Identification of Important Key Regulators in Cervical Cancer Network

Asmita Das* and Kapil Jangra

Department of Biotechnology, Delhi Technological University, India

*Corresponding author: Asmita Das, Department of Biotechnology, Delhi Technological University, India, Email: asmitadas1710@dce.ac.in

Abstract

Cervical cancer is a major cause of death in women. The total number of genes that are associated with cervical cancer is 3099. We created a PPI network and then studied the topological properties of the network that gave evidence that the network is scale free, the degree distribution obeys power law and have some nodes (vertices) that play a crucial role during cancer progression. The aim of this project was to identify these nodes (also called key regulators) and the strategy used for this, is based on Barabasi-Albert model (which is a bottom up approach).The model suggested that the network we created, is evolved in nature by growth and preferential attachment concept, and have hub nodes. Finally, we concluded that there are seven important key regulators (viz. CDK2, E2F1, CDKN1A, CDKN2A, TP53, PTGS2 and CTNNB1) that are present in our network, which implies that if we are able to delete these key regulators, the whole network would collapse. So, by understanding the complex functionality and regulation of these fundamental key regulators, we can identify the diagnostic biomarkers and can develop early detection techniques and therapy for cervical cancer.

Keywords: Cancer network; Cervical cancer; Scale free network; Cancer biomarker; Therapies; CDK2; E2F1; CDKN1A; TP53; PTGS2; CTNNB1

Abbreviations: CCDB: Cervical Cancer Database; STRING: Search Tool for the Retrieval of Interacting Genes/Proteins; GEO: Gene Expression Omnibus; FC: Fold Change value

Introduction

Cervical cancer is cancer of the cervix, in women, the narrow opening into the uterus from the vagina. The normal "ectocervix" (the portion of the uterus extends into the vagina) is covered with flat, thin cells called squamous cells. The "endocervix" or cervical canal is made up of another kind of cell called columnar cells.

Most cervical cancers (80 to 90 percent) are squamous cell cancers. Adenocarcinoma is the second most common type of cervical cancer, accounting for the remaining 10 to 20 percent of cases. Adenocarcinoma develops from the glands that produce mucus in the endocervix. The risk factors for cervical cancer are Human papilloma virus infection [1], smoking, Immunosuppression, being overweight, long-term use of oral contraceptives (birth control pills), having multiple full-term pregnancies [2] and family history of cervical cancer.

There are 3099 genes that associated with cervical cancer progression (validated by comparing the gene

expression profiles of microarray data taken from gene expression omnibus, an online database of NCBI). By using Cytoscape tool we created a PPI network for cervical cancer. The next step of this project is to prove the network we created is scale-free. All real networks (real networks are not created, they evolved by time) exist in nature, are scale free network [3,4] The Scale free network is defined as the network that reproduces power law degree distribution [4,5], contain hubs nodes with a very high number of links and the distribution of node linkages follows a power law. In these networks, most nodes have a few connections (very low degree) and some nodes have a large number of links (very high degree) which means the network has no scale. Scale free network [3,4] has some features (degree distribution, clustering coefficient, contains communities, sub communities, hub nodes and their reliance on these hubs) over a random network (that reproduces Poisson degree distribution instead of power law distribution [6] and if a network also showing these features, it must be a scale free network. So, in order to prove our network is scale free, we have taken the help of Network Analyzer [7], a plug-in of Cytoscape tool. The next step is to find communities, sub communities [8,9] that are present in the network, by using R language. The purpose of finding communities and sub communities [8,10] is to find key regulators and to prove that our network is evolved in nature by growth process and preferential attachment. Barabasi-Albert model is based on growth process [5] (The network starts from few nodes and it grows as new nodes are added) and preferential attachment (The probability that a new node will connect to an existing node, is not uniform. The probability is proportional to degree of existing node) and it is a generative model proposed to understand the mechanism responsible for emergence of scale free network, in real world. Finally, we concluded that there are 7 important key regulators (CDK2, E2F1, CDKN1A, CDKN2A, TP53, PTGS2 and CTNNB1) present in our network that play an important role in cervical cancer progression. If we target these fundamental key regulators, the whole network will break into tiny non-communicating islands [11,12] that will be helpful in the cure of cervical cancer.

