

Article

Analysis of Topological Parameters of Complex Disease Genes Reveals the Importance of Location in a Biomolecular Network

Xiaohui Zhao and Zhi-Ping Liu*

Department of Biomedical Engineering, School of Control Science and Engineering, Shandong University, Jinan 250061, Shandong, China; zhaoxiaohui199612@163.com

* Correspondence: zpliu@sdu.edu.cn; Tel.: +86-531-8839-2205

Received: 12 January 2019; Accepted: 11 February 2019; Published: 14 February 2019

Abstract: Network biology and medicine provide unprecedented opportunities and challenges for deciphering disease mechanisms from integrative viewpoints. The disease genes and their products perform their dysfunctions via physical and biochemical interactions in the form of a molecular network. The topological parameters of these disease genes in the interactome are of prominent interest to the understanding of their functionality from a systematic perspective. In this work, we provide a systems biology analysis of the topological features of complex disease genes in an integrated biomolecular network. Firstly, we identify the characteristics of four network parameters in the ten most frequently studied disease genes and identify several specific patterns of their topologies. Then, we confirm our findings in the other disease genes of three complex disorders (i.e., Alzheimer's disease, diabetes mellitus, and hepatocellular carcinoma). The results reveal that the disease genes tend to have a higher betweenness centrality, a smaller average shortest path length, and a smaller clustering coefficient when compared to normal genes, whereas they have no significant degree prominence. The features highlight the importance of gene location in the integrated functional linkages.

Keywords: disease gene; network topological parameter; location in network; systems biology

1. Introduction

Complex diseases such as neurodegenerative disorder, metabolism syndrome and cancer are often hypothesized as the results of molecular perturbations and dysfunctions [1–3]. In systems biology and medicine, the onset of complex diseases is considered to be the consequence of abnormal interactions among multiple genes, gene products, and metabolic compounds [3,4]. The disease genes play a driving role, causing locally original dysfunctions, and then signaling pathways spread their affections and cause global misalignment, even leading to mortality [5]. The disease genes provide causal information of the dysfunctional occurrence and development.

With the screening of high-throughput technologies, more and more putative disease genes have been identified [6]. For instance, the genome-wide association study (GWAS) provides a systematic investigation of genetic variants in case-control population individuals to see if any variant in the genome is associated with a particular trait [7]. As we all know, the papers about biomedicine and health have been documented in the national library of medicine [8]. There is summary statistics about the most popular genes studied in these papers. It is found that the most popular genes are all related to complex diseases in oncology and immunology [9].

A network provides a mathematical framework for deciphering the relationship between biomolecules [10–12]. In which, nodes refer to the biomolecules and edges refer to their relationships.

The genes and their products perform their normal or dysfunctional roles by interacting with each other exactly in the form of a network [4,13]. Due to recent rapid increase in biomolecular interaction data, the study on biomolecular networks has been booming [14,15]. For instance, the number of protein interactions deposited in the STRING database has grown up to 1380 million in thousands of organisms [16]. In a network, its topological parameters define the quantitative patterns and measures of the nodes and edges [14]. For instance, the degree of a node refers to the number of edges incident to the node. The topological properties, such as degree and its distribution, provide detailed descriptions of the network's features. Thus, the network's localization and organization are reflected in its topological parameters.

The topological parameters also characterize the pattern of genes in a networked system [14]. There are some examples that the topological prominence implies biological essentiality in networks [17–19]. For instance, the hub genes often refer to the nodes with high degree in a network [20]. Interestingly, it has been proved that disease genes are often not hub genes for their functional importance in the network structure [21,22]. The non-conventional findings deepen our understanding about the network localization of disease genes. In a cell, an interactome often refers to the whole set of gene interactions [23]. The location of nodes in the interactome highly coordinate their specific dysfunctions during disease development and progression [24,25]. The interactome presents a map of functional layout among genes. The information flow transmitting from disease genes to normal genes are also determined by the local neighbors and environments [1,2]. Their location in the network is also the crucial determinant of dysfunctions from the systematic perspective. It is of paramount interest to explore these parameters of disease genes in the interactome.

In this paper, we focus our analysis on the topological parameters underlying disease genes on an integrated gene–gene interaction network. For generality, we firstly investigate several network topological properties of the top-ten popular disease genes. Four widely-used network properties (i.e., node degree, betweenness, clustering coefficient, and shortest path length) are identified for representing the network localization of disease genes. For specificity, we validate the observed features underlying these network properties in three typical complex diseases (i.e., Alzheimer's disease (neurodegeneration-related), diabetes mellitus (metabolism-related), and hepatocellular carcinoma (cancer-related)). We collect their disease-related genes and verify the identified topological patterns from the top-ten most popular genes in the three types of diseases, respectively. The differences of the network characteristics are discovered between disease genes and other randomly-selected normal genes. The topological distinctiveness between disease genes and their neighbors also indicate the specificity of network location in the pathogenesis of disease. For checking the independence between the parameters, their pairwise correlations are identified. The analysis highlights the importance of location of disease genes in a network.

2. Materials and Methods

2.1. Disease Genes

Firstly, we accessed the most popular genes from the list of “hot studies” in all human genomes [9]. The top-ten genes have been investigated in more than 40,000 papers. Table 1 lists their details. Apparently, these genes are all highly related to complex diseases [26]. Some of them are disease causal mutations (e.g., *TP53* (related to cancers) and *APOE* (related to Alzheimer's disease)). Some of them perform severe dysfunctions related to inflammation and abnormal phosphorylation in diseases (e.g., *TNF* and *AKT1*). For generality, we identified their topological parameters in our integrated biomolecular network.

