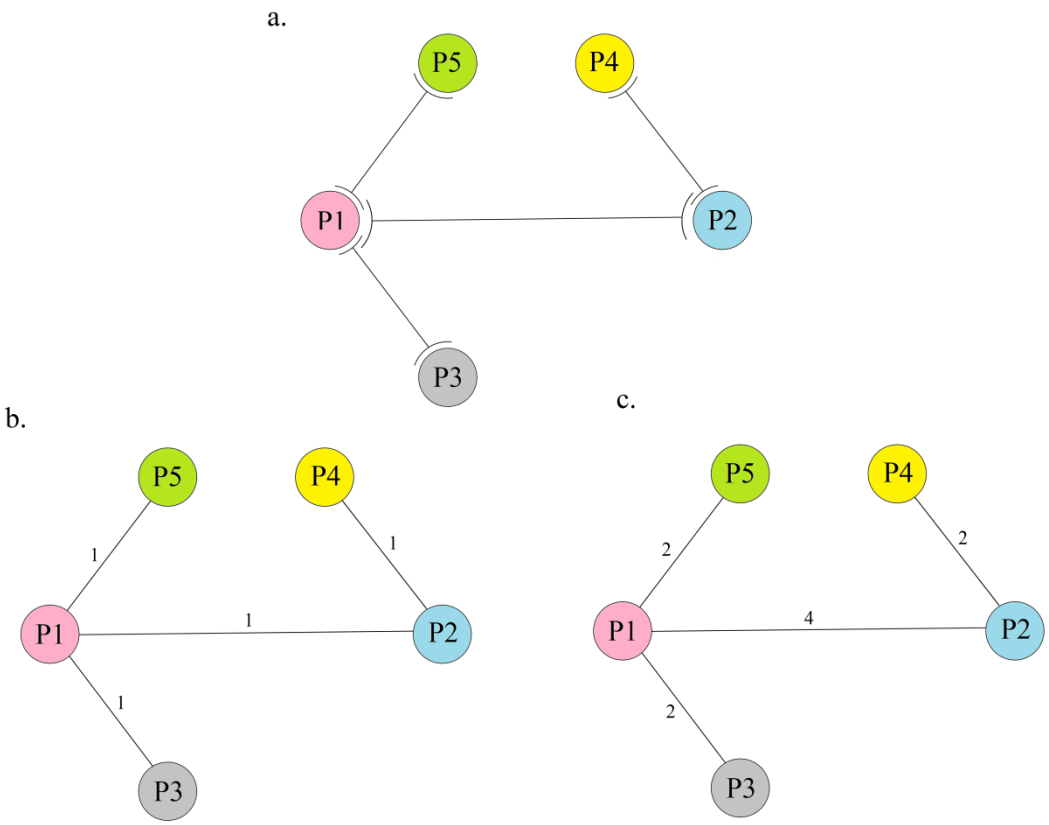

Figure 3-3 An example for assigning edge weights to represent mutually exclusive interactions based on overlapping binding site.

a) P1 protein interacts with P2, P3 and P5 proteins. P1-P5 and P1-P2 interactions and P1-P2 and P1-P3 have overlapping binding sites on P1 protein. P2 protein interacts with P1 and P4 proteins. P2-P1 protein and P2-P4 interaction has overlapping binding site on P2 protein. b) Network representation without overlapping binding site information. Edge weights are initialized to 1. c) After initializing all edges weights to 1, corresponding edge weights of P1-P3, P1-P5 and P1-P2 interactions are increased by 1 since the number of mutually exclusive interactions of them are 1. And corresponding edge weight of P1-P2 interaction is increased by 3 since this interaction cannot co-occur with P1-P3, P1-P5 and P2-P4 interactions.

3.3.2 Flow Betweenness Computation on weighted graph representation of SPPIN

To calculate flow betweenness centrality of p53 SPPIN generated by weight assignment, we converted the edge weights that are assigned by using the method given in the subsection 3.2.1 to edge capacities by applying the following formula.

$$c_e = \frac{1}{w_e} \ ,$$

where w_e is the weight of the edge e and c_e is the calculated capacity of it. This way of capacity assignment gives edges with higher weights lower capacities since edges with higher weights represent bigger number of mutually exclusive interactions.

Proteins exist in a cell in a certain concentration. Different samples of a type of protein may have different binding partners and the interactions they make may be mutually exclusive interactions. We thought that flow betweenness can be a better measure to calculate betweenness based on that fact since flow betweenness is based on flows. Adjusting capacities on the edges with combining mutually exclusive interaction count and expression levels of genes can be a more realistic approach to calculate betweenness as a future study.

3.3.3 Random Walk Betweenness Computation on weighted graph representation of SPPIN

Random walk betweenness of a vertex i is defined as the number of times that a random walk from s to t passes through i, averaged over all s and t. Assume that we have a random walk starting from a source node and making random moves till it finds the target node t. At some point in this walk, assume a node m is visited. Next node to move is selected among adjacent nodes of m with uniformly distributed probability. Therefore, the probability that a node n will be selected in the next move is given by

$$p_{nm} = \frac{1}{k_m} \ ,$$

where k_m is the number of nodes adjacent to m [15].

To reflect structural information, we used edge weights in random walk betweenness calculation. The probability that that a node n will be selected in the next move is distributed proportional to edge weights. It is given by

$$p_{nm} = \frac{w_{mn}}{K_m} ,$$

Where w_{mn} is the weight for edge between m and n and $K_m = \sum_i w_i$ is the sum of the edge weights that are adjacent to m.

Edge weights that are assigned by using the method given in the subsection 3.2.1 are recalculated with the formula below so that adjacent nodes with higher valued edges to reach them will get lower chance of being selected in the random walk betweenness calculation.

Therefore, edge weights for random walk betweenness calculation of structural network is defined as:

$$w'_e = \frac{1}{w_e} ,$$

where w_e is the weight of the edge e and w'_e is the new weight for random walk betweenness computation.

3.3.4 Structural Protein-Protein Interaction Network by Random Sampling

Some of the interactions are categorized as mutually exclusive by using overlapping binding sites on the interfaces predicted by PRISM. To analyze the effect of this structural information on betweenness centrality results, we generated random samples of the network by selecting a set of interactions that can co-occur for each sample.

Each network sample is generated as follows: First, we generated a list (L) of interactions. Each of the interactions in the list has at least one mutually exclusive interaction. Secondly, an interaction is selected among the interactions in that list randomly. Then, we constructed an interaction set S consisting of the interaction selected and the interactions that cannot co-occur with it. Next, we selected an interaction I from S. This selection is done by taking energy results of PRISM into

consideration; that is interactions that have lower energy score get higher chance of being selected since lower energy values reflect the higher reliance on prediction of PRISM. Next, the network is reformed by eliminating the interactions that cannot co-occur with I. After reforming the network, selected interaction I and its mutual exclusive interactions are deleted from the list. Selection an interaction and reforming the network according to this selection steps are repeated till the interaction list is empty.

Therefore, we get a sample network with the interactions that can simultaneously occur. In the Figure 3.3, these steps are shown as a flowchart.

After generating a number of network samples, betweenness centrality is calculated for each of them and results are analyzed.

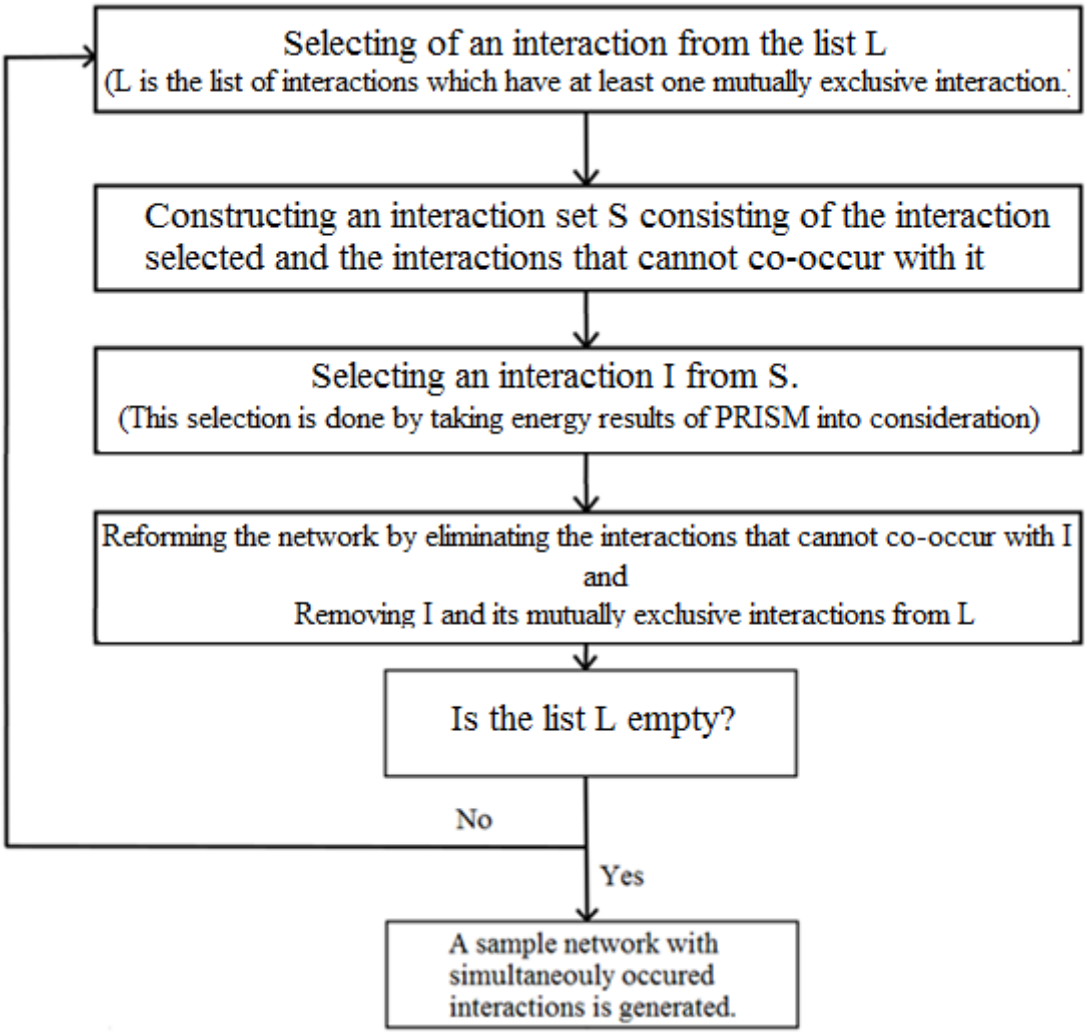

Figure 3-4 Flowchart that shows the steps of generating a network sample with only simultaneously occurred interactions by randomly selection using Prism's energy results

3.4 Definition of Hub and Bottleneck Proteins

Proteins that are in the top 20% of the degree distribution are defined as hubs. Similarly, Bottleneck proteins are defined as the top %20 with highest betweenness scored nodes for all types of betweenness centrality algorithms used in this study. They correspond to ~16 out of 81 proteins in p53 interaction network. Additionally, proteins are categorized

as hub-bottlenecks, hub-nonbottleneck, nonhub-bottleneck and nonhub-nonbottleneck by using hub and bottleneck definitions.

3.5 Generation of Randomized p53 Networks.

We generated five randomized version of p53 networks -by using Randomnetworks plugin of Cytoscape. All of these networks are consisting of 81 nodes. Random Networks plugin generated randomized versions of p53 network by randomly swapping the edges. To assign weights on the edges, we generated random weights according to Gaussian distribution with the mean 0 and variances 1, 4 and 8. Thus, each randomized network has an unweighted and three types of weighted versions.

3.6 Correlation Analysis of Betweenness Algorithms on p53 PPI and SPPI Networks

To analyze the correlation of the flow and random walk betweenness with degree and shortest path betweenness, we found the correlation coefficients and generated scatter plots in the MATLAB.

3.7 Correlation Analysis of Betweenness with Mutations on Interfaces

To analyze the correlation of betweenness with mutation on interfaces, we extracted mutation data from cBioPortal and calculated the number of the mutations on interfaces using the PRISM interface data [37, 38]. Then, we found the correlation coefficients in MATLAB for the correlation of mutations with betweenness scores coming from three betweenness algorithm.

Chapter 4
RESULTS AND DISCUSSION

This chapter includes results and analysis of betweenness centrality algorithms. First correlation of flow and random-walk betweenness with other measures is discussed. Then, correlation of mutations on interfaces with betweenness measures are given. Next, comparison of bottlenecks according to network and algorithm type is given. Finally, biological analysis of high betweenness scored proteins is discussed.

4.1 Correlation of Flow and Random-Walk Betweenness with Other Measures

We compared flow betweenness and random walk betweenness algorithms' results with results of shortest path betweenness and analyzed their correlation with degree distribution in this section.

In the Figure 4.1, correlation of flow betweenness with shortest path betweenness and degree distribution on p53 PPIN are shown as scatter plot.

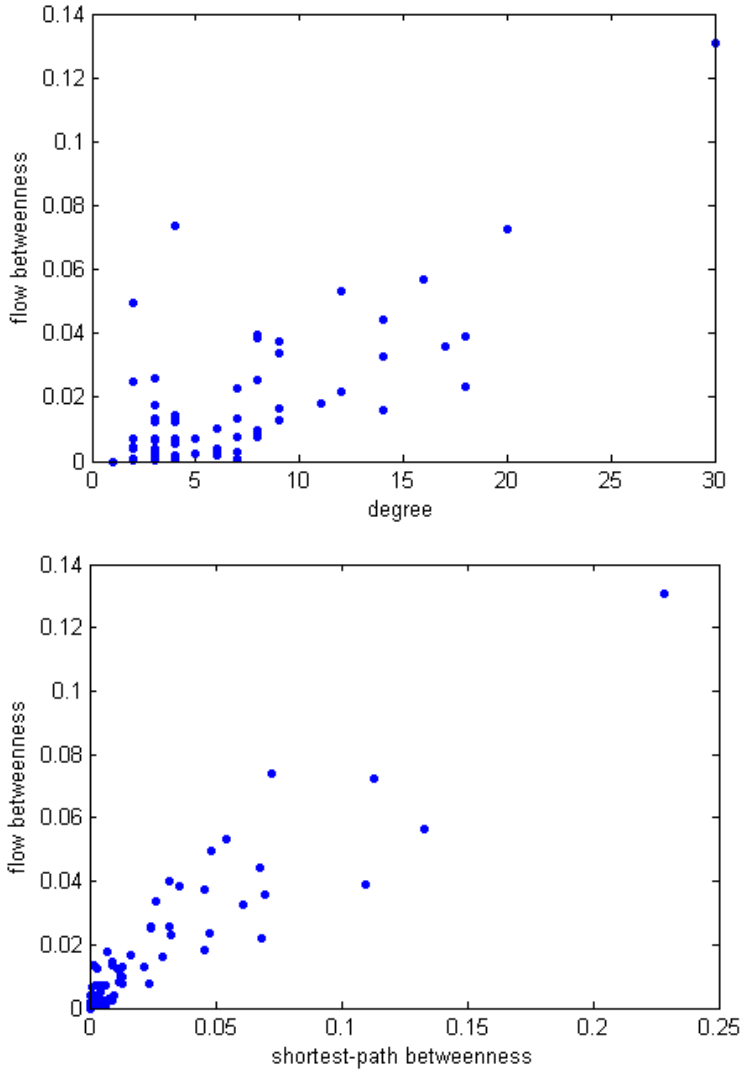

Figure 4-1 Scatter plots of flow betweenness vs degree and flow betweenness vs. shortest-path betweenness of p53 PPIN.