Materials and Method Used

Initially we get the list of all genes associated with cervical cancer from CCDB (Cervical Cancer Database). Cervical cancer gene database is a manually curated database [13] (available online), compiled 537 genes that are associated with different stages of cervical carcinogenesis. (The total number of references is 192). In order to find all interacting partners of each gene, we used STRING (Search Tool for the Retrieval of Interacting
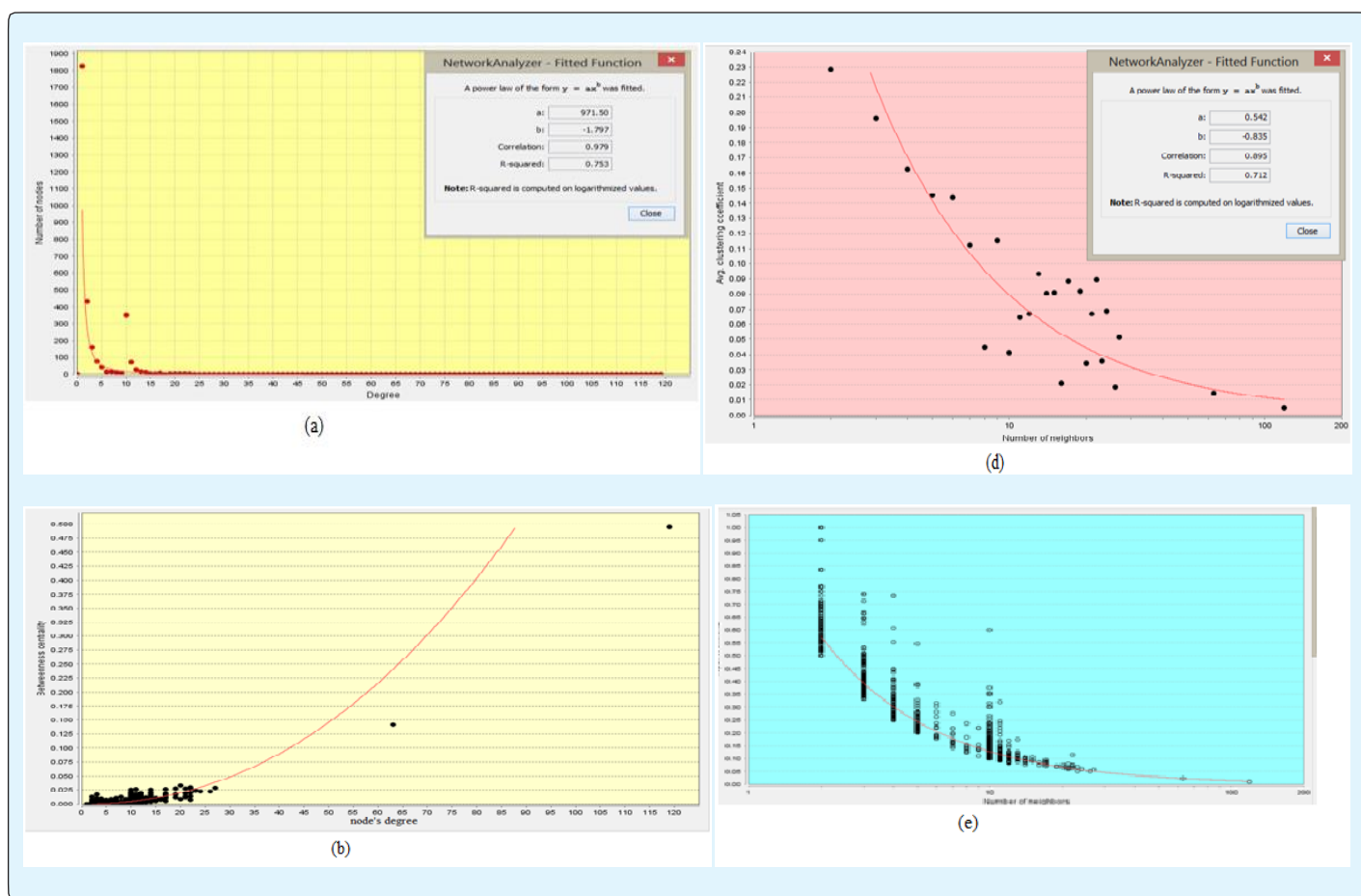
Genes/Proteins). STRING is a biological database of known and predicted protein interactions. The database is freely accessible and updated regularly. For Validation purpose we used GEO (Gene Expression Omnibus) and GEO2R. Gene Expression Omnibus at National Center for Biotechnology Information (NCBI) is a freely available repository that contains microarray data, next-generation sequencing (NGS), and other forms of high-throughput functional genomics data sets.GEO2R is a web tool of Gene Expression Omnibus database that permits users to compare two or more groups of samples in a GEO series in order to find those genes that are differentially expressed across experimental conditions. For network construction Cytoscape tool is used. Cytoscape is an open source software platform in bioinformatics that is used for molecular interaction network visualization and linked with high-throughput expression data and other molecular states into a consolidated conceptual framework. In order to study the topological properties of constructed network, Network Analyzer [7] (a plug-in of Cytoscape) is used. Network Analyzer is a very useful Cytoscape plug-in that requires no dexterous knowledge in graph theory from the user. By using Network Analyzer a perspicuous set of topological properties and centrality measures for both directed as well as undirected networks, which includes the number of nodes, edges, and connected components, the network diameter, radius, density, centralization, heterogeneity, clustering coefficient and the characteristic path length can be obtained. Besides topological parameters, Network Analyzer plots charts of the degree distribution of nodes, neighborhood connectivity's, average clustering coefficients and shortest path lengths. Finally, we used R programming language in order to find communities and sub communities [8] in the cervical cancer network. R is a programming language, developed at Bell Laboratories, the USA by John Chambers and colleagues provide an environment or platform for statistical analysis and graphics. R covers a broad area of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering) and graphical field and highly flexible. R Studio is a freely available IDE (integrated development environment) for R language.