Table 1. The top-ten most popular genes.

Symbol	Gene ID	Function
<i>TP53</i>	7157	Associated with the tumor suppressor protein p53. Mutation causes more than half of human cancers.
<i>TNF</i>	7124	An immune molecule associated with tumor necrosis factor, as a major drug target in inflammatory diseases.
<i>EGFR</i>	1956	Associated with the epidermal growth factor receptor, a membrane-bound receptor protein. Mutations may lead to drug-resistant cancer.
<i>VEGFA</i>	7422	Associated with vascular endothelial growth factor A, helping promote the growth of blood vessels.
<i>APOE</i>	348	Associated with cholesterol and lipoprotein metabolism, namely apolipoprotein E.
<i>IL6</i>	3569	Associated with an immune molecule namely interleukin 6, helps to stimulate and suppress inflammation.
<i>TGFB1</i>	7045	Associated with transforming growth factor β 1, used to regulate cell proliferation and differentiation.
<i>MTHFR</i>	4524	Associated with process amino acids, an enzyme named methylene-tetrahydrofolate reductase.
<i>ESR1</i>	2099	Associated with a nuclear receptor protein, estrogen receptor 1, which plays an important role in breast, ovarian, and endometrial cancers.
<i>AKT1</i>	207	Associated with a signaling protein namely kinase, helping activate other proteins by phosphorylation.

For specificity and justification, we also verified the findings of the top-ten-popular genes in three other complex diseases (i.e., Alzheimer’s disease (AD), diabetes mellitus (DM), and hepatocellular carcinoma (HCC)). The disease genes of these complex diseases were composited from KEGG [27], GWASdb [6], and GWAS Catalog [8]. The number of disease genes for AD, DM, and HCC were 171, 46, and 168, respectively. They are available from Supplementary material S1.

2.2. Integrated Biomolecular Network

For building up an interactome, we downloaded the documented biomolecular networks from STRING [16], BIND [28], MINT [29], BioGrid [30], IntAct [31], DIP [32], and HPRD [33]. In these databases, multiple types of biomolecular interactions are included in the integrated network. For completeness, the interactions referring to gene–gene co-expression, gene co-occurrence, gene fusion, gene regulation, and annotated pathways were all contained in the interactome. In total, the biomolecular network contained 7018 nodes and 224,127 edges. Figure 1 illustrates the global view of a typical part of the integrated gene–gene functional linkage network.

$$L = \frac{2}{N(N+1)} \sum_{i \leq j, i, j \in G} d_{ij}, \quad (1)$$

where N represents the number of nodes in the network $G(V,E)$, node i and node j are in the network G . Here, the distance from a node to itself is defined to be zero.

2.3.3. Clustering Coefficient

Clustering coefficient (CC) reflects the aggregation property underlying the nodes in a network [14], which refers to the tendency of gathering together of these nodes. In the network, CC depicts the average value of the ratio of the actual edge of a node in the complex network to all the possible edges, in essence,

$$CC = \frac{n}{\binom{k}{2}} = \frac{2n}{k(k-1)}, \quad (2)$$

where n is the number of edges connected to the node and its first-order neighbors, and $\binom{k}{2}$ is the number of its adjacent neighbors. The key of calculating CC is to obtain this value. We employed the node neighbor subgraph algorithm to achieve it from the adjacency matrix of the network. We found all neighbors of a location element in the adjacency matrix and got the number of edges. In an undirected graph, because of the symmetry of its adjacency matrix, the number of neighbors is two times of the number of edges. Therefore, the actual value should be divided by 2 [14].

2.3.4. Betweenness Centrality

Based on shortest paths, betweenness centrality (BC) is a measure of centrality for assessing the importance of individual nodes in a network [14]. The BC value of a node is the ratio of all the shortest paths between the other nodes to the total number of these shortest paths that pass through the node, in essence,

$$BC(v) = \sum_{i \neq v \neq j} \frac{\delta_{ij}(v)}{\delta_{ij}} \quad (3)$$

δ_{ij} is the number of the shortest paths from node i to node j . $\delta_{ij}(v)$ refers to the shortest paths from node i to node j , through node v .

2.4. Comparison Statistics

We compared the topological parameters of disease genes and the randomly-sampled same-size genes (no overlap between them). In accessing the statistical significance of their difference, we used the Mann–Whitney U test to obtain a p -value between the parameters in the two gene sets.

To assess the relationship between two topological parameters, we used the Spearman rank correlation coefficient to measure their association. Spearman rank correlation is a nonparametric measure of statistical dependence between two variables [37].

3. Results

3.1. The Values of Network Topological Parameters

In the integrated biomolecular network, we identified the values of network property parameters for the top-ten popular disease genes respectively. Table 2 lists the values of their topological properties. For illustrating the specificities of parameters in these disease genes, we

randomly chose ten other genes and calculated their topological parameters correspondingly. We repeated the random processes ten times. Figure 2 demonstrates the comparisons of the four parameters in the popular-studied genes with those in the same number of randomly-selected genes. In each subfigure, the genes were sorted based on the values of the parameters.

Table 2. The values of network topology parameters in the top-ten popular genes.