Flow betweenness on p53 PPIN is correlated with shortest path betweenness with correlation coefficient 0.924 and degree with correlation coefficient 0.758 as seen in the Figure 4.1.

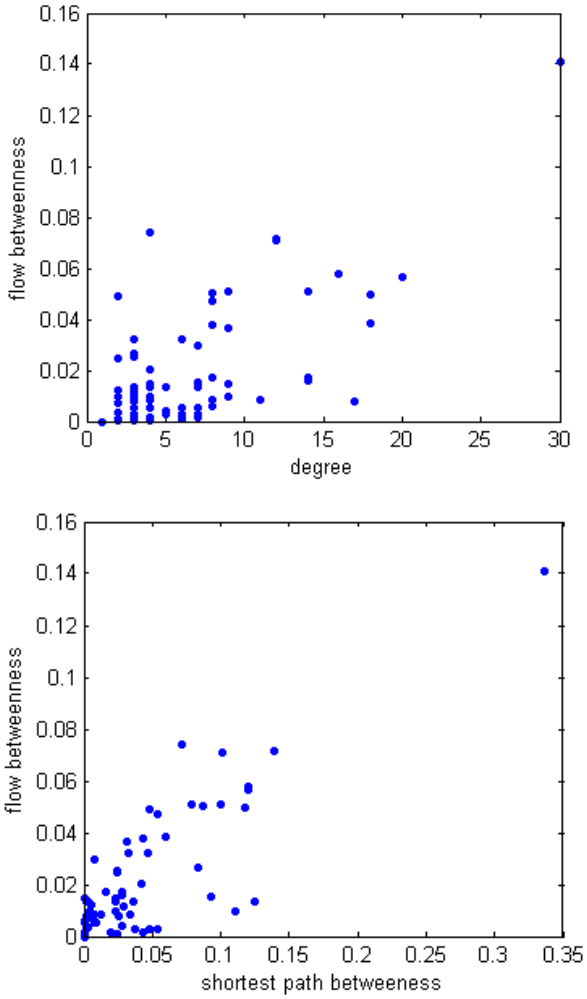

Figure 4-2 Scatter plots of flow betweenness vs degree and flow betweenness vs. shortest-path betweenness of p53 SPPIN.

Flow betweenness on p53 SPPIN is correlated with shortest path betweenness with correlation coefficient 0.848 and degree with correlation coefficient 0.704 as seen in the Figure 4.2.

Shortest Path Betweenness is highly correlated with degree in most networks [39, 40]. There are a small number of vertices which their betweenness scores and degrees are quite different and betweenness is a measure to find these kind of vertices [15]. The correlation between flow betweenness and degree is quite lower than the correlation

between shortest path betweenness and degree. Therefore, we conclude that flow betweenness finds more non-hub bottleneck nodes than shortest path betweenness.

When we look at correlation of random-walk betweenness with degree and shortest path betweenness, we see that it is highly correlated with degree and shortest path betweenness with correlation coefficients 0.956 and 0.945 of p53 PPIN as shown in Figure 4.3.

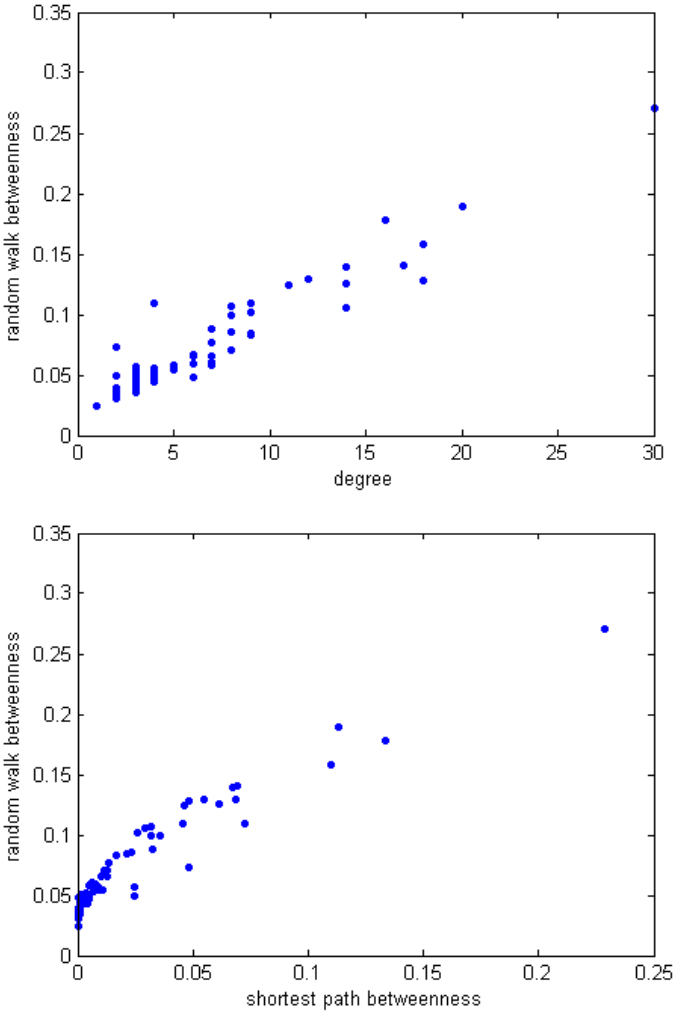

Figure 4-3 Scatter plots of random walk betweenness vs degree and random walk betweenness vs. shortest-path betweenness of p53 PPIN.

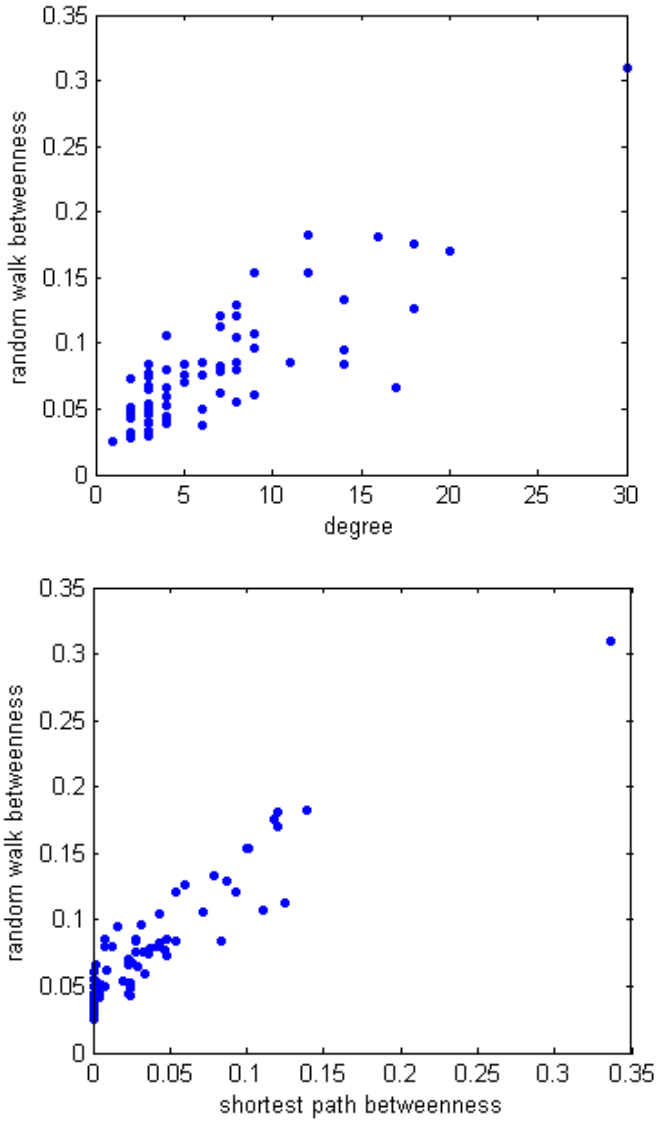

Figure 4-4 Scatter plots of random walk betweenness vs degree and random walk betweenness vs. shortest-path betweenness of p53 SPPIN.

Random-walk betweenness is also highly correlated with degree and shortest path betweenness with correlation coefficients 0.858 and 0.932 of p53 SPPIN as shown in Figure 4.4.

4.2 Correlation of Betweenness with Mutations on Interfaces

We analyzed the correlation of mutations on interfaces with shortest path, flow and random walk betweenness algorithms on unweighted and weighted forms of p53 PPI networks. To do this analysis, we used cBioPortal [37, 38].

All cancer mutations are extracted for each gene using cBioPortal. Next, the mutations on interfaces are identified using PRISM interface data. Finally, we found the correlation of mutations count with various types of betweenness measures on unweighted and weighted p53 networks. Below, in the table 4.1, correlation coefficients are shown.

Unweighted(Nonstructural)	Shortest Path Betw.	0.69
	Flow Betw.	0.63
	Random Walk Betw.	0.69
Weighted(Structural)	Shortest Path Betw.	0.56
	Flow Betw.	0.60
	Random Walk Betw.	0.59

Table 4-1 Correlation coefficients of mutations on interfaces and betweeness measures.

We also calculated the correltion between percentage of mutations on interface residues and betweenness but we didn't observe correlation.

4.3 Analysis of Betweenness Results

Bottleneck nodes on p53 PPIN and SPPIN and random PPIN and SPPIN which are found by using shortest path, flow and random-walk betweenness algorithms are analyzed in this section. Additionally, the results of random sampling method are analyzed and compared with shortest path betweenness algorithm results on p53 SSPIN generated by weight assignment. Next**,** we analyzed Gene Ontology Biological Process

and Molecular Functions Annotations of bottlenecks and compared the importance of bottlenecks and hubs on essentiality in this section.

4.3.1 Comparison of Bottlenecks According to Algorithm Type

Shortest path betweenness algorithm can count some nodes more than ones since many of the paths may share the same set of nodes. To eliminate this, flow betweenness algorithm can be applied as flow betweenness calculation counts only edge disjoint paths [9]. In addition, some nodes have low betweenness score since they are close to high betweenness scored nodes that take all shortest path traffic on themselves as a result of shortest path betweenness algorithm. With flow betweenness, these kind of nodes can be detected [12].

Bottlenecks on Unweighted p53 PPI Network					
Unweighted SP Betw.			**Unweighted Flow Betw.**		
APC	0,228487731	Hub	APC	0,130787084	Hub
CCNE1	0,133329397	Hub	CHEK1	0,073760549	Non-Hub
CDKN1B	0,112881364	Hub	CDKN1B	0,072493459	Hub
CDK2	0,109811812	Hub	CCNE1	0,056732663	Hub
CHEK1	0,072192696	Non-Hub	RPA3	0,05325874	Hub
E2F1	0,069472154	Hub	CDKN2A	0,049367089	Non-Hub
PARP1	0,068608224	Hub	XRCC5	0,044385416	Hub
XRCC5	0,06750204	Hub	**CKS1B**	0,039796187	Non-Hub
RFC5	0,061030623	Hub	CDK2	0,038917714	Hub
RPA3	0,054635561	Hub	RAF1	0,038699895	Non-Hub
CDKN2A	0,048148148	Non-Hub	GADD45A	0,037520344	Hub
MYC	0,047971529	Hub	E2F1	0,03591294	Hub
CDKN2D	0,045999957	Hub	**CDK4**	0,033673117	Hub
GADD45A	0,045801936	Hub	RFC5	0,032782119	Hub
RAF1	0,035941671	Non-Hub	**CDK5**	0,025896624	Non-Hub
SFN	0,032369354	Non-Hub	**SKP2**	0,025574518	Non-Hub

Table 4-2 Bottleneck nodes on p53 PPIN as a result of shortest path and flow betweenness algorithm. Different genes are bolded.

Nodes that entered the top %20 high betweenness score list with flow betweenness algorithm on p53 PPIN are CKS1B, CDK4, CDK5 and SKP2 and on p53 SPPIN are CDKN2A, RAF1, CDK4 and RELA.

Bottlenecks on Unweighted p53 PPI Network					
Unweighted SP Betw.			**Unweighted Random Walk Betw.**		
APC	0,228487731	Hub	APC	0.2709490529891378	Hub
CCNE1	0,133329397	Hub	CDKN1B	0.18924280795780773	Hub
CDKN1B	0,112881364	Hub	CCNE1	0.17844539119457375	Hub
CDK2	0,109811812	Hub	CDK2	0.15839400473831144	Hub
CHEK1	0,072192696	Non-Hub	E2F1	0.14075031412249678	Hub
E2F1	0,069472154	Hub	XRCC5	0.1388126268618467	Hub
PARP1	0,068608224	Hub	RPA3	0.12982533794110993	Hub
XRCC5	0,06750204	Hub	PARP1	0.12902421383261617	Hub
RFC5	0,061030623	Hub	MYC	0.12863249622166056	Hub
RPA3	0,054635561	Hub	RFC5	0.1252206960159592	Hub
CDKN2A	0,048148148	Hub	CDKN2D	0.12403692960092805	Hub
MYC	0,047971529	Hub	GADD45A	0.10997658468142453	Hub
CDKN2D	0,045999957	Hub	CHEK1	0.1088880936310862	Non-Hub
GADD45A	0,045801936	Hub	**SKP2**	0.10626625874283549	Non-Hub
RAF1	0,035941671	Non-Hub	**EP300**	0.10510129926439911	Hub
SFN	0,032369354	Non-Hub	**CDK4**	0.10168422003080997	Hub

Table 4-3 Bottleneck nodes on p53 PPIN as a result of shortest path and random walk betweenness algorithm. Different genes are bolded.

Nodes that are in bottleneck list of random walk betweenness results but not in shortest path betweenness results of p53 PPIN are SKP2, EP300 and CDK4. On p53 SPPIN only difference between bottleneck list of shortest path and random walk betweenness is RAF1.

4.3.2 Comparison of Bottlenecks According to Network Type

Betweenness centrality results are calculated on non-structural p53 network, structural p53 network modelled by assigning weights, five random networks generated by randomly swapping edges on weighted p53 network and scale free network by using shortest path, flow and random walk betweenness centrality algorithms. Only bottleneck nodes which corresponds the top %20 of the nodes with highest betweenness are considered in this analysis.