Results and Discussion

The total number of genes associated with cervical cancer (as mentioned in Cervical Cancer Gene Database) is 537, also known as source genes [13]. In order to construct network, all target genes should be known, so that a complete network of genes associated with cervical cancer can be obtained. By using online tool STRING (Search Tool for Retrieval of Interacting Genes/Proteins (https://string-db.org/)), we have found all the

interacting partners (called Target Genes) of 537 genes. Now, the total number of genes is 3241. Only those genes are included in the list whose STRING interacting score greater than 0.90. After finding the target genes with the help of STRING database, validation of target genes through micro array data was done, in order to check the genes, we found, whether expressed in cervical cancer cells as well as in normal cervix cells or not. We Downloaded the gene expression microarray data of normal cervix cell and cervical cancer cell having GEO id GSE30758 and GSE27469 respectively from GEO database (https://www.ncbi.nlm.nih.gov/geo/) and check how many genes (source and target genes) are common in these two microarray data. After comparison, we found that 3300 genes are common in normal cervix gene micro array data and Target genes and 3245 genes are common in Target Genes and cervical cancer gene microarray data. From this comparison, we concluded that the common genes we found, are associated with cervical cancer progression because the genes are expressed in cervical cancer cells wi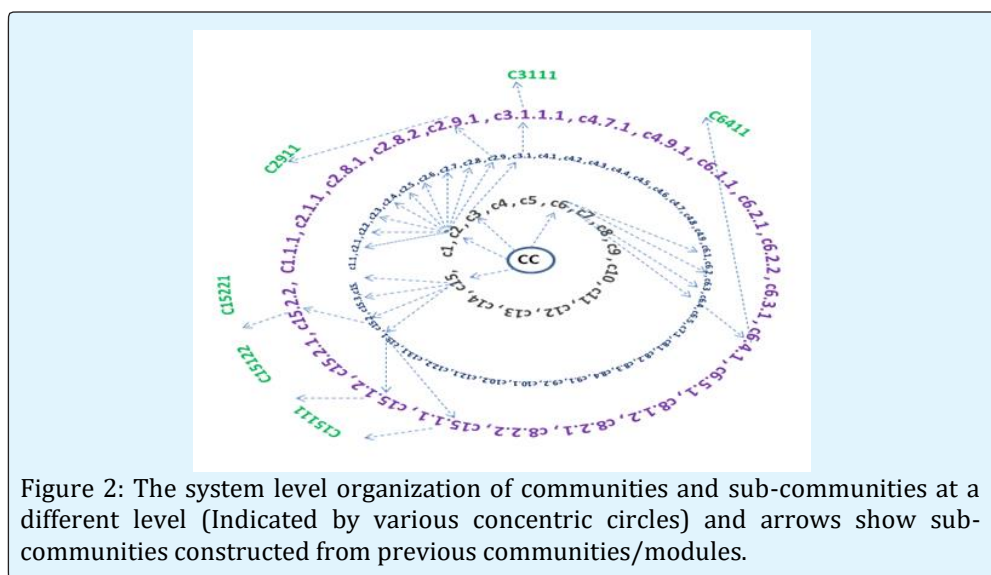th different log FC (Fold Change value) as compared to normal cervix cells, which means that the change in log FC of common genes (3241) is responsible for changing the state of normal cervix cell into cervical cancer cell. The validation process concluded that when a normal cervix cell changed into a cervical cancer cell, the fold change (log FC) of 3241 genes is changed accordingly. We prepared the PPI network of all genes associated with cervical cancer by using Cytoscape tool (http://www.cytoscape.org/). In the network, node represents a gene or protein, and line represents an interaction of the two nodes. Degree of each node is equal to the number of nodes that interact with this node. Initially, the network contains 3241 nodes and 5211 edges. Removed all non-interacted regions, self-loops and duplicated edges from the network, in order to make Cervical Cancer network more significant and relevant. So, the final network contains 3099 nodes and 4932 edges. Once the network was obtained, Network Analyzer [7] (plug-in of Cytoscape) used for the topological studies of the network (Figure 1a-f).