Gene	Topology Parameter			
	Degree	Average Shortest Path Length	Clustering Coefficient	Betweenness Centrality
<i>TP53</i>	178	2.57	0.39	1.17E-01
<i>TNF</i>	137	2.12	0.47	1.64E-02
<i>EGFR</i>	228	2.38	0.42	3.06E-02
<i>VEGFA</i>	85	2.23	0.45	8.05E-03
<i>APOE</i>	68	2.50	0.40	1.03E-02
<i>IL6</i>	100	2.18	0.46	1.45E-02
<i>TGFB1</i>	96	2.46	0.41	3.59E-02
<i>MTHFR</i>	102	2.45	0.41	6.63E-02
<i>ESR1</i>	122	2.55	0.39	6.41E-02
<i>AKT1</i>	141	1.99	0.50	1.23E-01

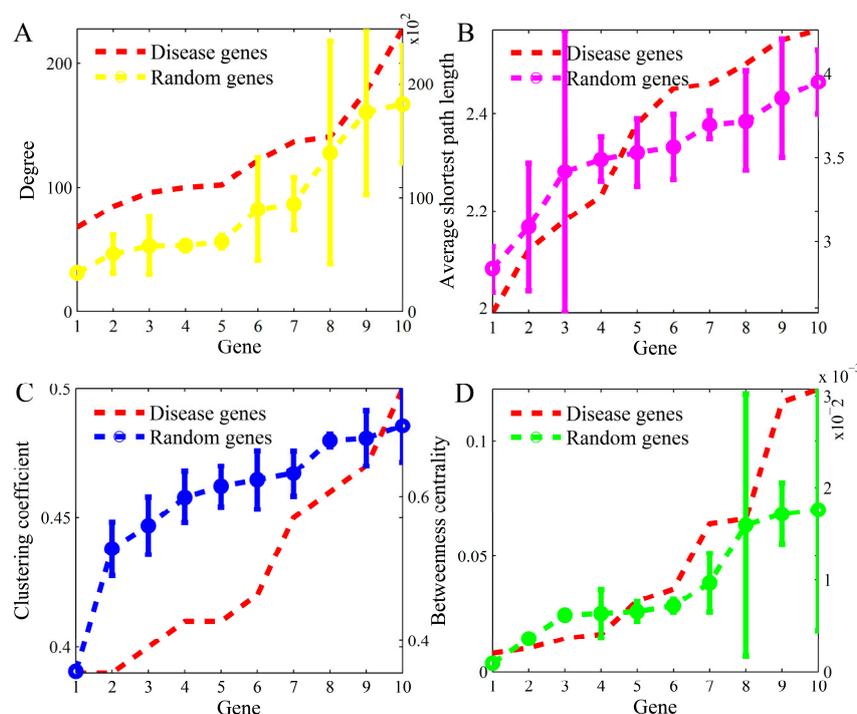


Figure 2. The comparisons of four topology parameters between disease genes and randomly-selected normal genes: (A) degree, (B) average shortest path length, (C) clustering coefficient, and (D) betweenness centrality. The x-axis is for ten genes, the left y-axis is for “mean” and the right y-axis is for “standard deviation”.

As shown in Figure 2, we found that the degree and BC of disease genes were, on average, higher than those of random genes. Whereas the average shortest path length and CC of disease genes were, on average, lower than those of random genes. In fact, degree is also a kind of centrality measure referring to the centrality of node in a network. The higher centrality of degree and betweenness indicated the crucial roles of genes in the network. Moreover, the lower average shortest path length also proved the importance of disease genes. Interestingly, the CC of random genes was generally higher than those of disease genes. This means that disease genes did not tend to be the central genes in network modules. The results provide evidence that the disease genes were located in critical positions in the network. Their locations reflect their functional importance in the interactome.

We also certified the comparison results in larger number of random samplings. We randomly selected ten groups of genes with ten genes per group, respectively. Then, the four topological parameters were calculated. The results are shown in Figure 3. The comparisons illustrated that the disease genes tended to have a higher degree and betweenness centrality, but they also tended to have smaller average shortest path length and clustering coefficient. These are consistent with those shown in Figure 2.