Bottleneck Nodes - Shortest Path Betweenness					
Unweighted			**Weighted**		
APC	0,228487731	Hub	APC	0,336441237	Hub
CCNE1	0,133329397	Hub	RPA3	0,139476645	Hub
CDKN1B	0,112881364	Hub	**RPA1**	0,125139295	Non-Hub
CDK2	0,109811812	Hub	CDKN1B	0,120344327	Hub
CHEK1	0,072192696	Non-Hub	CCNE1	0,11946163	Hub
E2F1	0,069472154	Hub	MYC	0,117693592	Hub
PARP1	0,068608224	Hub	**RB1**	0,110337875	Hub
XRCC5	0,06750204	Hub	PARP1	0,101458745	Hub
RFC5	0,061030623	Hub	GADD45A	0,100323629	Hub
RPA3	0,054635561	Hub	**CDK7**	0,092480955	Non-Hub
CDKN2A	0,048148148	Non-Hub	**SKP2**	0,086666956	Non-Hub
MYC	0,047971529	Hub	**BRCA1**	0,083094136	Non-Hub
CDKN2D	0,045999957	Hub	XRCC5	0,078957532	Hub
GADD45A	0,045801936	Hub	CHEK1	0,071296296	Non-Hub
RAF1	0,035941671	Non-Hub	CDK2	0,059476044	Hub
SFN	0,032369354	Non-Hub	**CKS1B**	0,054224966	Non-Hub

Table 4-4 Bottleneck nodes with betweenness scores and hub information on p53 PPIN and SPPIN as a result of shortest path algorithm. Nodes that are in bottleneck nodes of unweighted but not in bottleneck nodes of weighted and vice versa are bolded to show the difference.

In the Table 4.4, bottleneck nodes are listed as the results of shortest path betweenness centrality algorithm on p53 PPIN and SPPIN. % 37.5 of genes that are present in the top 16 genes with high shortest path betweenness list is different from the genes found by unweighted betweenness analysis. The betweenness scores of APC, PARP1, RPA3, MYC and GADD45A genes on SPPIN increased remarkably compared to scores of same genes on p53 PPIN .

Assigning weights to edges of p53 network to represent structural feature changed the hub characteristics of bottleneck nodes. The number of nonhub-bottlenecks on weighted p53 network in shortest path betweenness analysis is higher compared to unweighted p53 network. Five out of six nodes that entered the bottleneck node list after assigning weights are non-hub nodes. In the figure 4.5 below, the positions of six nodes that entered the bottleneck list after assigning weight are shown as highlighted.

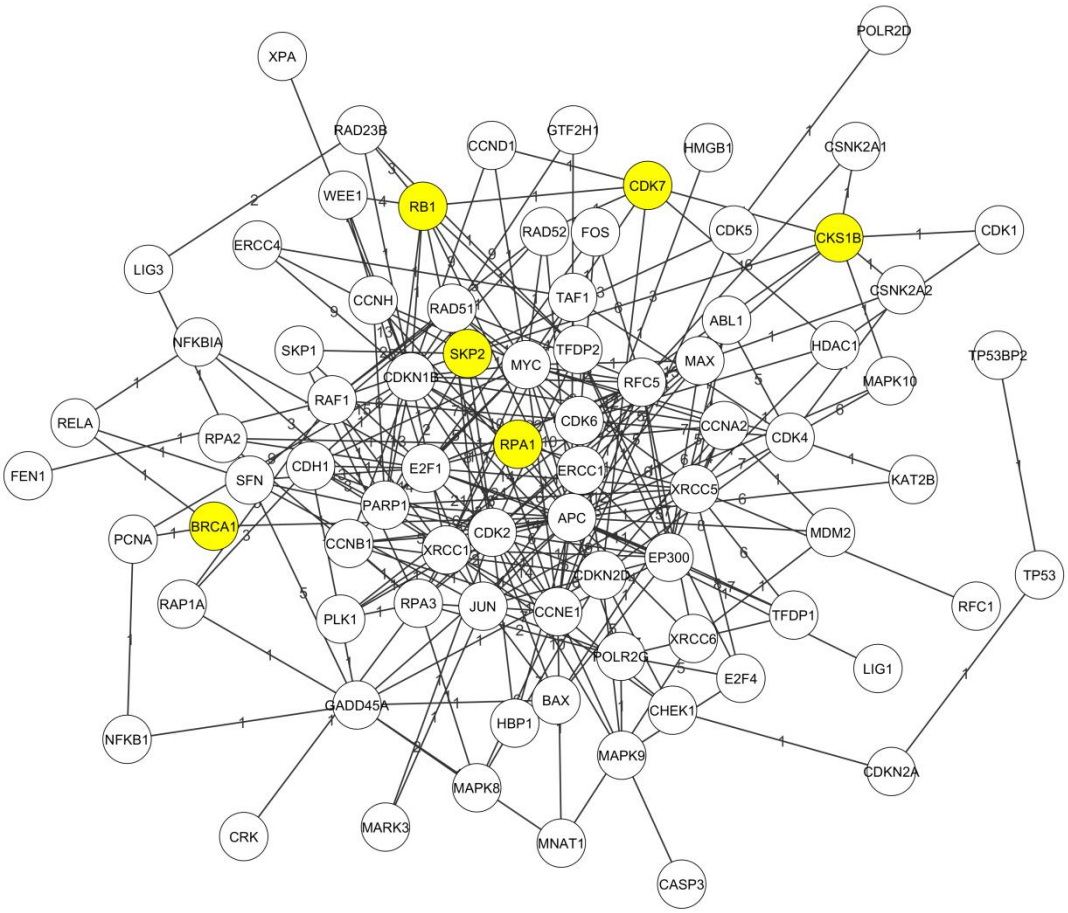

Figure 4-5 Positions of six nodes that entered the bottleneck list after assigning weight with shortest path betweenness analysis

Highlighted nodes RPA1, RB1, CDK7, CKS1B, SKP2 and BRCA1 in Figure 4.5 are in the bottleneck nodes on p53 SPPIN which are not in the bottleneck node list of p53 PPIN. All of these nodes except for RB1 are non-hubs. In other words, these nodes have low number of neighbors but have critical positions and act as bottlenecks in the network and are detected taking weights into consideration in betweenness analysis.

Bottleneck Nodes -Flow Betweenness					
Unweighted			**Weighted**		
APC	0,130787084	Hub	APC	0,140650985	Hub
CHEK1	0,073760549	Non-Hub	CHEK1	0,07393809	Non-Hub
CDKN1B	0,072493459	Hub	RPA3	0,071440439	Hub
CCNE1	0,056732663	Hub	**PARP1**	0,070805883	Hub
RPA3	0,05325874	Hub	CCNE1	0,057766875	Hub
CDKN2A	0,049367089	Non-Hub	CDKN1B	0,056746937	Hub
XRCC5	0,044385416	Hub	XRCC5	0,050968402	Hub
CKS1B	0,039796187	Non-Hub	GADD45A	0,050665316	Hub
CDK2	0,038917714	Hub	SKP2	0,05041865	Non-Hub
RAF1	0,038699895	Non-Hub	**MYC**	0,049601459	Hub
GADD45A	0,037520344	Hub	CDKN2A	0,049367089	Non-Hub
E2F1	0,03591294	Hub	CKS1B	0,047033604	Non-Hub
CDK4	0,033673117	Hub	CDK2	0,038572002	Hub
RFC5	0,032782119	Hub	RAF1	0,037712229	Non-Hub
CDK5	0,025896624	Non-Hub	CDK4	0,036415156	Hub
SKP2	0,025574518	Non-Hub	**RELA**	0,032353563	Non-Hub

Table 4-5 Bottleneck nodes with betweenness scores and hub information on p53 PPIN and p53 SPPIN as a result of flow betweenness algorithm. Nodes that are in bottleneck nodes of unweighted but not in bottleneck nodes of weighted and vice versa are bolded to show the difference.

In the Table 4.5, bottleneck nodes are listed as the results of flow betweenness centrality algorithm on p53 PPIN and SPPIN. % 18.75 of genes that are present in the top 16 genes with high flow betweenness list is different from the bottleneck genes found by unweighted betweenness analysis.

Bottleneck Nodes -Random Walk Betweenness					
Unweighted			**Weighted**		
APC	0.2709490529891378	Hub	APC	0.3089864506355171	Hub
CDKN1B	0.18924280795780773	Hub	RPA3	0.18314570438311695	Hub
CCNE1	0.17844539119457375	Hub	CCNE1	0.18162897597629435	Hub
CDK2	0.15839400473831144	Hub	MYC	0.17636714622085156	Hub
E2F1	0.14075031412249678	Hub	CDKN1B	0.16999421034689813	Hub
XRCC5	0.1388126268618467	Hub	PARP1	0.1542223111368365	Hub
RPA3	0.12982533794110993	Hub	GADD45A	0.15320895921605712	Hub
PARP1	0.12902421383261617	Hub	XRCC5	0.13337254936160972	Hub
MYC	0.12863249622166056	Hub	SKP2	0.12862624009629184	Non-Hub
RFC5	0.1252206960159592	Hub	CDK2	0.12652540030781836	Hub
CDKN2D	0.12403692960092805	Hub	**CDK7**	0.12059937110891275	Non-Hub
GADD45A	0.10997658468142453	Hub	**CKS1B**	0.12045372501107059	Non-Hub
CHEK1	0.1088880936310862	Non-Hub	**RPA1**	0.11278828548896348	Non-Hub
SKP2	0.10626625874283549	Non-Hub	**RB1**	0.10752909028993692	Hub
EP300	0.10510129926439911	Hub	CHEK1	0.10551668992226874	Non-Hub
CDK4	0.10168422003080997	Hub	**RAF1**	0.10516915911193937	Non-Hub

Table 4-6 Bottleneck nodes with betweenness scores and hub information on p53 PPIN and SPPIN as a result of random walk betweenness algorithm. Nodes that are in bottleneck nodes of nonstructural but not in bottleneck nodes of structural network and vice versa are bolded to show the difference.

In the Table 4.6, bottleneck nodes are listed as the results of random walk betweenness centrality algorithm on p53 PPIN and SPPIN. % 31.25 of genes that are present in the top 16 genes with high random walk betweenness list is different from the genes found by unweighted betweenness analysis. In addition, almost all of the nodes that entered the bottleneck node list after weight assignment are the non-hub nodes. So, we conclude

that by assigning weights on p53 PPIN to represent mutually exclusive interactions, four nodes that are non-hubs are defined as bottleneck with random-walk betweenness algorithm.

Shortest path, flow and random walk betweenness centralities are computed for unweighted and weighted forms of five random networks generated by swapping edges of p53 network. Edge weights of these networks are determined by Gaussian distribution with variances 1, 4 and 8. The table 4.7 shows the percentage difference between bottleneck nodes of unweighted and weighted forms of these networks.

	Shortest Path Betw.			**Flow Betw.**			**Random Walk Betw.**		
	Var=1	**Var=4**	**Var=8**	**Var= 1**	**Var=4**	**Var=8**	**Var=1**	**Var=4**	**Var=8**
Random Network 1	%12,5	%31,25	%25	%6,25	%12,5	%12,5	%6,25	%25	%12,5
Random Network 2	%12,5	%31,25	%31,25	%0	%12,5	%25	%0	%12,5	%12,5
Random Network 3	%12,5	%43,75	%18,75	%12,5	%6,25	%31,25	%0	%31,25	%18,75
Random Network 4	%6,25	%25	%31,25	%6,25	%18,75	%12,5	%0	%18,75	%25
Random Network 5	%12,5	%31,25	%31,25	%6,25	%25	%6,25	%0	%18,75	%6,25

Table 4-7 Table shows the percentage difference between bottleneck nodes of weighted and unweighted forms of randomized networks. Weights of edges are determined randomly by Gaussian distribution with the variances 1, 4, 8 and percentage differences for all of them are listed in the table separately.

For random networks generated by randomly swapping edges of p53 network, the percentage differences are similar to those of p53 network.

4.3.3 Analysis of Random Sampling Method's Results

We created 20000 networks by applying the random sampling method explained in chapter 3 and ran shortest path betweenness algorihtm on each one of these networks. For each network, we ordered the nodes according to betweenness centrality scores in the descending order and found the occurences count for each gene in the top 20 high betweenness scored gene list. This analysis is also done for top 1, 5 and 10 occurence count. For each occurrence analysis new 20000 networks were created and betweenness of their nodes calculated.

Occurrence Count in the Top Twenty Lists			
Gene	**Count**	**Gene**	**Count**
RPA3	20000	CCNA2	4258
APC	20000	TFDP1	4134
RPA1	20000	MAPK10	4119
GADD45A	20000	RAD52	3614
CHEK1	20000	PLK1	2467
MYC	20000	CDKN2D	2447
CDKN1B	20000	MARK3	1088
CCNE1	19998	SFN	1039
CDK7	19609	RAD23B	894
PARP1	19581	HDAC1	829
CDKN2A	19554	BAX	790
XRCC5	19276	E2F1	380
CKS1B	19137	CDH1	222
CDK2	18169	TAF1	136
RB1	17871	ERCC1	117
SKP2	17599	CDK6	94
BRCA1	15694	CSNK2A2	87
RAF1	10984	TFDP2	81
RFC5	7984	XRCC1	73
POLR2G	7196	JUN	71
CCNB1	6005	MNAT1	7
EP300	5091	NFKBIA	5
RELA	4992		
CDK4	4308		

Table 4-8 Occurrence counts of the genes in the top 20 high betweenness scored gene lists of 20000 networks generated by applying random sampling method to p53 PPIN

Occurrence Count in the Top Ten Lists		Occurrence Count in the Top Five Lists	
Gene	**Count**	**Gene**	**Count**
APC	20000.0	APC	20000.0
RPA3	19938.0	RPA3	14998.0
CDKN1B	19815.0	PARP1	14576.0
MYC	19799.0	CCNE1	13939.0
GADD45A	19161.0	MYC	13366.0
CCNE1	17898.0	CDKN1B	10065.0
PARP1	17360.0	GADD45A	3944.0
RPA1	14137.0	**SKP2**	2332.0
CHEK1	11700.0	CDK2	1915.0
SKP2	9809.0	**RPA1**	1748.0
CDK2	7656.0	**RB1**	1550.0
RB1	6644.0	**BRCA1**	1259.0
XRCC5	5316.0	XRCC5	179.0
BRCA1	4983.0	RELA	52.0
CDK7	1853.0	TFDP1	32.0
TFDP1	1763.0	EP300	21.0
RELA	753.0	POLR2G	13.0
EP300	338.0	CCNB1	5.0
POLR2G	331.0	CHEK1	4.0
CCNB1	286.0	CDK7	2.0
RFC5	202.0		
CKS1B	123.0		
CDKN2D	92.0		
E2F1	19.0		
CCNA2	14.0		
RAF1	7.0		
PLK1	2.0		
SFN	1.0		

Table 4-9Occurrence counts of the genes in the top 5 and 10 high betweenness scored gene lists of 20000 networks generated by applying random sampling method to p53 PPIN

In the table 4.8 and 4.9, occurrence counts of the genes in the top 20, 10 and 5 high betweenness scored gene lists of 20000 networks generated by random sampling of p53 PPIN. The genes which are not bottlenecks as a result of shortest path betweenness centrality on unweighted p53 network but have high number of occurrence count in top 20, 10 or 5 lists are highlighted. RPA1, CDK7, CKS1B, RB1, SKP2 and BRCA1 are

the nodes which are detected as bottlenecks with shortest path betweenness centrality algorithm after assigning weights on p53 PPIN. RAF1 is found as bottleneck with random walk betweenness centrality on weighted p53 PPIN. Therefore, we conclude that weights assignment is a successful way to detect bottleneck on the p53 PPIN with mutually exclusive interactions as a structural feature. In addition APC gene always has the highest betweenness score on each type of betweenness centrality measures of all 20000 networks.