(a)

(d)

(b)

(e)

Asmita Das and Kapil Jangra. Identification of Important Key Regulators in Cervical Cancer Network. Ann Immunol Immunother 2019, 1(1): 000101.

Copyright© Asmita Das and Kapil Jangra.

Figure 1: (a) The node degree distribution of cervical cancer associated gene network follows the power law [6] and the value of ɣ is 1.8. Which means that network we constructed is scale free and network has hub nodes that have important role in CC network [14,15]. (b) Graph plotted between betweenness centrality of each node and their respective degree. Nodes with high betweenness may have great influence on network by virtue of their control over information passing between other nodes, and removal of these nodes will disrupt communications between other nodes. The value of betweenness centrality for each node lies between 0 and 1. There are only few nodes in CC network that have high betweenness centrality that conformed their influence or dominance on network. (c) Graph of closeness centrality v/s number of neighbors. There are few nodes that have closeness centrality value greater than average, shown in graph, hence confirmed their dominance in network. Closeness is reciprocal of the farness. The closeness centrality of a node can be any number between 0 and 1 (0 for isolated node and 1 for highly centred node in a network). (d) Graph-plot of avg. clustering coefficient and number of neighbors. The curved line represents that graph follow power law. (e) The graph of topological coefficients v/s number of neighbors. It can be used to examine the ability of the node to have shared neighbors, in the network. (f) Graph for stress centrality distribution. If a node has high stress centrality value, it has more influence in network. This figure clearly shown that there are some nodes present, in our network have high stress centrality.

After analyzing the topological properties of cervical cancer associated gene network we conclude that the network made by us is a scale free network and it is evolved in nature because it fulfilled all the conditions that are necessary for being a scale free biological network. The network contains a few nodes that are dominant or showing huge influence in the network. The next step was to find communities and sub-communities [8] present in the network. R Studio (an IDE based on R language) used for this (Figure 2-4).