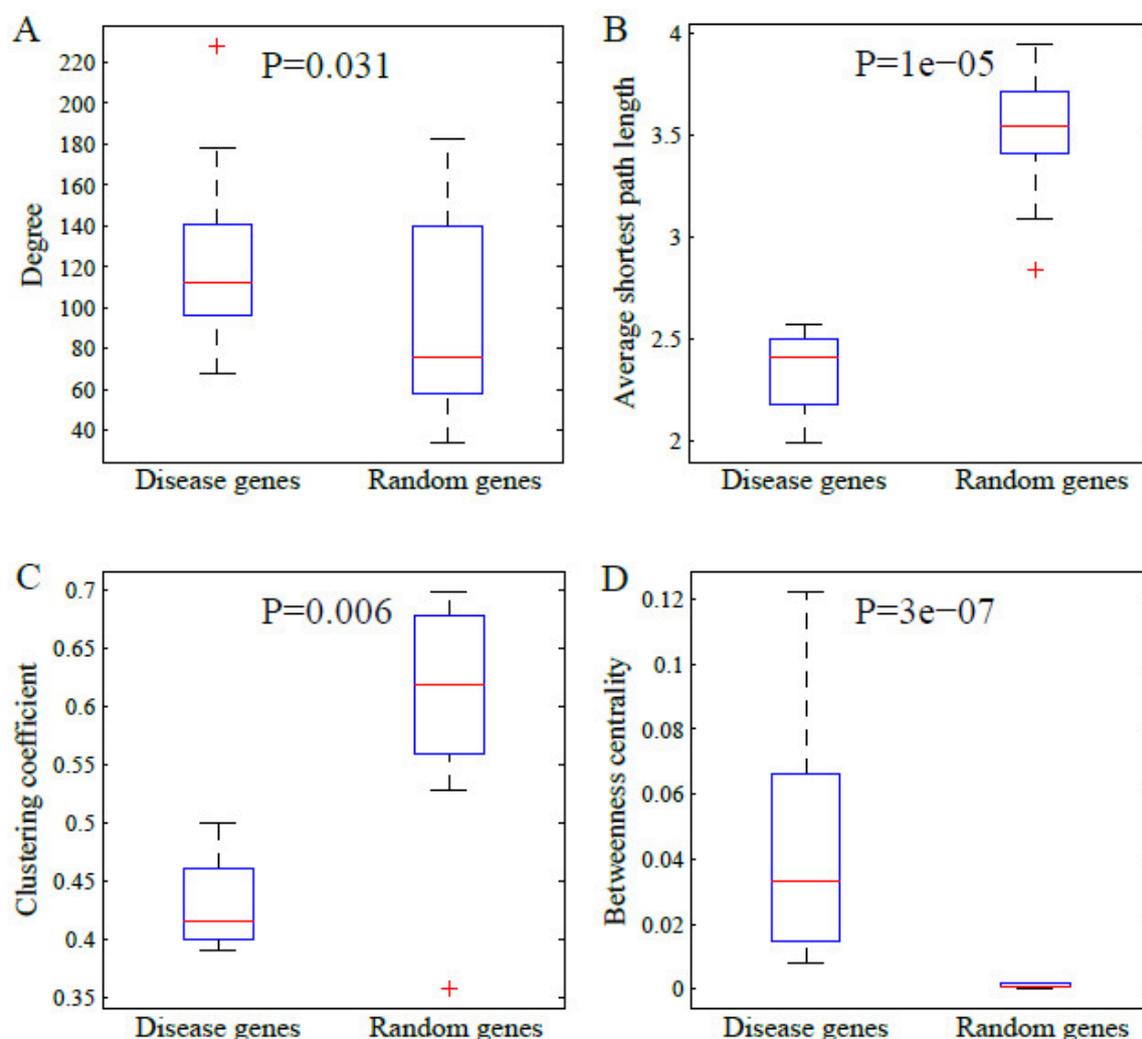


Figure 3. The boxplots of the topological parameters in disease genes and groups of random genes: (A) degree, (B) average shortest path length, (C) clustering coefficient, and (D) betweenness centrality. The p -value in each subfigure refers to the statistical significance of difference via Mann-Whitney U-test.

3.2. The Parameters in the First-Order Neighbour Genes

The former parameters indicated the importance of network localization in the ten most popular genes. It is of interest to identify the four parameters of their first-order neighbor genes for their closeness with disease genes. Figure 4 demonstrates the ten genes and their first-order neighbors in the interactome. The boxplots of their topological parameters are shown in Figure 5. For comparison, we also plotted the corresponding parameters of the ten genes. As shown in Figure 5, we found that the network parameters of the first-order genes were distinct with those of the ten disease genes. Compared with Figure 3, we found that the parameter patterns in the first-order neighbor genes were similar to those in the random genes. This proves the disease genes had different topological parameters to their first-order neighbor genes. The result indicates that the disease genes were located

in special positions in the interactome and the network localization was a major determinant factor for their malfunctional roles.