4.3.4 Essential Genes

Yu et al analyzed the importance of bottleneck and hub proteins and their correlation with gene essentiality in the yeast PPI interaction and regulatory networks [31]. They have showed that bottlenecks are tend to be essential genes in both regulatory and interaction networks. They also found that the betweenness is stronger determinant on essentiality than degree on regulatory networks, whereas, in contrast to regulatory network, hubs tend to be more essential than bottlenecks in interaction networks.

Kar et al analyzed human cancer PPI network from a structural perspective [34]. They conducted a similar analysis as Yu et al did on PPI and structural PPI network. They classified proteins into four types; hub-bottleneck, hub-nonbottleneck, nonhub-bottleneck and nonhub-nonbottleneck and calculated the fraction of essential genes. They have observed that hub-nonbottlenecks are more essential than nonhub-bottlenecks. Therefore, they concluded that degree is a more important determinant on essentiality in their networks as well.

We also analyzed unweighted and weighted p53 PPIN in the same way Kar et al and Yu et al did. Since we have three different betweenness centrality measures which are shortest path, flow, random walk betweenness, we analyzed bottlenecks found by each of them separately in terms of essentiality.

In the Figure 4.6 and 4.7, fractions of essential genes found by shortest path, flow and random walk betweenness algorithms on p53 PPIN and SPPIN are seen. We observed that for both p53 PPIN and SPPIN, fraction of essential genes in bottlenecks is not higher than that of non-bottlenecks. Therefore, we cannot state that bottlenecks tend to be essential in both p53 PPIN and SPPIN.

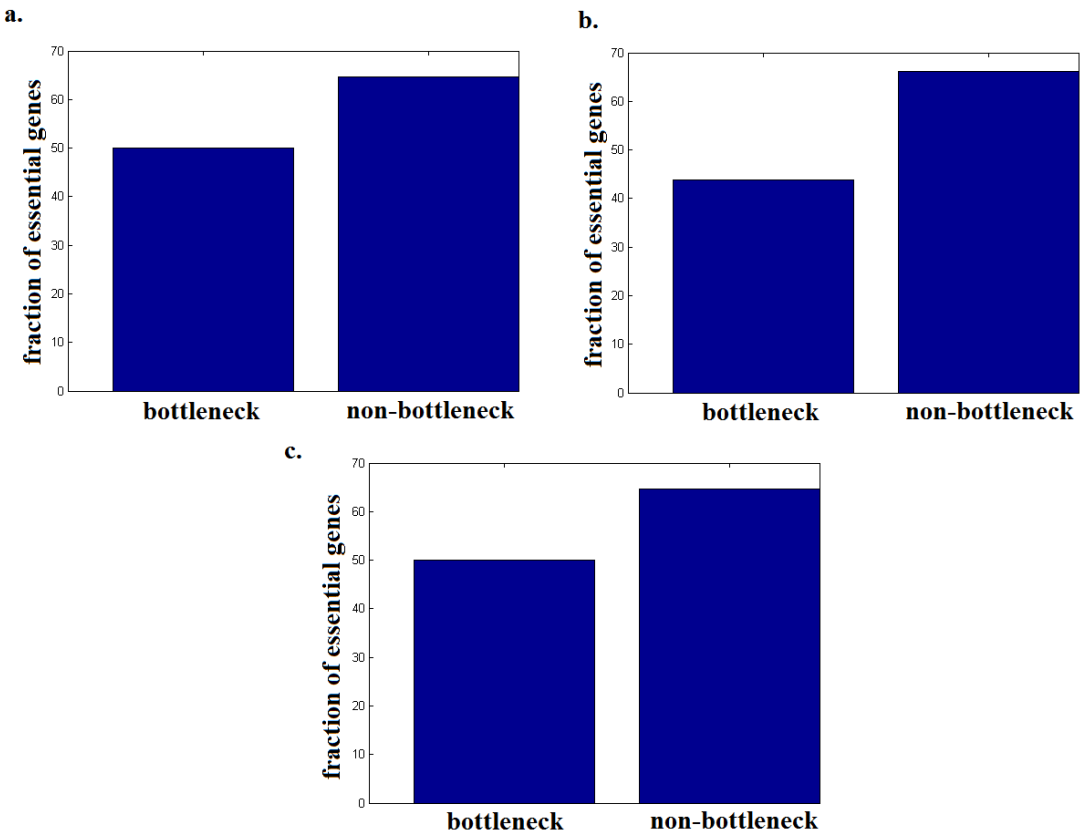

Figure 4-6 Fractions of essential genes in bottlenecks and non-bottlenecks found by a) shortest path b) flow c) random walk betweenness algorithms on p53 PPIN

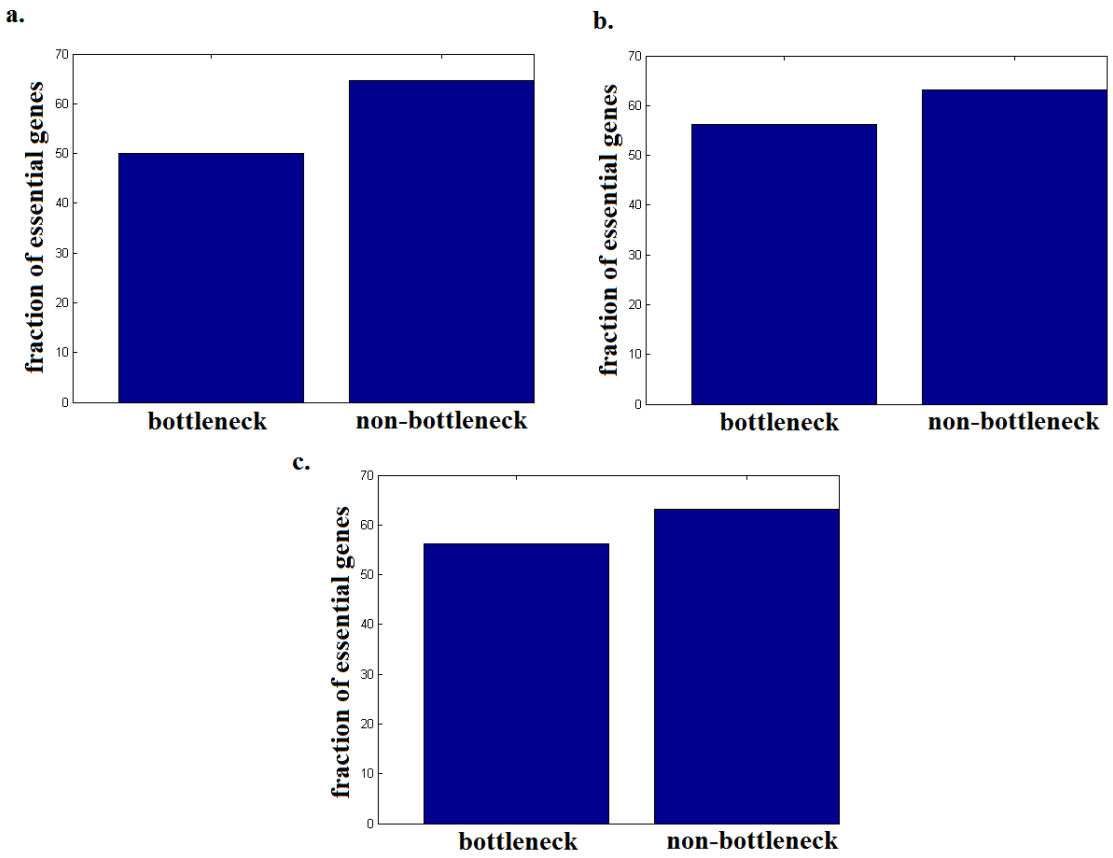

Figure 4-7 Fractions of essential genes in bottlenecks and non-bottlenecks found by a) shortest path b) flow c) random walk betweenness algorithms on p53 network SPPIN

4.3.5 Gene Ontology Biological Process and Molecular Function Annotations of Bottlenecks

We analyzed Gene Ontology Biological Process and Molecular Functions of the bottleneck genes that found by shortest path, flow and random walk betweenness algorithms on unweighted and weighted forms of p53 PPI networks. To do this analysis, we used BINGO plugin of Cytoscape [35, 41]. These biological processes and molecular functions are in a hierarchical tree structure.

The common biological functions of the bottleneck genes found by shortest path betweenness centrality algorithm on unweighted p53 PPI network are related to cell cycle, DNA stimulus. On weighted p53 PPI network, DNA repair is included as a critical biological function of the bottleneck nodes while on unweighted network, only parent of DNA repair function exists.

When we do the same analysis for flow betweenness algorithm results, the common biological functions of bottlenecks of both unweighted and weighted p53 network are generally related to cell cycle.

In random walk betweenness centrality results, cell cycle is again a critical biological function of bottlenecks of both unweighted and weighted p53 network and DNA repair and stimulus are included as common biological functions for bottlenecks of weighted p53 network.

We also analyzed the Molecular Function Annotations of the bottleneck genes found by running shortest path, flow and random walk betweenness algorithms on both unweighted and weighted p53 PPI network. Kinase activity is the most common molecular function for the bottlenecks found by shortest path algorithm on unweighted network. For weighted network, bottleneck nodes have hormone receptor binding and transcription activation besides kinase activity as molecular functions.

For flow betweenness, kinase activity is again most common molecular function for both weighted and unweighted network's bottlenecks. NF-kappaB, promoter and DNA regulatory region binding are defined as molecular function of bottlenecks of only weighted p53 network as a result of that RELA is in the bottleneck list.

For random walk betweenness, molecular function analysis of bottleneck nodes revealed quite similar results with shortest path betweenness. Thus, the bottleneck nodes of both unweighted and weighted networks have kinase activity molecular function. However, several hormone binding functions is seen for only weighted network's bottlenecks.

4.4 Case Studies

4.4.1 SKP2: S-phase kinase-associated protein 2

SKP2 is an example non-hub protein in p53 PPIN network which entered the bottleneck list with flow and random-walk betweenness calculation. SKP2 functions in cell cycle regulation [42]. Specifically, SKP2 plays role in the cyclin A-CDK2 S-phase kinase and it recognizes phosphorylated cyclin-dependent kinase inhibitor 1B (CDNK1B) [43-45].

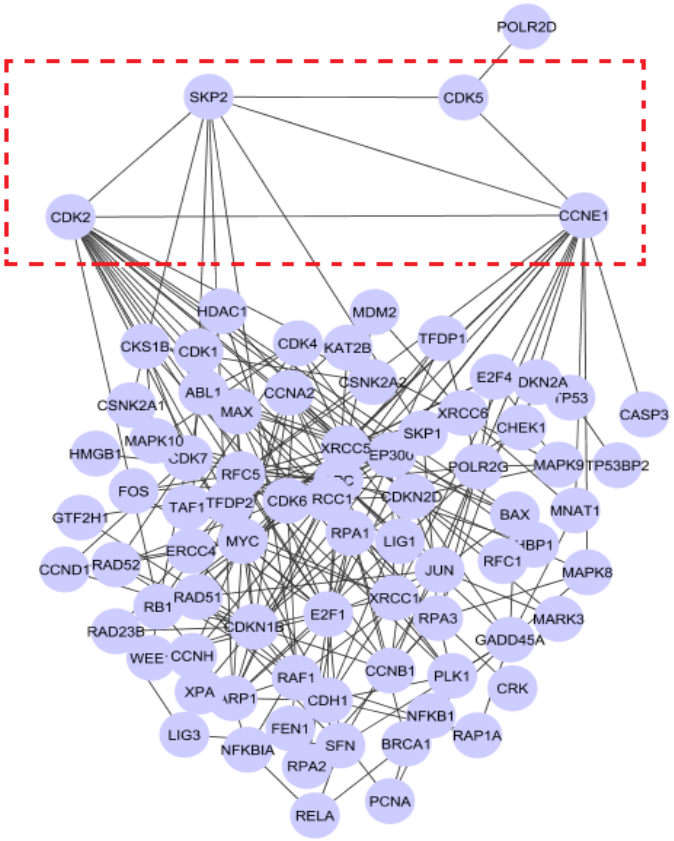

Figure 4-8 Position of SKP2 in p53 PPIN

In the Figure 4.7 position of SKP2 in the p53 is seen. SKP2 interacts with two hub-bottleneck proteins CCNE1 and CDK2. With shortest path betweenness measure, SKP2

is not found as a bottleneck since shortest-path betweenness algorithm selects the edge between CCNE1 and CDK2 as a shortest path and the path going through SKP2 is missed. According to flow and random-walk betweenness measures, SKP2 is bottleneck on p53 PPIN.

In p53 signaling pathway, SKP2 acts in degradation of cyclin-dependent kinase inhibitor 1B (CDKN1B). CDKN1B is an enzyme inhibitor which inhibits cyclin E-CDK2 (CCNE1-CDK2) complexes and blocks the cell cycle in the G0/G1 phase [46]. In the figure 4.7, SKP2's role in p53 signaling pathway is seen schematically.

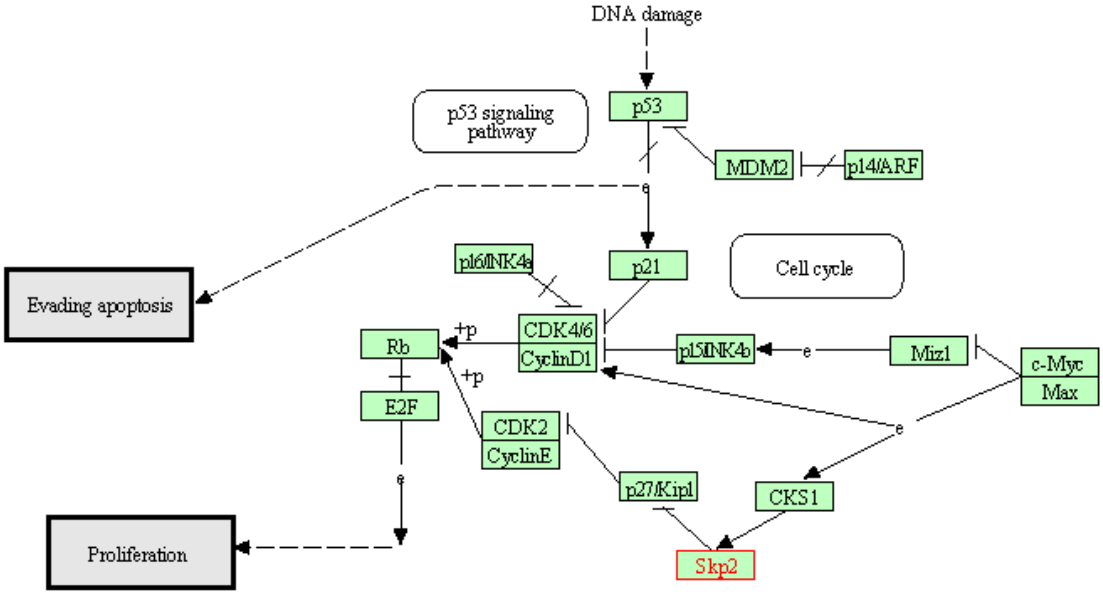

Figure 4-9 SKP2 in p53 signaling pathway from KEGG database [47].