Figure 2: The system level organization of communities and sub-communities at a different level (Indicated by various concentric circles) and arrows show sub-communities constructed from previous communities/modules.
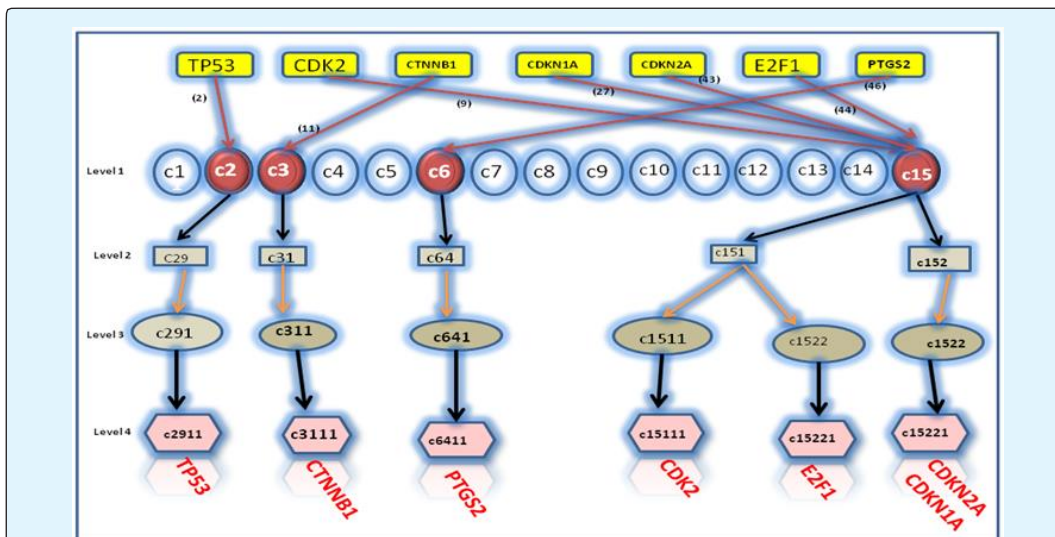
Figure 3: Paths of Important key regulators from complete network to motif through various modules/sub-modules at various level of organization.
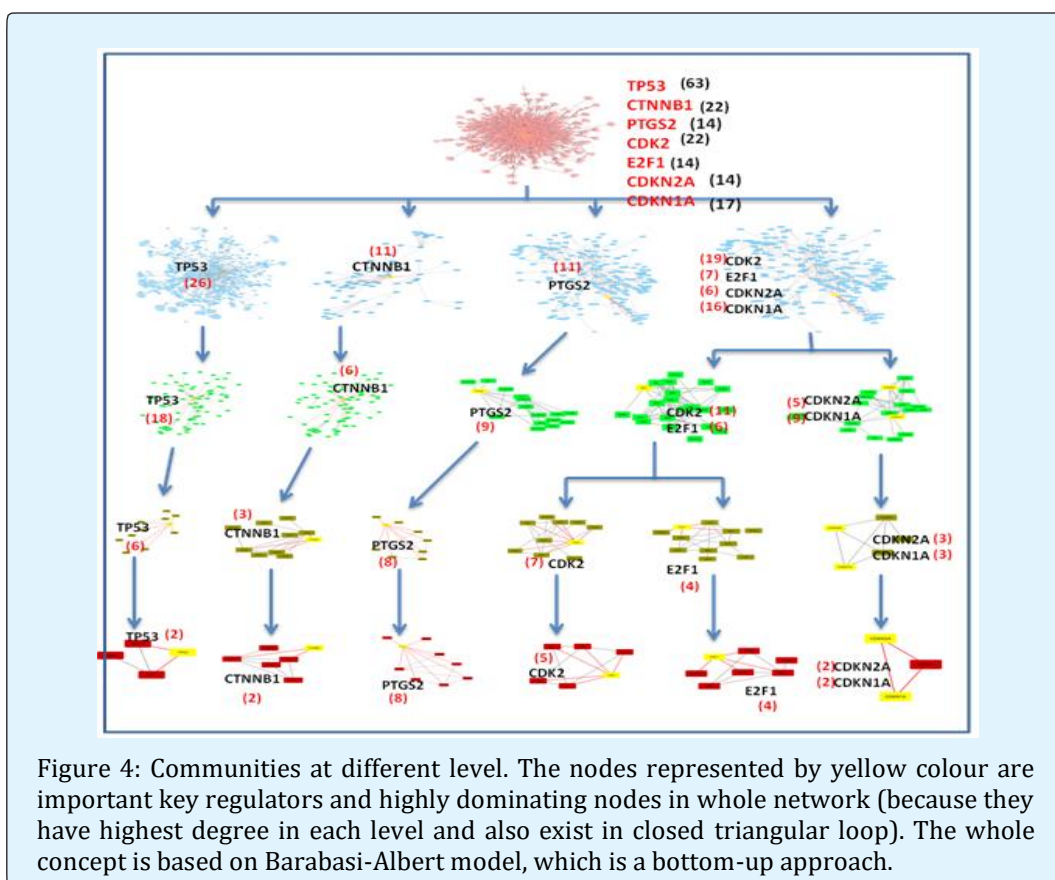

Figure 4: Communities at different level. The nodes represented by yellow colour are important key regulators and highly dominating nodes in whole network (because they have highest degree in each level and also exist in closed triangular loop). The whole concept is based on Barabasi-Albert model, which is a bottom-up approach.

Figure 4 shows that initially, motif combined to each other and form sub-communities [8,9] Sub-communities forms communities [9] and communities form meta-communities [8,9] and finally, the network is obtained.