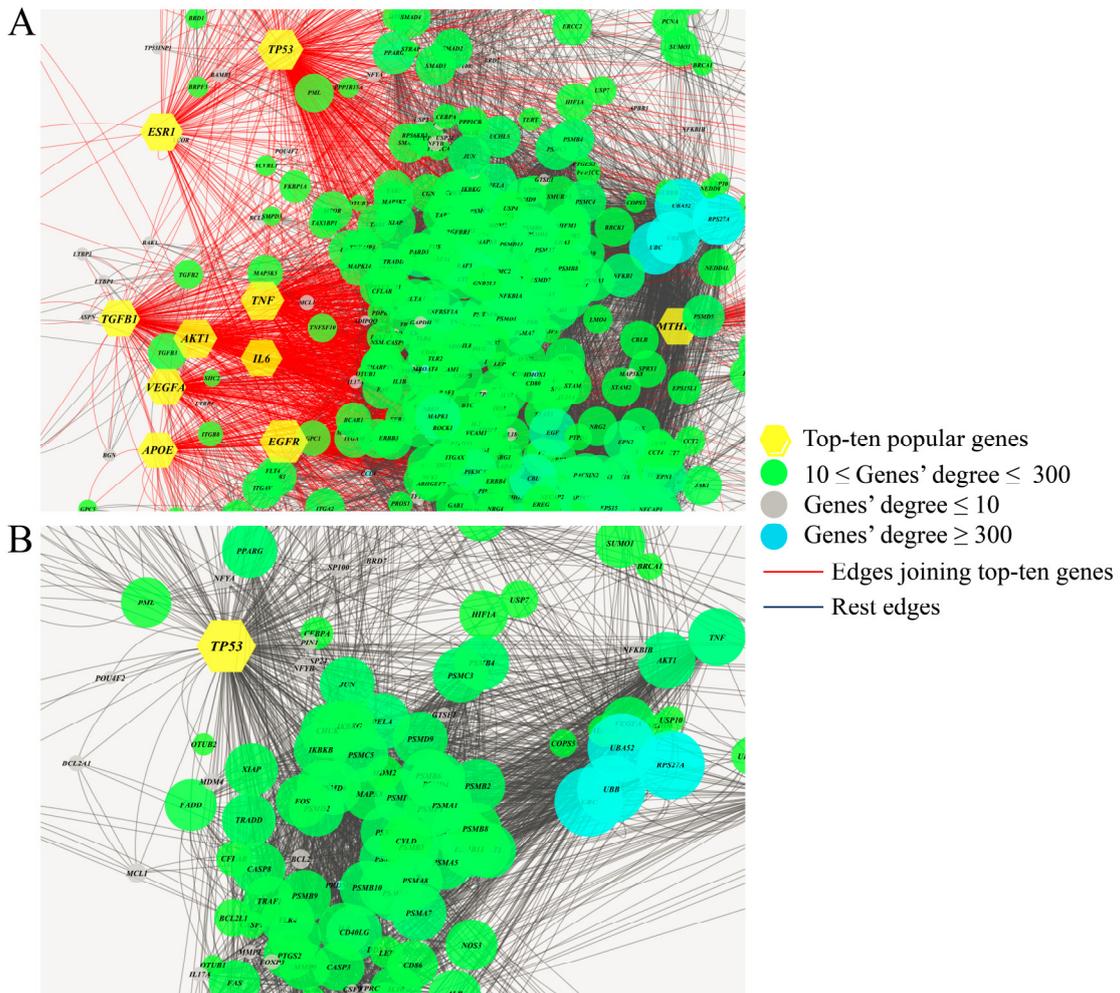


Figure 4. The network of ten popular disease genes and their first-order neighbor genes. (A) The top-ten popular genes and their nearest neighbors. (B) TP53 and its neighbor genes.

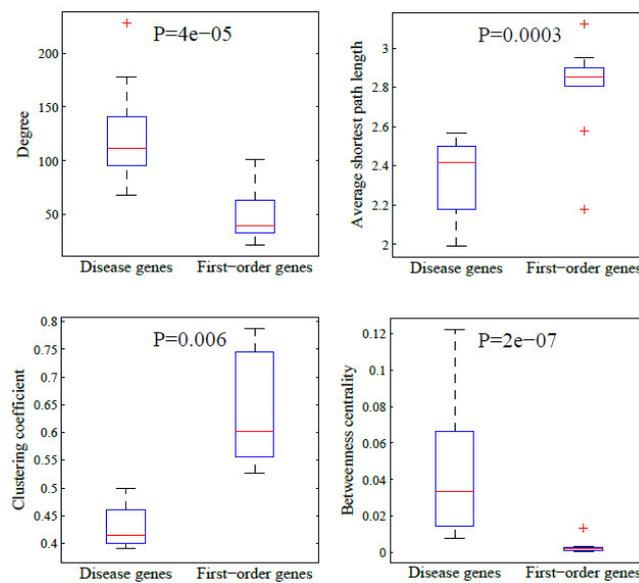


Figure 5. The comparison of the four network parameters in the top-ten popular genes and those in their first-order neighbor genes.

3.3. The Parameters in the Other Disease Genes

We identified the specific network topological patterns underlying disease genes from the most popular genes. For justifying these findings in particular diseases, we further studied the topological parameters of the other disease-related genes in three major types of complex diseases (i.e., AD, DM, and HCC). For the different numbers of disease genes, we divided these genes into 10 groups with equal sizes in the three diseases. Similar to the ten most-popular genes, we calculated the statistics for the ten-gene groups in the three diseases individually. The results of AD, DM, and HCC are shown in Figure 6, Figure 7, and Figure 8, respectively.

The results in the three figures are presented in parallel manners. As shown in Figure 6, the subfigures plot the node degree, average shortest path length, clustering coefficient, and betweenness centrality of AD genes, respectively. The topological parameters in the same size of randomly selected gene sets are also shown. The results in Figures 7 and 8 are very similarly presented.

From the comparison studies, we found that the results in the three complex diseases were consistent with those identified in the top-ten popular genes, except the parameter pattern about degree. The degree of the top-ten popular genes was significantly higher than that of the random genes, whereas it was marginally higher in the three diseases than in the random genes. Even, the degree median in the three diseases was slightly lower when compared to normal genes. The independent validations from the other diseases mutually justify our findings that the disease genes in the biomolecular network often locate in important positions with special topology parameters. The location in an interactome plays an important role in determining the dysfunction of a gene in complex diseases.