4.4.2 RB1: Retinoblastoma Protein

Retinoblastoma Protein RB1 is a hub protein in p53 PPIN that is not bottleneck list without adding structural information on p53 network. After generating structural p53 PPIN, it is detected as a bottleneck with shortest-path and random-walk betweenness algorithms.

All interactions of RB1 in our p53 PPIN can simultaneously occur. Therefore, the weights of representative edges of these interactions are set to 1. As a result of this, shortest-path betweenness algorithm selects more paths going through RB1 so that it enters the bottleneck list. RB1 is also detected as a bottleneck by random-walk betweenness algorithm and has high occurrence counts in the top 5 and 10 high betweenness scored gene lists of 20000 networks generated by applying random sampling method to p53 PPI network.

RB1 protein is a tumor suppressor protein and it has dysfunctional role in several major cancers [48]. RB1 inhibits cell cycle progression and prevents excessive cell growth. RB1's activity is inhibited by CDK4/CDK6-CCDN1 and CDK2-CCNE1 complexes, when a cell is to enter S phase [49-51].

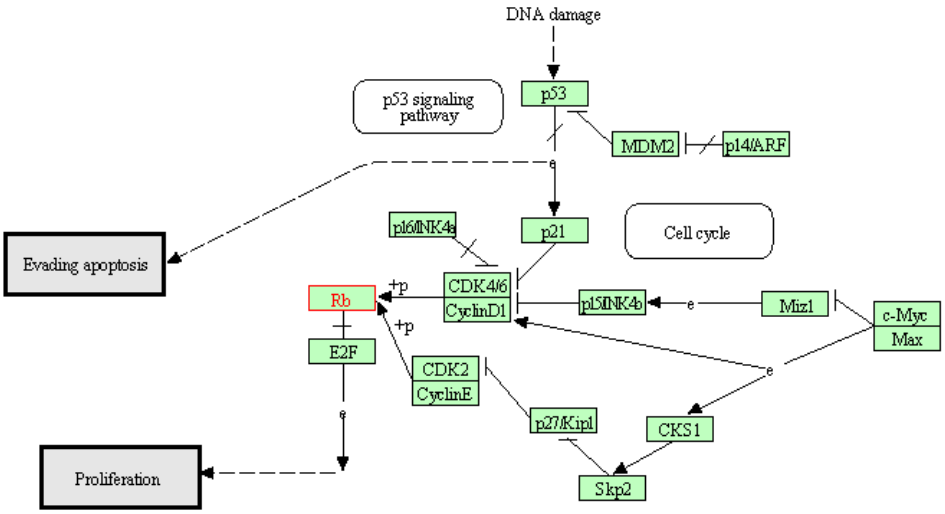

Figure 4-10 RB1 in p53 signaling pathway from KEGG database [47].

4.4.3 RPA1: Replication Protein A

RPA1 is a non-hub protein that is not bottleneck list with shortest-path and random-walk betweenness calculations in p53 PPIN. It is detected as bottleneck with shortest-path and random-walk betweenness computation on p53 SPPIN.
RPA1 protein is important for DNA replication and repair and mutations on its coding gene can cause cancer [52].

Chapter 5
CONCLUSION

Proteins interact each other to function in biological process and act in molecular functions. Protein-protein interaction networks are abstractions to represent the map of interactions among proteins. Betweenness centrality measure helps to understand importance of proteins in an PPIN.

A PPIN does not reflect structural and chemical features of interactions. In literature, there a studies on PPINs taking structural information into consideration. Kim et al [4]. reflected structural information on protein-protein interactions and Tuncbag et al. analyzed mutually exclusive and simultanesouly occurable interactions using PRISM algorithm [5, 6]. Demircioglu et al. modelled p53 structural protein-protein interaction network for betweenness centrality analysis [7].

In this work, we aimed modelling structural form of p53 network usind the data from Demircioglu's study and analyze betweenness on it using shortest path, flow and random walk betweenness centrality algortihms.

We found that bottleneck nodes in p53 SPPIN are different from p53 PPIN found by three different betweenness centrality algoritm. Therefore, we concluded that reflecting structural information on p53 PPIN changed the betweeness characteristics of nodes. Bottleneck nodes found on p53 SPPIN by shortest path betweenness algorithm are different from those found on p53 PPIN by % 37.5. We observed this percentage as %18.75 for flow betweenness algorithm and % 31.5 for random walk betweenness algorithm. In addition, flow betweenness algorithm detected more non-hub bottleneck nodes than shortest path betweenness algorithm.

We also created 20000 random network samples in which there are only simultaneous interactions and run shortest path betweenness algorithm on each of them to get the proteins which mostly have high betweenness score. We observed that those proteins compromise with bottleneck proteins found on p53 SSPIN.

Yu et al have showed that bottlenecks are tend to be essential genes in interaction networks [31]. We also analyzed betweenness centrality on p53 PPIN and SPPIN in terms of essentiality. We did not observe that fraction of essential genes in bottlenecks is higher than that of non-bottlenecks for both p53 PPIN and SPPIN.

The correlation betweenness with number of mutations on interface structures are also analyzed. Results showed that shortest path, flow and random walk betweenness measures are positively correlated with number of mutations on interfaces.

We gave SKP2 as an example non-hub protein in p53 PPIN which entered the bottleneck list with flow and random-walk betweenness calculation. SKP2 plays role in the cyclin A-CDK2 S-phase kinase and it recognizes phosphorylated cyclin-dependent kinase inhibitor 1B (CDNK1B) [42-45].

Retinoblastoma Protein RB1 and Replication Protein RPA1 are hub proteins detected as a bottleneck with shortest-path and random walk betweenness algorithms after generating p53 SPPIN. RB1 has important roles in cancers and cell progression [48-51] and RPA1 is important for DNA replication and repair [52].

APPENDIX

Table A-1 p53 SPPIN generated by weight assignment

Node	Edge Weight	Node	Node	Edge Weight	Node
CCNE1	5	APC	JUN	5	POLR2G
CHEK1	7	CDKN2D	CHEK1	4	CCNE1
MAPK8	7	CDKN2D	CDK5	5	CCNE1
CDK2	13	XRCC5	CDK2	14	CCNE1
CDK4	9	CDKN2D	CDK2	1	MARK3
CDK4	7	XRCC5	CDK2	5	CCNB1
MAPK9	8	CDKN2D	NFKBIA	13	E2F1
CDK6	4	CKS1B	SFN	13	E2F1
CDK2	4	CKS1B	APC	3	PARP1
RAP1A	5	RAF1	E2F4	5	POLR2G
E2F4	5	TFDP2	TP53	1	TP53BP2
CDK6	7	RAF1	APC	1	HBP1
RAD51	8	JUN	APC	1	MAPK10
EP300	16	E2F1	APC	1	XRCC1
XRCC5	4	CCNA2	APC	1	GTF2H1
APC	9	CCNH	APC	1	RAD52
SKP2	2	SKP1	APC	1	FOS
APC	5	MAX	APC	1	CCNB1
CDK6	7	CDKN2D	APC	1	BRCA1
CCNA2	5	EP300	APC	1	HMGB1
RFC5	4	RAD51	APC	1	LIG1
SFN	7	RAD51	APC	1	TFDP1
PARP1	2	SKP2	APC	1	RFC5
E2F1	18	MAX	APC	1	RB1
CDKN2D	14	CDKN1B	CDK2	1	SKP2
RB1	5	MYC	CDK2	1	PLK1
CCNH	6	PARP1	CDK2	1	HBP1
RFC5	19	CDKN1B	CDKN1B	1	CCNE1
GTF2H1	9	CDKN1B	CDKN1B	1	CCNA2
ERCC4	9	CDKN1B	CDKN1B	1	RB1
ERCC1	10	CDKN1B	CDKN1B	1	CCNB1

XRCC1	9	CDKN1B	CDKN1B	1	WEE1
CCND1	9	CDKN1B	CDKN1B	1	CDH1
E2F1	19	JUN	CDKN1B	1	RAD52
ABL1	5	CDK2	CDKN1B	1	XPA
RELA	1	NFKBIA	CDKN1B	1	RAD23B
EP300	6	JUN	CDKN1B	1	RPA1
EP300	7	TFDP1	MYC	1	CCNH
MYC	10	APC	MYC	1	RAD52
XRCC6	4	APC	MYC	1	CCNB1
APC	11	JUN	MYC	1	CDKN1B
XRCC5	3	RAD51	MYC	1	MAX
CCNA2	2	CDKN2D	MYC	1	CDH1
BRCA1	1	RELA	MYC	1	CCNE1
SFN	1	LIG3	MYC	1	CCNA2
RPA3	8	CDK2	MYC	1	E2F1
MAX	6	EP300	MYC	1	ERCC1
RFC5	7	XRCC6	MYC	1	CCND1
LIG3	2	RAD23B	MYC	1	XRCC1
CDK6	7	RAD51	GADD45A	1	APC
APC	10	ERCC1	GADD45A	1	NFKB1
CDH1	3	RAP1A	GADD45A	1	CDKN2D
EP300	5	BAX	GADD45A	1	RAP1A
XRCC5	4	ABL1	GADD45A	1	BAX
CDK4	1	CSNK2A2	GADD45A	1	MNAT1
CDKN2D	2	CCNB1	GADD45A	1	PLK1
RFC5	11	EP300	CCNE1	1	MNAT1
CDKN2D	6	CDH1	CCNE1	1	EP300
XRCC1	1	CDH1	CCNE1	1	RPA3
ERCC1	5	MAX	CCNE1	1	XRCC1
SKP2	3	CDK5	CCNE1	1	CCNB1
SKP2	3	CKS1B	PARP1	1	RB1
XRCC5	6	TFDP1	PARP1	1	CHEK1
CCNE1	12	RFC5	JUN	1	MARK3
MAPK8	6	CCNE1	CDK7	1	RAD52
SKP2	5	RAF1	CDK7	1	MYC
E2F4	4	MAPK9	CDK7	1	CDKN2D
CDK2	8	SFN	CDK7	1	CKS1B
E2F1	15	SKP1	CDK7	1	CCND1

CSNK2A1	6	RFC5	CDK7	1	RB1
SFN	5	GADD45A	RFC5	1	CSNK2A2
CCNE1	5	CDK6	CDK4	1	KAT2B
JUN	6	ERCC1	CDK4	1	TAF1
XRCC5	8	E2F4	RAF1	1	FEN1
EP300	1	XRCC1	RAF1	1	RAD52
ABL1	5	CDK4	RAF1	1	BAX
FOS	6	RFC5	RAF1	1	PARP1
E2F1	24	APC	SKP2	1	CCNE1
CDK2	14	APC	SKP2	1	ERCC4
E2F1	16	XRCC5	E2F1	1	CDH1
CDK2	7	BAX	CCNA2	1	MDM2
CDK2	21	E2F1	CCNA2	1	RPA1
EP300	5	ABL1	RPA1	1	RPA3
RFC5	13	MAX	RPA1	1	XRCC5
E2F1	13	WEE1	RPA1	1	RPA2
XRCC5	6	PARP1	RPA1	1	E2F1
CCNA2	6	APC	CKS1B	1	MAPK10
RAD51	5	RAF1	CKS1B	1	CSNK2A2
RAD23B	3	RB1	CKS1B	1	CSNK2A1
RPA2	14	E2F1	TFDP2	1	RB1
RFC5	6	PLK1	TFDP2	1	EP300
PCNA	9	CDKN1B	TFDP2	1	RAD23B
MAPK9	10	CDK2	POLR2G	1	TFDP1
MYC	12	RFC5	POLR2G	1	MAPK9
APC	6	KAT2B	RELA	1	SFN
PARP1	13	CDKN1B	XRCC5	1	BAX
CDK4	7	CDK6	RPA3	1	SKP1
POLR2G	2	RPA3	RPA3	1	PLK1
E2F1	14	PARP1	RPA3	1	CCNB1
GADD45A	2	MAPK8	RPA3	1	TAF1
MAPK10	6	CDK4	RPA3	1	ERCC1
APC	8	MDM2	RPA3	1	CRK
APC	11	EP300	RPA3	1	MAPK8
MYC	5	PARP1	RPA3	1	RPA2
RFC5	22	E2F1	EP300	1	ERCC1
POLR2G	9	APC	XRCC6	1	MNAT1
CDH1	3	PARP1	XRCC6	1	MDM2

WEE1	4	RB1	PCNA	1	BRCA1
CASP3	3	CCNE1	PCNA	1	NFKB1
EP300	7	CDK2	CDKN2A	1	CHEK1
CDK6	12	RFC5	CDKN2A	1	TP53
CDK2	6	TFDP2	POLR2D	1	CDK5
CDK2	7	RB1	ERCC4	1	TAF1
NFKBIA	3	PARP1	PLK1	1	XRCC1
SFN	8	JUN	PLK1	1	CDH1
E2F1	14	FOS	CDK7	1	HDAC1
XRCC5	3	RFC1	CDK6	1	HDAC1
CDK4	7	MYC	XRCC5	1	HDAC1
APC	1	RPA1	HDAC1	1	CDK1
JUN	9	XRCC5	CKS1B	1	CDK1

Table A-2 Genes and Pdb Files

Gene	**Pdb Files**
RELA	1nfiAC
PCNA	1axcACE
RAD23B	1pve 1p1a
CDK1	1lc9
RFC5	1lfs
CDKN1B	1jsuC
HMGB1	2yrq
NFKBIA	1nfiEF
ERCC4	1z00B
KAT2B	1zs5
BAX	1f16
HDAC1	1tyi
ABL1	2abl 1zzp
FEN1	1ul1XYZ
MAPK10	1jnk
CDK5	1unhAB
XRCC1	1cdz 2d8m 1xna
CRK	2eyz
RPA2	2z6kAB
XRCC5	1jeqB
MDM2	2vjeAC 2c6a

XRCC6	1jeqA
CCNA2	1vywBD
TP53	1gzhA 1ycsA
SKP1	2astA
GADD45A	2kg4
CCNE1	1w98B
JUN	1fosFH
CKS1B	1buhB
RB1	1gh6B 2qdj
CSNK2A2	3e3b
PLK1	1q4kABC 2ou7
LIG1	1x9n
RAF1	1c1yB
MARK3	2qnjAB
CDH1	2o72
NFKB1	1nfiBD
WEE1	1x8b
MAPK9	3e7oAB
XPA	1xpa
CHEK1	1nvqA
BRCA1	1t15A 1jm7A
CDKN2A	1a5e
CDK6	1blxA
CDKN2D	1blxB
FOS	1fosEG
RFC1	2ebu
GTF2H1	1pfj
MAPK8	1ukhA
PARP1	1wokABCD 2riq 2cr9 2cok
POLR2D	2c35ACEG
POLR2G	2c35BDFH
RAP1A	1c1yA
TP53BP2	1ycsB
SKP2	2astB
MNAT1	1g25
EP300	2k8fA 1p4qB 3biy
MYC	1nkpAD
CSNK2A1	1jwhAB