With the help of R scripts, we found 16 communities at level 1 and found sub communities in further levels. We continued the process of finding communities until or unless we get motif. The motif is the closed triangular

Asmita Das and Kapil Jangra. Identification of Important Key Regulators in Cervical Cancer Network. Ann Immunol Immunother 2019, 1(1): 000101.

Copyright© Asmita Das and Kapil Jangra.

shaped tread of three nodes. As we know that biological network is not constructed, they evolved with time and follow growth and preferential attachment concepts. So, we can say that cervical cancer network is also evolved and formation of network starts from a motif that is shown in figures 3, 4. Now the most important node that is also known as a key regulator of a particular network is that have a maximum degree at each level which means such nodes have great dominancy at each level of network and play a crucial role in network construction because real network follows preferential attachment4 which means richer gets richer. Here, there are 7 important key regulators (CDK2, E2F1, CDKN1A, CDKN2A, TP53, PTGS2 and CTNNB1) that are the highest degree at each level, so that's why they are important for network and transferring of biological information throughout the network. If these key regulators are destroyed the whole network would collapse [11,12] Hence these important key regulators can be used as early detection biomarkers and helpful in network medicine and design new therapeutic strategies for cervical cancer.

Conclusion

As we know that cervical cancer is responsible for a large number of death in females, worldwide. More than 3000 genes are associated with cervical cancer progression. Here, we proved that the network made by us is a scale free network, which is the property of network which is real and evolved in nature by the passage of time. Scale free networks have some specific features and one of them is they contain hub nodes. By using R language we obtained communities, sub communities [8,10] at different levels and also identify the most crucial nodes of cervical cancer PPI network called important key regulators, the process of retrieving key regulators of the network is based on Barabasi-Albert model (generative model of PPI network, a backward approach). Because our network is scale free, it is robust against random node failure but vulnerable to deliberate attacks, which means if we target these important key regulators (CDK2, E2F1, CDKN1A, CDKN2A, TP53, PTGS2 and CTNNB1)(deliberate attacks [3], the whole network will break into tiny non-communicating islands [11] that will be helpful in inhibition of cervical cancer progression. Hence we can design early diagnostic markers as well as therapeutics strategies for cervical cancer, by identifying new markers.

References

1. Franceschi S, Rajkumar T,Vaccarella S, Gajalakshmi V, Sharmila A, et al. (2003) Human papilloma virus and risk factors for cervical cancer in Chennai, india: a case study 107(1): 127-133.

2. Louie KS, Sanjose S de, Diaz M, Castellsague X, Herrero R, et al. (2009) Early age at first sexual intercourse and early pregnancy are riskfactors for cervical cancer in developing countries 1191-1197.

3. Barabási AL, Bonabeau E (2003) Scale free networks 288(5).

4. Barabási AL (2009) Scale-Free Networks: A Decade and Beyond 325(5939): 412- 413.

5. Barabasi AL, Albert R (1999) Emergence of Scaling in Random Networks: 286(5439): 509-512.

6. Adamic LA, Lukose RM, Puniyani AR, Huberman BA, Labs HP, et al. (2001) Search in power-law networks 64: 046135.

7. Assenov Y, Ramırez F, Schelhorn SE, Lengauer T, Albrecht M (2007) Computing topological parameters of biological networks 24(2): 282-284.

8. Girvan M, Newman ME J (2001) Community Structure in Social and Biological Networks: 1-8.

9. Erzsebet Ravasz, Albert-Laszlo Barabasi (2003) Hierarchical organization in complex networks. Phys Rev E Stat Nonlin Soft Matter Phys 67(2 Pt 2): 026112.

10. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical Organization of Modularity in Metabolic Networks 297(5586): 1551-1555.

11. Albert R, Jeong H, Barabasi AL (2000) Error and attack tolerance of complex networks, Hierarchical organization in complex networks 406: 378-382.

12. Wang XF, Chen G (2002) Synchronization in Scale-Free Dynamical Networks: Robustness and Fragility 49(1): 54-62.

13. Agarwal SM, Raghav D, Singh H, Raghava GPS, (2010) CCDB: A curated database of genes involved in cervix cancer.

14. Jeong H, Mason SP, Barabási AL, Oltvai ZN (2001) Lethality and centrality in protein networks 411: 41-42.

15. Liu yy, Slotine JJ, Barabasi AL (2011) Controllability of complex networks 473: 167-173.