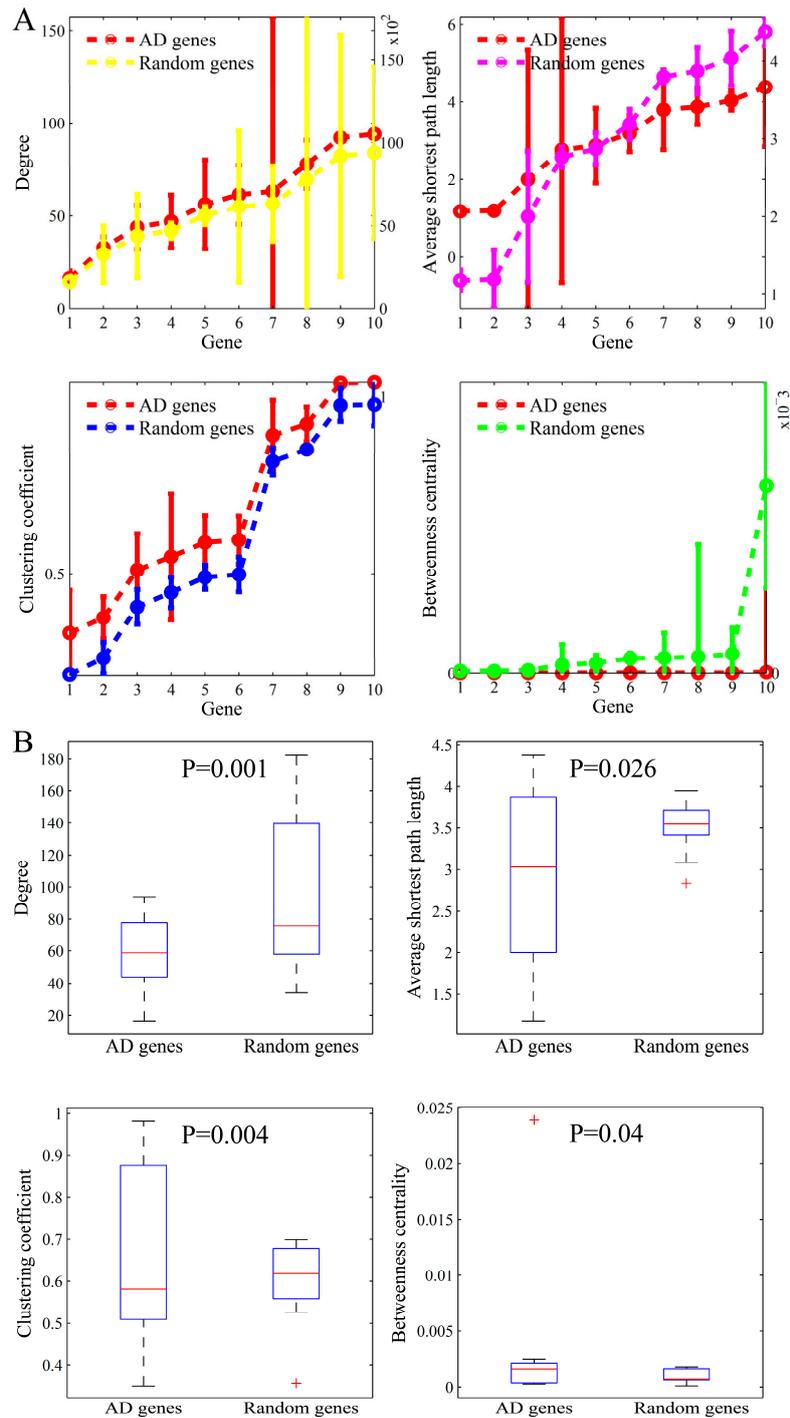


Figure 6. The comparison of topological parameters between AD (Alzheimer’s disease) genes and randomly-selected genes. **(A)** The means (left y-axis) and standard deviations (right y-axis) of the four parameters in the ten AD gene groups and those corresponding in the randomly-selected ten gene groups. **(B)** Boxplots of these parameters.