CCNB1	2b9rAB
E2F4	1cf7A
ERCC1	2a1i 1z00A
CASP3	1cp3AB
E2F1	2azeB
CCND1	2w96A
MAX	1nkpBE
RPA1	2b3gA 1fguAB 1l1oCF
RPA3	2z6kCD
HBP1	2e6o
SFN	1ywtAB
TAF1	1eqf
CDK2	1vywAC
CDK4	2w96B
RAD52	1h2i
RAD51	1n0wA 1b22
CDK7	1ua2ABCD
LIG3	1uw0 1imo
APC	1debAB 1m5i
CCNH	1jkw
TFDP2	1cf7B
TFDP1	2azeA

Table A-3 Betweenness Centrality Rasults on p53 PPIN

Node Betweenness of Unweighted p53 Network					
Shortest-Path		**Flow**		**Random Walk**	
APC	0,228487731	APC	0,1307871	APC	0.2709490529891378
CCNE1	0,133329397	CHEK1	0,0737605	CDKN1B	0.18924280795780773
CDKN1B	0,112881364	CDKN1B	0,0724935	CCNE1	0.17844539119457375
CDK2	0,109811812	CCNE1	0,0567327	CDK2	0.15839400473831144
CHEK1	0,072192696	RPA3	0,0532587	E2F1	0.14075031412249678
E2F1	0,069472154	CDKN2A	0,0493671	XRCC5	0.1388126268618467
PARP1	0,068608224	XRCC5	0,0443854	RPA3	0.12982533794110993
XRCC5	0,06750204	CKS1B	0,0397962	PARP1	0.12902421383261617
RFC5	0,061030623	CDK2	0,0389177	MYC	0.12863249622166056

RPA3	0,054635561	RAF1	0,0386999	RFC5	0.1252206960159592
CDKN2A	0,048148148	GADD45A	0,0375203	CDKN2D	0.12403692960092805
MYC	0,047971529	E2F1	0,0359129	GADD45A	0.10997658468142453
CDKN2D	0,045999957	CDK4	0,0336731	CHEK1	0.1088880936310862
GADD45A	0,045801936	RFC5	0,0327821	SKP2	0.10626625874283549
RAF1	0,035941671	CDK5	0,0258966	EP300	0.10510129926439911
SFN	0,032369354	SKP2	0,0255745	CDK4	0.10168422003080997
SKP2	0,031810955	TP53	0,025	RAF1	0.09987695208559336
CKS1B	0,031659247	MYC	0,0233595	CKS1B	0.09890120575057622
EP300	0,028894682	SFN	0,022956	SFN	0.08790689929994837
CDK4	0,026141526	PARP1	0,0218243	CDK6	0.08600454836556677
CDK5	0,024382716	CDKN2D	0,0180396	RB1	0.08410885759332347
TP53	0,024382716	PCNA	0,0175272	JUN	0.08304732744424387
CDK6	0,023524631	JUN	0,0166571	CDK7	0.07733838578473938
RB1	0,021446932	EP300	0,0160222	CDKN2A	0.0728395061728395
JUN	0,016642854	RAD23B	0,0146761	CCNA2	0.07088763358800103
CDK7	0,013187848	HDAC1	0,0137176	CDH1	0.07068137773561024
CDH1	0,013024345	RELA	0,0135624	POLR2G	0.0662941033114174
RPA1	0,012759682	CDK7	0,0132437	RAD51	0.06594869533810424
POLR2G	0,012425536	RB1	0,0130732	RPA1	0.06567840546900255
CCNA2	0,01153403	XRCC6	0,0125	CCNB1	0.06117668987480337
BRCA1	0,010891117	BRCA1	0,0123418	ERCC1	0.0595316400155366
RAD51	0,009863283	POLR2G	0,0103376	PLK1	0.059427042076819675
HDAC1	0,009185238	CDH1	0,009679	XRCC1	0.05819298090896352
RAD23B	0,009050453	CCNA2	0,0081379	TFDP2	0.05808267084686691
RAD52	0,008955628	CDK6	0,0078605	CDK5	0.05725002948958264
BAX	0,008857037	RPA1	0,0076171	BAX	0.05642750520319859
ERCC1	0,007714753	E2F4	0,0072253	HDAC1	0.05595812142393979
PLK1	0,007355281	TFDP2	0,0071466	RAD23B	0.05475580577112744
PCNA	0,006688166	TAF1	0,0070675	RAD52	0.054280613285613614
CCNB1	0,006190251	ERCC4	0,0070314	BRCA1	0.0538641535521129
TFDP2	0,006170352	NFKB1	0,0068565	PCNA	0.05367047071768541
NFKBIA	0,005113904	NFKBIA	0,0065354	XRCC6	0.05164822498306482
ERCC4	0,005039271	MAPK9	0,0052479	RELA	0.05103770040681539
XRCC1	0,0046572	LIG3	0,0045434	MAPK8	0.050193239906941
MAPK9	0,004645318	MDM2	0,0041139	E2F4	0.04933529128802585
MAPK10	0,00440962	CDK1	0,0041139	TP53	0.049074074074074076
TAF1	0,003712998	RAD51	0,0040676	MAX	0.04829342817008102

XRCC6	0,003298391	MNAT1	0,0039557	MAPK9	0.04822180577134028
MAPK8	0,002987195	PLK1	0,0030666	TAF1	0.04763541282712336
MNAT1	0,002492326	XRCC1	0,0027065	ERCC4	0.047362578506186026
E2F4	0,002358918	ERCC1	0,0026538	NFKBIA	0.04711564724663456
RELA	0,001902408	RAP1A	0,0025392	MNAT1	0.04604768638925154
RAP1A	0,001784404	RAD52	0,0024654	TFDP1	0.04433316757366516
SKP1	0,001631091	BAX	0,0021538	ABL1	0.04429987809025114
TFDP1	0,001423154	TFDP1	0,001988	RAP1A	0.043911419142901864
CSNK2A2	0,001411599	MAX	0,0016417	MAPK10	0.04359674564472339
NFKB1	0,001260288	RPA2	0,0016115	SKP1	0.04308711204744235
ABL1	0,001121495	CSNK2A2	0,0014768	CSNK2A2	0.04172501632063064
LIG3	0,001033046	MAPK10	0,0013186	MDM2	0.039288302891288124
KAT2B	0,000925375	CCND1	0,00129	NFKB1	0.03910328540554266
RPA2	0,000590112	MAPK8	0,0012371	RPA2	0.037915138374562875
CSNK2A1	0,000539101	SKP1	0,0011821	CCND1	0.037307028222432356
CCND1	0,00053669	CSNK2A1	9,49E-04	WEE1	0.03719677651165821
WEE1	0,000505001	ABL1	9,33E-04	LIG3	0.036702783577096595
GTF2H1	0,000421186	CCNB1	8,14E-04	CCNH	0.036520785017658985
MAX	0,000406661	WEE1	6,19E-04	FOS	0.03613978465300501
CDK1	0,000352734	MARK3	5,80E-04	CDK1	0.03459159120765722
MDM2	0,000298721	KAT2B	5,27E-04	CSNK2A1	0.034162734298206314
MARK3	0,000131918	CCNH	2,90E-04	KAT2B	0.034120540923887346
HBP1	0	FOS	2,33E-04	GTF2H1	0.03183798360449791
TP53BP2	0	HBP1	5,27E-05	HBP1	0.031358640249369525
POLR2D	0	GTF2H1	1,67E-05	MARK3	0.030695972737303123
XPA	0	TP53BP2	0	CRK	0.024691358024691357
CRK	0	POLR2D	0	TP53BP2	0.024691358024691357
FOS	0	XPA	0	FEN1	0.024691358024691357
RFC1	0	CRK	0	POLR2D	0.024691358024691357
LIG1	0	RFC1	0	HMGB1	0.024691358024691357
CASP3	0	LIG1	0	RFC1	0.024691358024691357
HMGB1	0	CASP3	0	XPA	0.024691358024691357
CCNH	0	HMGB1	0	LIG1	0.024691358024691357
FEN1	0	FEN1	0	CASP3	0.024691358024691357

Table A-4 Betweenness Centrality Rasults on p53 SPPIN

Node Betweenness of Weighted p53 Network					
Shortest-Path		**Flow**		**Random Walk**	
APC	0,3364412	APC	0,140651	APC	0.3089864506355171
RPA3	0,1394766	CHEK1	0,0739381	RPA3	0.18314570438311695
RPA1	0,1251393	RPA3	0,0714404	CCNE1	0.18162897597629435
CDKN1B	0,1203443	PARP1	0,0708059	MYC	0.17636714622085156
CCNE1	0,1194616	CCNE1	0,0577669	CDKN1B	0.16999421034689813
MYC	0,1176936	CDKN1B	0,0567469	PARP1	0.1542223111368365
RB1	0,1103379	XRCC5	0,0509684	GADD45A	0.15320895921605712
PARP1	0,1014587	GADD45A	0,0506653	XRCC5	0.13337254936160972
GADD45A	0,1003236	SKP2	0,0504187	SKP2	0.12862624009629184
CDK7	0,092481	MYC	0,0496015	CDK2	0.12652540030781836
SKP2	0,086667	CDKN2A	0,0493671	CDK7	0.12059937110891275
BRCA1	0,0830941	CKS1B	0,0470336	CKS1B	0.12045372501107059
XRCC5	0,0789575	CDK2	0,038572	RPA1	0.11278828548896348
CHEK1	0,0712963	RAF1	0,0377122	RB1	0.10752909028993692
CDK2	0,059476	CDK4	0,0364152	CHEK1	0.10551668992226874
CKS1B	0,054225	RELA	0,0323536	RAF1	0.10516915911193937
RAD52	0,053482	POLR2G	0,0320641	CDK4	0.09624382633140238
CDKN2A	0,0481481	SFN	0,0297015	EP300	0.09546457598707797
PLK1	0,047541	BRCA1	0,0265314	CDKN2D	0.08557695596442227
RELA	0,0461655	CDK5	0,0252755	CCNA2	0.08530062359100272
RAF1	0,0427985	TP53	0,025	PLK1	0.0849501539559862
CCNB1	0,0425719	HDAC1	0,0202823	RFC5	0.0839260911335062
HDAC1	0,0416145	EP300	0,0174934	BRCA1	0.08347353766343933
XRCC1	0,0367802	CCNA2	0,0170779	RAD52	0.08342313392058351
TAF1	0,0364679	RFC5	0,0162123	CCNB1	0.08238470781566654
TFDP1	0,0337122	CDK7	0,0152171	CDH1	0.08032591081150194
POLR2G	0,0323325	RAD23B	0,0148501	HDAC1	0.07991437542565831
CDK4	0,0309965	JUN	0,0145131	SFN	0.07978034187704096
MNAT1	0,0288698	RPA1	0,0134886	XRCC1	0.07912209074586633
BAX	0,0282689	XRCC6	0,0134283	RELA	0.076718159385166
RFC5	0,0277016	PCNA	0,0133768	POLR2G	0.07605541884407428
CCNA2	0,0275372	TFDP2	0,0132409	BAX	0.07535675530841222
CSNK2A2	0,0254821	TAF1	0,0132392	TAF1	0.07390666221489474
CDK5	0,0243827	NFKB1	0,0119559	CDKN2A	0.0728395061728395

TP53	0,0243827	MNAT1	0,0118497	TFDP2	0.06977282369419252
HBP1	0,0242964	MAPK9	0,009867	CSNK2A2	0.06777729477143683
TFDP2	0,0233609	MDM2	0,0098148	E2F1	0.0668977624454145
MARK3	0,0231481	RB1	0,0097319	RAD23B	0.06660067584508579
RAD23B	0,0231006	MARK3	0,0096141	MNAT1	0.06479042785329145
MAPK10	0,0191133	CDH1	0,0086072	ERCC1	0.06257161294010967
EP300	0,0161225	CDKN2D	0,0084844	JUN	0.060849566345041205
CDH1	0,0126306	TFDP1	0,0083338	TFDP1	0.0592087569410028
ERCC1	0,00886	E2F1	0,0077279	CDK6	0.055823533749903544
CDKN2D	0,0076351	CSNK2A2	0,0076176	MAPK10	0.05372427907537132
SFN	0,0074618	LIG3	0,007181	PCNA	0.052916050788853516
ERCC4	0,0073349	CDK6	0,005928	CDK5	0.05285607446204238
NFKB1	0,0056639	E2F4	0,0056141	XRCC6	0.05225878108513521
MDM2	0,0043158	ERCC1	0,0054241	NFKB1	0.05081776323099193
MAPK9	0,0039914	ERCC4	0,0052847	ERCC4	0.050338553169124514
LIG3	0,0039352	RAD51	0,0051543	RAD51	0.050078453759129844
CDK1	0,0033333	BAX	0,0043006	TP53	0.049074074074074076
XRCC6	0,0025669	CDK1	0,0032428	NFKBIA	0.04774552874087249
E2F1	0,002308	NFKBIA	0,0027838	LIG3	0.04578774813429356
NFKBIA	0,0015432	RAD52	0,0027569	MDM2	0.045345529678303166
PCNA	0,0015432	PLK1	0,0027221	CDK1	0.04443246205873119
RAD51	0	XRCC1	0,0026995	MAPK8	0.04409390341767557
CDK6	0	MAPK10	0,0017531	MARK3	0.04376129265735431
WEE1	0	CCNB1	0,0013826	HBP1	0.042620964976958796
GTF2H1	0	ABL1	0,0012512	MAPK9	0.042352765768443795
RPA2	0	MAX	8,81E-04	RPA2	0.039999426373873055
JUN	0	SKP1	8,61E-04	SKP1	0.0397503371848626
E2F4	0	RAP1A	7,62E-04	ABL1	0.03891354519149272
TP53BP2	0	HBP1	7,30E-04	CCND1	0.0389091642325172
ABL1	0	CSNK2A1	6,03E-04	E2F4	0.038825451397583505
POLR2D	0	RPA2	5,97E-04	RAP1A	0.03873163524662314
RAP1A	0	FOS	5,58E-04	MAX	0.03707310525491497
MAPK8	0	KAT2B	5,37E-04	CCNH	0.03307202786782178
XPA	0	MAPK8	5,29E-04	KAT2B	0.03211128091531921
MAX	0	CCND1	3,54E-04	FOS	0.030970609648053597
CRK	0	WEE1	3,05E-04	WEE1	0.030059061998040277
KAT2B	0	CCNH	2,38E-04	CSNK2A1	0.029831055238721944
FOS	0	GTF2H1	6,91E-06	GTF2H1	0.027827854011384108