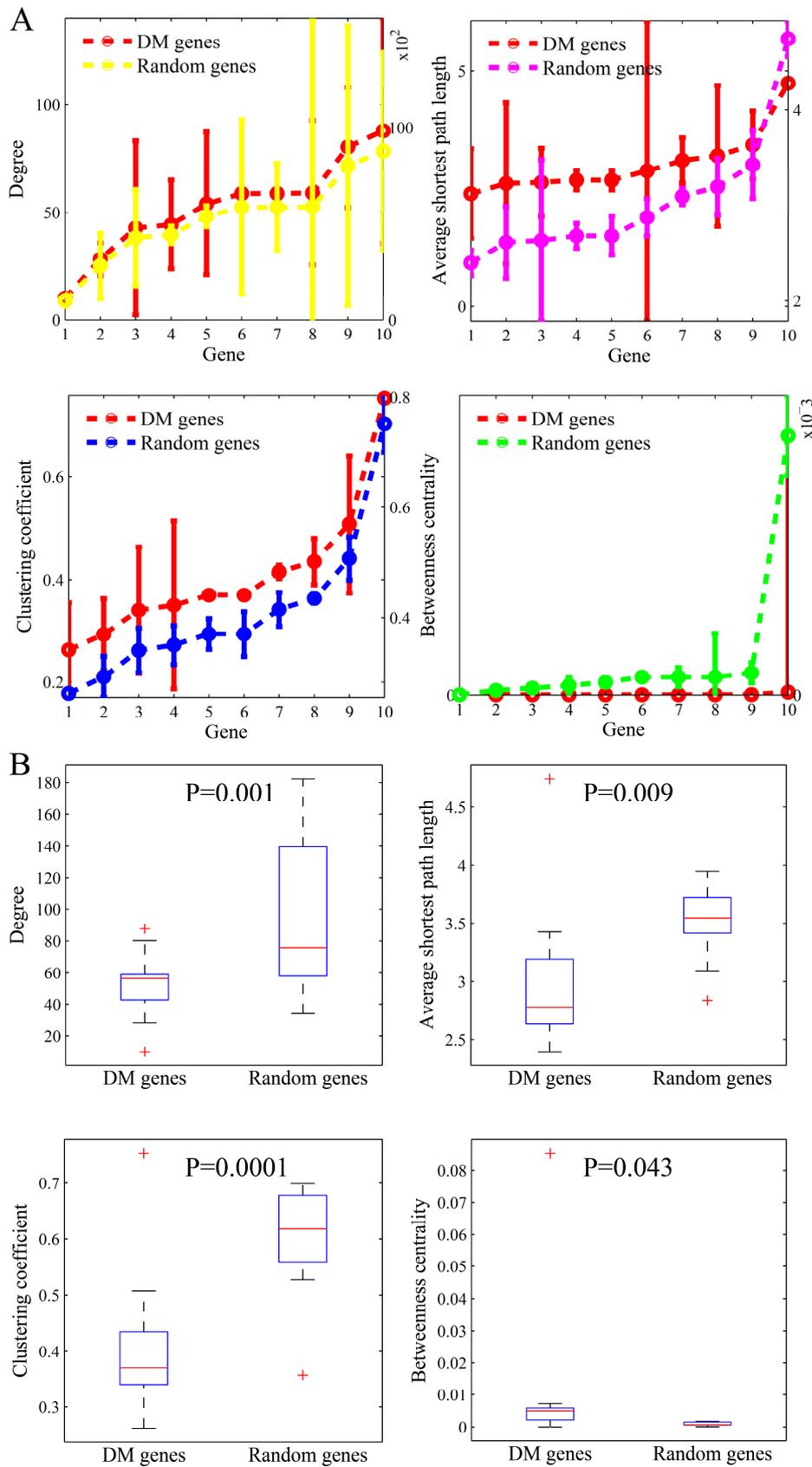


Figure 7. The comparison of topological parameters between DM (diabetes mellitus) genes and randomly-selected genes. **(A)** The means (left y-axis) and standard deviations (right y-axis) of the four parameters in the ten AD (Alzheimer’s disease) gene groups and those corresponding in the randomly-selected ten gene groups. **(B)** Boxplots of these parameters.

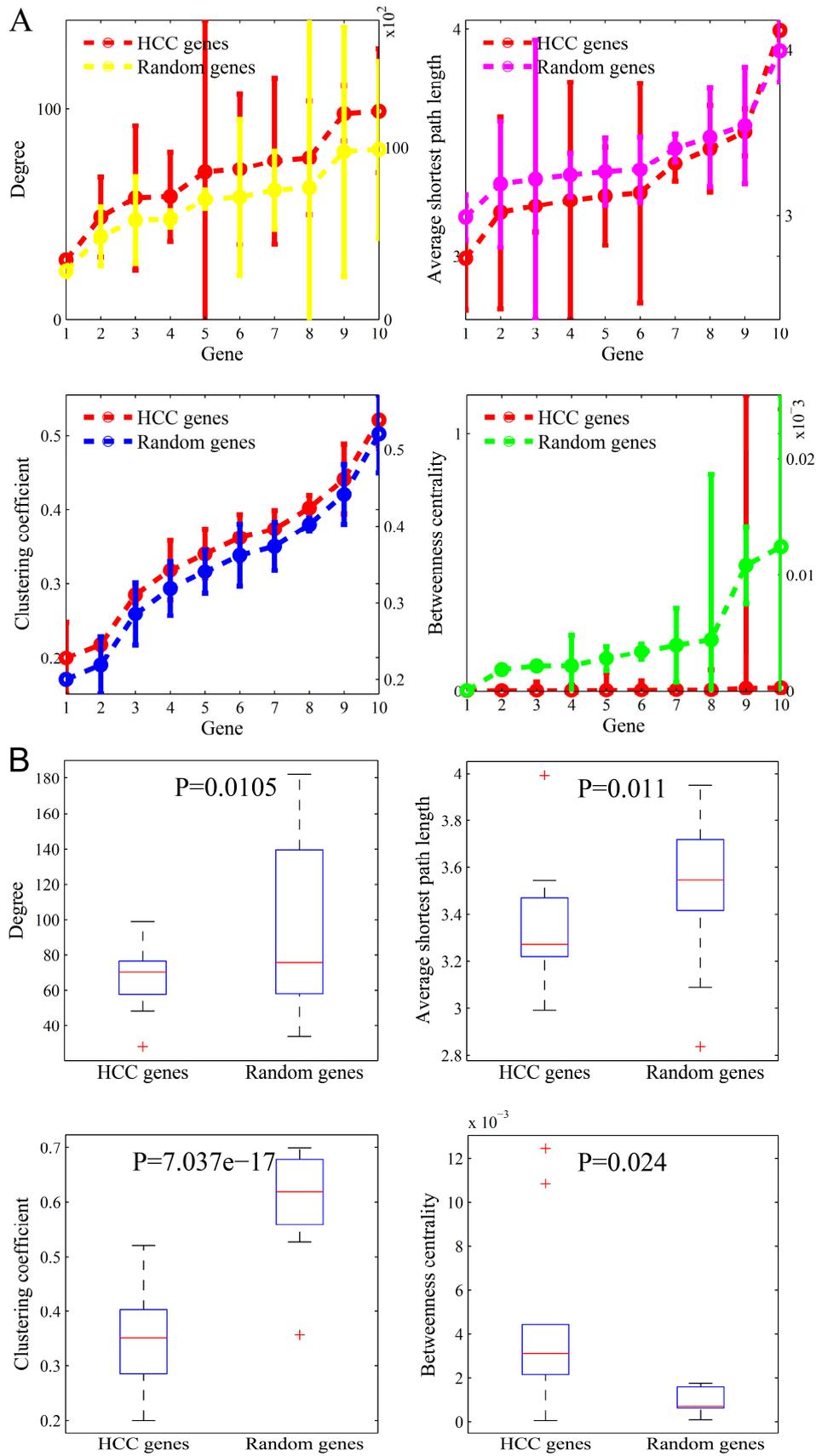


Figure 8. The comparison of topological parameters between HCC (hepatocellular carcinoma) genes and randomly-selected genes. **(A)** The means (left y-axis) and standard deviations (right y-axis) of the four parameters in the ten AD (Alzheimer’s disease) gene groups and those corresponding in the randomly-selected ten gene groups. **(B)** Boxplots of these parameters.

3.4. Correlation between Parameters

The four topological parameters are often employed to describe network properties [4]. To investigate their interrelations, it is of interest to check their correlation coefficients. In the same philosophy, we implemented the calculations of parameter correlations for the ten genes and justified the findings in the disease genes of AD, DM, and HCC. The results are shown in the following figures.