CSNK2A1	0	TP53BP2	0	CRK	0.024691358024691357
RFC1	0	POLR2D	0	TP53BP2	0.024691358024691357
LIG1	0	XPA	0	FEN1	0.024691358024691357
SKP1	0	CRK	0	POLR2D	0.024691358024691357
CASP3	0	RFC1	0	HMGB1	0.024691358024691357
CCND1	0	LIG1	0	RFC1	0.024691358024691357
HMGB1	0	CASP3	0	XPA	0.024691358024691357
CCNH	0	HMGB1	0	LIG1	0.024691358024691357
FEN1	0	FEN1	0	CASP3	0.024691358024691357

Table A-5 GO Biological Process Analysis on p53 PPIN

GO Biological Process - Shortest Path Betweeenness - Unweighted			
GO-ID	**p-value**	**Description**	**Genes in test set**
79	1.7677E-13	regulation of cyclin-dependent protein kinase activity	CDKN2D|CDKN1B|APC|CDKN2A|GADD45A|CHEK1|SFN
51726	5.3246E-12	regulation of cell cycle	CDKN2D|CDKN1B|APC|CDKN2A|GADD45A|MYC|CDK2|CHEK1|E2F1|SFN
51329	8.2584E-12	interphase of mitotic cell cycle	CDKN2D|CDKN1B|CDKN2A|GADD45A|CCNE1|CDK2|E2F1
51325	1.0874E-11	interphase	CDKN2D|CDKN1B|CDKN2A|GADD45A|CCNE1|CDK2|E2F1
22402	7.4009E-11	cell cycle process	CDKN2D|CDKN1B|APC|CDKN2A|GADD45A|CCNE1|MYC|CDK2|CHEK1|E2F1
6974	8.4129E-11	response to DNA damage stimulus	RFC5|CDKN2D|PARP1|APC|XRCC5|GADD45A|RPA3|CHEK1|SFN
22403	1.9410E-10	cell cycle phase	CDKN2D|CDKN1B|APC|CDKN2A|GADD45A|CCNE1|CDK2|CHEK1|E2F1
51716	5.0685E-10	cellular response to stimulus	RFC5|CDKN2D|PARP1|APC|XRCC5|GADD45A|CCNE1|RPA3|CHEK1|E2F1|SFN
6259	8.9675E-10	DNA metabolic process	RFC5|CDKN2D|PARP1|XRCC5|CDKN2A|GADD45A|RPA3|CDK2|CHEK1
7050	1.2128E-9	cell cycle arrest	CDKN2D|CDKN1B|APC|CDKN2A|GADD45A|MYC
GO Biological Process - Flow Betweeenness - Unweighted			
GO-ID	**p-value**	**Description**	**Genes in test set**
51329	6.2046E-14	interphase of mitotic cell cycle	CDKN1B|CDKN2A|GADD45A|CCNE1|CDK4|CDK2|E2F1|SKP2
51325	8.5156E-14	interphase	CDKN1B|CDKN2A|GADD45A|CCNE1|CDK4|CDK2|E2F1|SKP2
22403	4.1522E-12	cell cycle phase	CDKN1B|APC|CDKN2A|GADD45A|CCNE1|CDK4|CHEK1|CDK2|E2F1|SKP2

51726	5.3209E-12	regulation of cell cycle	CDKN1B|APC|CDK5|CDKN2A|GADD45A|CDK4|CHEK1|CDK2|E2F1|CKS1B
79	3.2646E-11	regulation of cyclin-dependent protein kinase activity	CDKN1B|APC|CDKN2A|GADD45A|CHEK1|CKS1B
7049	4.8708E-11	cell cycle	CDKN1B|APC|CDKN2A|GADD45A|CCNE1|CDK4|CHEK1|CDK2|E2F1|SKP2|CKS1B
278	5.8200E-11	mitotic cell cycle	CDKN1B|APC|CDKN2A|GADD45A|CCNE1|CDK4|CDK2|E2F1|SKP2
22402	7.3958E-11	cell cycle process	CDKN1B|APC|CDKN2A|GADD45A|CCNE1|CDK4|CHEK1|CDK2|E2F1|SKP2
82	2.2081E-9	G1/S transition of mitotic cell cycle	CDKN1B|CDKN2A|CCNE1|CDK4|SKP2
51301	2.2174E-8	cell division	APC|CDK5|CDKN2A|CCNE1|CDK4|CDK2|CKS1B
GO Biological Process - Random Walk Betweeenness - Unweighted			
GO-ID	**p-value**	**Description**	**Genes in test set**
51329	6.2046E-14	interphase of mitotic cell cycle	CDKN2D|CDKN1B|GADD45A|CCNE1|CDK4|CDK2|E2F1|SKP2
51325	8.5156E-14	interphase	CDKN2D|CDKN1B|GADD45A|CCNE1|CDK4|CDK2|E2F1|SKP2
7049	1.1690E-12	cell cycle	CDKN2D|CDKN1B|APC|GADD45A|CCNE1|CDK4|MYC|CDK2|CHEK1|E2F1|EP300|SKP2
22402	1.6729E-12	cell cycle process	CDKN2D|CDKN1B|APC|GADD45A|CCNE1|CDK4|MYC|CDK2|CHEK1|E2F1|SKP2
22403	4.1522E-12	cell cycle phase	CDKN2D|CDKN1B|APC|GADD45A|CCNE1|CDK4|CDK2|CHEK1|E2F1|SKP2
278	5.8200E-11	mitotic cell cycle	CDKN2D|CDKN1B|APC|GADD45A|CCNE1|CDK4|CDK2|E2F1|SKP2
51726	2.4217E-10	regulation of cell cycle	CDKN2D|CDKN1B|APC|GADD45A|CDK4|MYC|CDK2|CHEK1|E2F1
51716	5.0647E-10	cellular response to stimulus	RFC5|CDKN2D|PARP1|APC|XRCC5|GADD45A|CCNE1|RPA3|CHEK1|E2F1|EP300

82	2.2081E-9	G1/S transition of mitotic cell cycle	CDKN2D|CDKN1B|CCNE1|CDK4|SKP2
6974	3.4074E-9	response to DNA damage stimulus	RFC5|CDKN2D|PARP1|APC|XRCC5|GADD45A|RPA3|CHEK1

Table A-6 GO Biological Process Analysis on p53 SPPIN

GO Biological Process - Shortest Path Betweeenness - Weighted			
GO-ID	**p-value**	**Description**	**Genes in test set**
7049	1.1699E-12	cell cycle	RB1|CDKN1B|APC|GADD45A|CCNE1|MYC|CHEK1|CDK2|RPA1|BRCA1|SKP2|CKS1B
51726	5.3246E-12	regulation of cell cycle	RB1|CDK7|CDKN1B|APC|GADD45A|MYC|CHEK1|CDK2|BRCA1|CKS1B
79	3.2660E-11	regulation of cyclin-dependent protein kinase activity	CDK7|CDKN1B|APC|GADD45A|CHEK1|CKS1B
22402	7.4009E-11	cell cycle process	RB1|CDKN1B|APC|GADD45A|CCNE1|MYC|CHEK1|CDK2|RPA1|SKP2
6974	8.4129E-11	response to DNA damage stimulus	CDK7|PARP1|APC|GADD45A|XRCC5|RPA3|CHEK1|RPA1|BRCA1

22403	1.9410E-10	cell cycle phase	RB1|CDKN1B|APC|GADD45A|CCNE1|CHEK1|CDK2|RPA1|SKP2
6281	3.7953E-10	DNA repair	CDK7|PARP1|GADD45A|XRCC5|RPA3|CHEK1|RPA1|BRCA1
51329	8.5650E-10	interphase of mitotic cell cycle	RB1|CDKN1B|GADD45A|CCNE1|CDK2|SKP2
6259	8.9675E-10	DNA metabolic process	CDK7|PARP1|GADD45A|XRCC5|RPA3|CHEK1|CDK2|RPA1|BRCA1
51325	1.0825E-9	interphase	RB1|CDKN1B|GADD45A|CCNE1|CDK2|SKP2
GO Biological Process - Flow Betweeenness - Weighted			
GO-ID	p-value	Description	Genes in test set
51329	8.2544E-12	interphase of mitotic cell cycle	CDKN1B|GADD45A|CDKN2A|CCNE1|CDK4|CDK2|SKP2
51325	1.0869E-11	interphase	CDKN1B|GADD45A|CDKN2A|CCNE1|CDK4|CDK2|SKP2

79	3.2646E-11	regulation of cyclin-dependent protein kinase activity	CDKN1B|APC|GADD45A|CDKN2A|CHEK1|CKS1B
7049	4.8708E-11	cell cycle	CDKN1B|APC|GADD45A|CDKN2A|CCNE1|CDK4|MYC|CHEK1|CDK2|SKP2|CKS1B
22402	7.3958E-11	cell cycle process	CDKN1B|APC|GADD45A|CDKN2A|CCNE1|CDK4|MYC|CHEK1|CDK2|SKP2
22403	1.9398E-10	cell cycle phase	CDKN1B|APC|GADD45A|CDKN2A|CCNE1|CDK4|CHEK1|CDK2|SKP2
51726	2.4217E-10	regulation of cell cycle	CDKN1B|APC|GADD45A|CDKN2A|CDK4|MYC|CHEK1|CDK2|CKS1B
82	2.2081E-9	G1/S transition of mitotic cell cycle	CDKN1B|CDKN2A|CCNE1|CDK4|SKP2
278	2.4623E-9	mitotic cell cycle	CDKN1B|APC|GADD45A|CDKN2A|CCNE1|CDK4|CDK2|SKP2
42127	6.7996E-8	regulation of cell proliferation	CDKN1B|APC|CDKN2A|CDK4|MYC|CHEK1|CDK2|RAF1|RELA
GO Biological Process - Random Walk Betweeenness - Weighted			

GO-ID	p-value	Description	Genes in test set
79	3.2660E-11	regulation of cyclin-dependent protein kinase activity	CDK7|CDKN1B|APC|GADD45A|CHEK1|CKS1B
7049	4.8744E-11	cell cycle	RB1|CDKN1B|APC|GADD45A|CCNE1|MYC|CDK2|CHEK1|RPA1|SKP2|CKS1B
22402	7.4009E-11	cell cycle process	RB1|CDKN1B|APC|GADD45A|CCNE1|MYC|CDK2|CHEK1|RPA1|SKP2
22403	1.9410E-10	cell cycle phase	RB1|CDKN1B|APC|GADD45A|CCNE1|CDK2|CHEK1|RPA1|SKP2
51726	2.4231E-10	regulation of cell cycle	RB1|CDK7|CDKN1B|APC|GADD45A|MYC|CDK2|CHEK1|CKS1B
51329	8.5650E-10	interphase of mitotic cell cycle	RB1|CDKN1B|GADD45A|CCNE1|CDK2|SKP2
51325	1.0825E-9	interphase	RB1|CDKN1B|GADD45A|CCNE1|CDK2|SKP2
6974	3.4092E-9	response to DNA damage stimulus	CDK7|PARP1|APC|GADD45A|XRCC5|RPA3|CHEK1|RPA1

6281	1.6197E-8	DNA repair	CDK7|PARP1|GADD45A|XRCC5|RPA3|CHEK1|RPA1
6259	2.7508E-8	DNA metabolic process	CDK7|PARP1|GADD45A|XRCC5|RPA3|CDK2|CHEK1|RPA1

Table A-7 GO Molecular Function Analysis on p53 PPIN

GO Molecular Function - Shortest Path Betweeenness - Unweighted			
GO-ID	**p-value**	**Description**	**Genes in test set**
30291	2.3312E-9	protein serine/threonine kinase inhibitor activity	CDKN2D|CDKN1B|CDKN2A|SFN
19887	2.1021E-8	protein kinase regulator activity	CDKN2D|CDKN1B|APC|CDKN2A|SFN
19207	3.8383E-8	kinase regulator activity	CDKN2D|CDKN1B|APC|CDKN2A|SFN
4860	4.9723E-8	protein kinase inhibitor activity	CDKN2D|CDKN1B|CDKN2A|SFN
19210	6.1847E-8	kinase inhibitor activity	CDKN2D|CDKN1B|CDKN2A|SFN
4861	1.4983E-7	cyclin-dependent protein kinase inhibitor activity	CDKN2D|CDKN1B|CDKN2A
16538	7.3771E-7	cyclin-dependent protein kinase regulator activity	CDKN2D|CDKN1B|CDKN2A
5515	3.4065E-5	protein binding	RFC5|CDKN2D|CDKN1B|PARP1|XRCC5|CDKN2A|GADD45A|APC|CCNE1|RPA3|MYC|CDK2|CHEK1|E2F1|SFN|RAF1
4857	1.5992E-4	enzyme inhibitor activity	CDKN2D|CDKN1B|CDKN2A|SFN
19901	7.1874E-4	protein kinase binding	CDKN2D|APC|CDKN2A
GO Molecular Function - Flow Betweeenness - Unweighted			
GO-ID	**p-value**	**Description**	**Genes in test set**
16538	7.3757E-7	cyclin-dependent protein kinase regulator activity	CDKN1B|CDKN2A|CKS1B
19887	1.6240E-6	protein kinase regulator activity	CDKN1B|APC|CDKN2A|CKS1B

19207	2.6175E-6	kinase regulator activity	CDKN1B|APC|CDKN2A|CKS1B
4693	3.2799E-6	cyclin-dependent protein kinase activity	CDK5|CDK4|CDK2
5515	3.4030E-5	protein binding	RFC5|CDKN1B|CDKN2A|XRCC5|GADD45A|CKS1B|APC|CDK5|CCNE1|CDK4|RPA3|CHEK1|CDK2|E2F1|RAF1|SKP2
4674	4.7320E-5	protein serine/threonine kinase activity	CDK5|CDK4|CHEK1|CDK2|RAF1
4861	5.5052E-5	cyclin-dependent protein kinase inhibitor activity	CDKN1B|CDKN2A
30291	1.5250E-4	protein serine/threonine kinase inhibitor activity	CDKN1B|CDKN2A
4672	2.0615E-4	protein kinase activity	CDK5|CDK4|CHEK1|CDK2|RAF1
2039	2.5141E-4	p53 binding	CDK5|CDKN2A
GO Molecular Function - Random Walk Betweeenness - Unweighted			
GO-ID	**p-value**	**Description**	**Genes in test set**
5515	3.4030E-5	protein binding	RFC5|CDKN2D|CDKN1B|PARP1|XRCC5|GADD45A|APC|CCNE1|CDK4|RPA3|MYC|CDK2|CHEK1|E2F1|EP300|SKP2
4861	5.5052E-5	cyclin-dependent protein kinase inhibitor activity	CDKN2D|CDKN1B
19887	9.1719E-5	protein kinase regulator activity	CDKN2D|CDKN1B|APC
19207	1.3070E-4	kinase regulator activity	CDKN2D|CDKN1B|APC
30291	1.5250E-4	protein serine/threonine kinase inhibitor activity	CDKN2D|CDKN1B
16538	1.5250E-4	cyclin-dependent protein kinase regulator activity	CDKN2D|CDKN1B
10843	1.9930E-4	promoter binding	XRCC5|MYC|EP300
44212	2.2087E-4	DNA regulatory region binding	XRCC5|MYC|EP300
4693	4.0199E-4	cyclin-dependent protein kinase activity	CDK4|CDK2
4860	6.5624E-4	protein kinase inhibitor activity	CDKN2D|CDKN1B

Table A-8 GO Molecular Function on p53 SPPIN

GO Molecular Function - Shortest Path Betweeenness - Weighted			
GO-ID	**p-value**	**Description**	**Genes in test set**
50681	2.0722E-8	androgen receptor binding	RB1|CDK7|CCNE1|BRCA1
35258	1.1162E-7	steroid hormone receptor binding	RB1|CDK7|CCNE1|BRCA1
35257	1.1020E-6	nuclear hormone receptor binding	RB1|CDK7|CCNE1|BRCA1
51427	1.8612E-6	hormone receptor binding	RB1|CDK7|CCNE1|BRCA1
5515	3.3998E-5	protein binding	RB1|CDKN1B|PARP1|GADD45A|XRCC5|RPA1|BRCA1|CKS1B|CDK7|APC|CCNE1|RPA3|MYC|CHEK1|CDK2|SKP2
16563	5.4816E-5	transcription activator activity	RB1|CDK7|CCNE1|MYC|BRCA1
3713	5.8321E-5	transcription coactivator activity	RB1|CDK7|CCNE1|BRCA1
19887	9.1737E-5	protein kinase regulator activity	CDKN1B|APC|CKS1B
19207	1.3072E-4	kinase regulator activity	CDKN1B|APC|CKS1B
8134	1.3879E-4	transcription factor binding	RB1|CDK7|PARP1|CCNE1|BRCA1
GO Molecular Function - Flow Betweeenness - Weighted			
GO-ID	**p-value**	**Description**	**Genes in test set**
16538	7.3757E-7	cyclin-dependent protein kinase regulator activity	CDKN1B|CDKN2A|CKS1B
19887	1.6240E-6	protein kinase regulator activity	CDKN1B|APC|CDKN2A|CKS1B
19207	2.6175E-6	kinase regulator activity	CDKN1B|APC|CDKN2A|CKS1B
5515	3.4030E-5	protein binding	CDKN1B|PARP1|XRCC5|GADD45A|CDKN2A|RELA|CKS1B|APC|CCNE1|CDK4|RP

			A3|MYC|CHEK1|CDK2|SKP2|RAF1
4861	5.5052E-5	cyclin-dependent protein kinase inhibitor activity	CDKN1B|CDKN2A
30291	1.5250E-4	protein serine/threonine kinase inhibitor activity	CDKN1B|CDKN2A
51059	1.5250E-4	NF-kappaB binding	CDKN2A|RELA
10843	1.9930E-4	promoter binding	XRCC5|MYC|RELA
44212	2.2087E-4	DNA regulatory region binding	XRCC5|MYC|RELA
4693	4.0199E-4	cyclin-dependent protein kinase activity	CDK4|CDK2
GO Molecular Function - Random Walk Betweeenness - Weighted			
GO-ID	**p-value**	**Description**	**Genes in test set**
50681	3.6428E-6	androgen receptor binding	RB1|CDK7|CCNE1
35258	1.2612E-5	steroid hormone receptor binding	RB1|CDK7|CCNE1
5515	3.3998E-5	protein binding	RB1|CDKN1B|PARP1|GADD45A|XRCC5|RPA1|CKS1B|CDK7|APC|CCNE1|RPA3|MYC|CDK2|CHEK1|SKP2|RAF1
35257	6.8789E-5	nuclear hormone receptor binding	RB1|CDK7|CCNE1
19887	9.1737E-5	protein kinase regulator activity	CDKN1B|APC|CKS1B
51427	1.0148E-4	hormone receptor binding	RB1|CDK7|CCNE1
19207	1.3072E-4	kinase regulator activity	CDKN1B|APC|CKS1B
16538	1.5252E-4	cyclin-dependent protein kinase regulator activity	CDKN1B|CKS1B
4693	4.0204E-4	cyclin-dependent protein kinase activity	CDK7|CDK2
43566	4.0686E-4	structure-specific DNA binding	XRCC5|RPA3|RPA1

Table A-9 Percentage of mutations on interface structures of genes

Gene	% of Mutations on Interfaces
CDKN1B	61.76470588235294
TFDP1	39.285714285714285
XRCC1	38.88888888888889
MAX	35.13513513513514
APC	34.5679012345679
RAP1A	33.33333333333333
CDK4	31.958762886597935
CDK5	31.03448275862069
BRCA1	30.76923076923077
RAD23B	28.57142857142857
MAPK10	27.27272727272727
CSNK2A2	26.31578947368421
CHEK1	26.31578947368421
GTF2H1	26.08695652173913
MAPK8	23.52941176470588
RELA	23.076923076923077
CDK6	22.47191011235955
SKP2	22.36842105263158
EP300	21.50537634408602
WEE1	21.428571428571427
E2F1	19.11764705882353
MYC	19.047619047619047
TFDP2	17.391304347826086
ERCC1	15.873015873015872
TP53BP2	15.789473684210526
CSNK2A1	15.789473684210526
CCNB1	15.625
ABL1	15.492957746478872
BAX	15.384615384615385
CCND1	15.384615384615385
CDKN2D	15.151515151515152
RPA3	14.705882352941178
PLK1	14.285714285714285
RAD51	14.102564102564102

PARP1	14.018691588785046
POLR2G	13.88888888888889
RAF1	13.793103448275861
CDK2	13.142857142857142
FOS	12.82051282051282
CASP3	12.5
XRCC6	12.121212121212121
CKS1B	12.121212121212121
SKP1	11.627906976744185
JUN	11.428571428571429
MARK3	11.11111111111111
CCNE1	10.569105691056912
RPA2	10.0
CCNH	9.523809523809524
NFKBIA	9.195402298850574
XRCC5	8.653846153846153
RFC1	6.25
SFN	4.705882352941177
GADD45A	4.166666666666666
LIG3	2.857142857142857
MAPK9	2.7027027027027026
MDM2	0.0
ERCC4	0.0
RPA1	0.0
E2F4	0.0
RFC5	0.0

BIBLIOGRAPHY

1. Albert, R., H. Jeong, and A.L. Barabasi, *Error and attack tolerance of complex networks.* Nature, 2000. **406**(6794): p. 378-382.
2. Barabasi, A.L. and R. Albert, *Emergence of scaling in random networks.* Science, 1999. **286**(5439): p. 509-512.
3. Valente, A.X. and M.E. Cusick, *Yeast Protein Interactome topology provides framework for coordinated-functionality.* Nucleic Acids Res, 2006. **34**(9): p. 2812-9.
4. Kim, P.M., et al., *Relating three-dimensional structures to protein networks provides evolutionary insights.* Science, 2006. **314**(5807): p. 1938-41.
5. Tuncbag, N., et al., *Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM.* Nature Protocols, 2011. **6**(9): p. 1341-1354.
6. Tuncbag, N., et al., *Towards inferring time dimensionality in protein-protein interaction networks by integrating structures: the p53 example.* Molecular Biosystems, 2009. **5**(12): p. 1770-1778.
7. Demircioğlu, D., Ö. Keskin, and A. Gursoy, *Modeling Structural Protein Interaction Networks for Betweenness Analysis*, in *Information Sciences and Systems 2014*, T. Czachórski, E. Gelenbe, and R. Lent, Editors. 2014, Springer International Publishing. p. 367-376.
8. Borgatti, S.P., *Centrality and network flow.* Social Networks, 2005. **27**(1): p. 55-71.
9. Borgatti, S.P. and M.G. Everett, *A graph-theoretic perspective on centrality.* Social Networks, 2006. **28**(4): p. 466-484.
10. Linton, C.F., *A Set of Measures of Centrality Based on Betweenness.* Sociometry, 1977. **40**(1): p. 35-41.
11. Pitts, F.R., *The medieval river trade network of Russia revisited.* Social Networks, 1978. **1**(3): p. 285-292.
12. Freeman, L.C., *Centrality in social networks conceptual clarification.* Social Networks, 1978. **1**(3): p. 215-239.
13. Brandes, U., *On variants of shortest-path betweenness centrality and their generic computation.* Social Networks, 2008. **30**(2): p. 136-145.
14. Brandes, U., *A faster algorithm for betweenness centrality.* Journal of Mathematical Sociology, 2001. **25**(2): p. 163-177.
15. Newman, M.E.J., *A measure of betweenness centrality based on random walks.* Social Networks, 2005. **27**(1): p. 39-54.
16. Freeman, L.C., S.P. Borgatti, and D.R. White, *Centrality in Valued Graphs - a Measure of Betweenness Based on Network Flow.* Social Networks, 1991. **13**(2): p. 141-154.
17. Zhang, Q., M. Sanner, and A.J. Olson, *Shape complementarity of protein-protein complexes at multiple resolutions.* Proteins-Structure Function and Bioinformatics, 2009. **75**(2): p. 453-467.
18. Grunberg, R., J. Leckner, and M. Nilges, *Complementarity of structure ensembles in protein-protein binding.* Structure, 2004. **12**(12): p. 2125-2136.
19. Cukuroglu, E., A. Gursoy, and O. Keskin, *HotRegion: a database of predicted hot spot clusters.* Nucleic Acids Res, 2012. **40**(Database issue): p. D829-33.

20. Tsai, C.J., et al., *Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect.* Protein Science, 1997. **6**(1): p. 53-64.
21. Keskin, O., N. Tuncbag, and A. Gursoy, *Characterization and prediction of protein interfaces to infer protein-protein interaction networks.* Current Pharmaceutical Biotechnology, 2008. **9**(2): p. 67-76.
22. Yan, C., et al., *Characterization of protein-protein interfaces.* Protein Journal, 2008. **27**(1): p. 59-70.
23. Tuncbag, N., O. Keskin, and A. Gursoy, *HotPoint: hot spot prediction server for protein interfaces.* Nucleic Acids Research, 2010. **38**: p. W402-W406.
24. Chen, J., et al., *Protein-Protein Interface Analysis and Hot Spots Identification for Chemical Ligand Design.* Current Pharmaceutical Design, 2014. **20**(8): p. 1192-1200.
25. Gavin, A.C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Faseb Journal, 2002. **16**(4): p. A523-A523.
26. Ho, Y., et al., *Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry.* Nature, 2002. **415**(6868): p. 180-183.
27. Rual, J.F., et al., *Towards a proteome-scale map of the human protein-protein interaction network.* Nature, 2005. **437**(7062): p. 1173-1178.
28. Stelzl, U., et al., *A human protein-protein interaction network: A resource for annotating the proteome.* Cell, 2005. **122**(6): p. 957-968.
29. Nikitin, A., et al., *Pathway studio - the analysis and navigation of molecular networks.* Bioinformatics, 2003. **19**(16): p. 2155-2157.
30. Barabasi, A.L., N. Gulbahce, and J. Loscalzo, *Network medicine: a network-based approach to human disease.* Nature Reviews Genetics, 2011. **12**(1): p. 56-68.
31. Yu, H.Y., et al., *The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics.* Plos Computational Biology, 2007. **3**(4): p. 713-720.
32. Joy, M.P., et al., *High-betweenness proteins in the yeast protein interaction network.* J Biomed Biotechnol, 2005. **2005**(2): p. 96-103.
33. Yang, J.S., et al., *SAPIN: a framework for the structural analysis of protein interaction networks.* Bioinformatics, 2012. **28**(22): p. 2998-9.
34. Kar, G., A. Gursoy, and O. Keskin, *Human Cancer Protein-Protein Interaction Network: A Structural Perspective.* Plos Computational Biology, 2009. **5**(12).
35. Shannon, P., et al., *Cytoscape: A software environment for integrated models of biomolecular interaction networks.* Genome Research, 2003. **13**(11): p. 2498-2504.
36. Pettersen, E.F., et al., *UCSF chimera - A visualization system for exploratory research and analysis.* Journal of Computational Chemistry, 2004. **25**(13): p. 1605-1612.
37. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.
38. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.
39. Goh, K.I., et al., *Betweenness centrality correlation in social networks.* Physical Review E, 2003. **67**(1).
40. Nakao, K., *Distribution of Measures of Centrality: Enumerated Distributions of Freeman's Graph Centrality Measures.* Connections, 1990. **13**(3): p. 10-22.

41. Maere, S., K. Heymans, and M. Kuiper, *BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks.* Bioinformatics, 2005. **21**(16): p. 3448-3449.
42. Murray, A.W., *Recycling the cell cycle: Cyclins revisited.* Cell, 2004. **116**(2): p. 221-234.
43. Zhang, H., et al., *P19(Skp1) and P45(Skp2) Are Essential Elements of the Cyclin a-Cdk2 S-Phase Kinase.* Cell, 1995. **82**(6): p. 915-925.
44. Carrano, A.C., et al., *SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27.* Nature Cell Biology, 1999. **1**(4): p. 193-199.
45. Tsvetkov, L.M., et al., *P27(Kip1) ubiquitination and degradation is regulated by the SCFSkp2 complex through phosphorylated Thr187 in p27.* Current Biology, 1999. **9**(12): p. 661-664.
46. Polyak, K., et al., *Cloning of P27(Kip1), a Cyclin-Dependent Kinase Inhibitor and a Potential Mediator of Extracellular Antimitogenic Signals.* Cell, 1994. **78**(1): p. 59-66.
47. Kanehisa, M. and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27-30.
48. Murphree, A.L. and W.F. Benedict, *Retinoblastoma - Clues to Human Oncogenesis.* Science, 1984. **223**(4640): p. 1028-1033.
49. Korenjak, M. and A. Brehm, *E2F-Rb complexes regulating transcription of genes important for differentiation and development.* Current Opinion in Genetics & Development, 2005. **15**(5): p. 520-527.
50. Munger, K. and P.M. Howley, *Human papillomavirus immortalization and transformation functions.* Virus Research, 2002. **89**(2): p. 213-228.
51. Bartkova, J., et al., *Cell-cycle regulatory proteins in human wound healing.* Archives of Oral Biology, 2003. **48**(2): p. 125-132.
52. Hass, C.S., L. Gakhar, and M.S. Wold, *Functional Characterization of a Cancer Causing Mutation in Human Replication Protein A.* Molecular Cancer Research, 2010. **8**(7): p. 1017-1026.

VITA

Cemal Yamak was born in Zonguldak, Turkey on June 9, 1990. He graduated from IMKB Anatolian Teacher's High School, Zonguldak 2008. He received his B.Sc. degree in Computer Engineering from Tobb University of Economics and Technology , Ankara, in 2013. Then, He started M.Sc. degree in Computer Science and Engineering and joinged the COSBI Lab at Koç University, Istanbul, 2013. During Master Degree study from 2013 till 2015, he worked as a teaching and research assistant at Koç University. He currently lives in Istanbul.