Figure 9 illustrates the all-against-all correlations ($C_4^2 = 6$) in the four parameters for the top-ten popular genes. We found that the association between average shortest path length and clustering coefficient achieved an outlier coefficient of -0.994 . The high correlation value indicated the consistency between the rankings of two topologies in the ten genes. There were no high correlations between the other pairs of four parameters. Figure 10 shows the correlations in the AD genes. The correlation between degree and betweenness centrality was as high as 0.844. The other pairs could not obtain higher correlations. In the cases of high correlation values of 0.844 and -0.715 , the betweenness centrality only achieved low values. As shown in Figure 11 for DM genes, the correlation between average shortest path length and betweenness centrality achieved the highest value of -0.660 . The relationship between degree and betweenness centrality was rather high at 0.590. In HCC genes shown in Figure 12, the correlation between average shortest path length and betweenness centrality achieved the highest value of 0.812. The outlier points in each subfigure also had no special corresponding rules between any two parameters. The results provide evidence that these parameters were relatively independent, except in a few isolated cases. Thus, they can be employed to represent the location properties in the interactome. They have specific meanings and references in describing network topologies. An interesting research direction is to investigate the differences between these network topologies and check their abilities in representing network localization. We employed all the four parameters to describe the network topology.

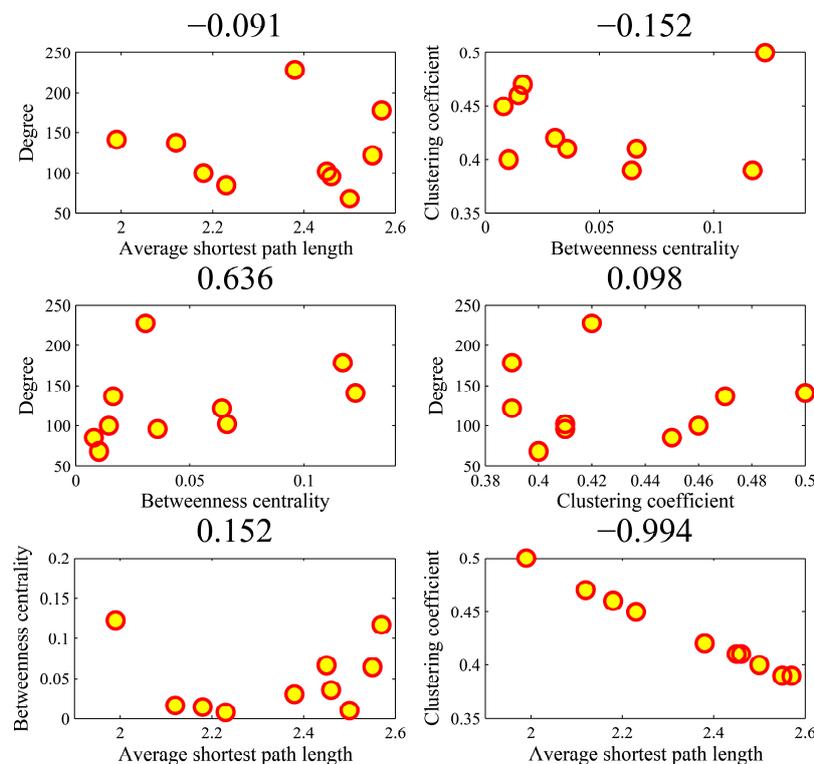


Figure 9. The Spearman rank correlations between the four parameters in the top-ten popular genes.

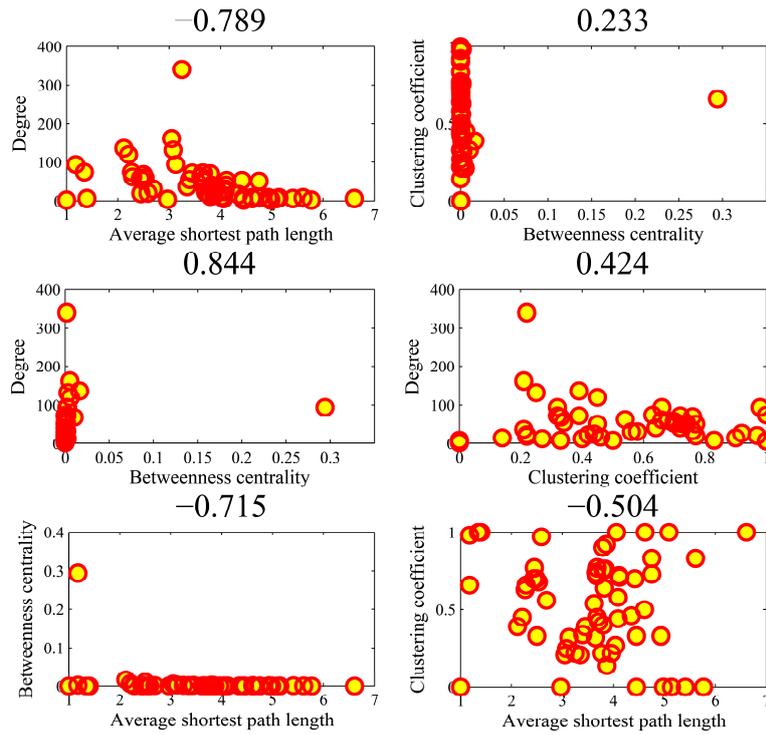


Figure 10. The Spearman rank correlations between the four parameters in the AD (Alzheimer’s disease) genes.

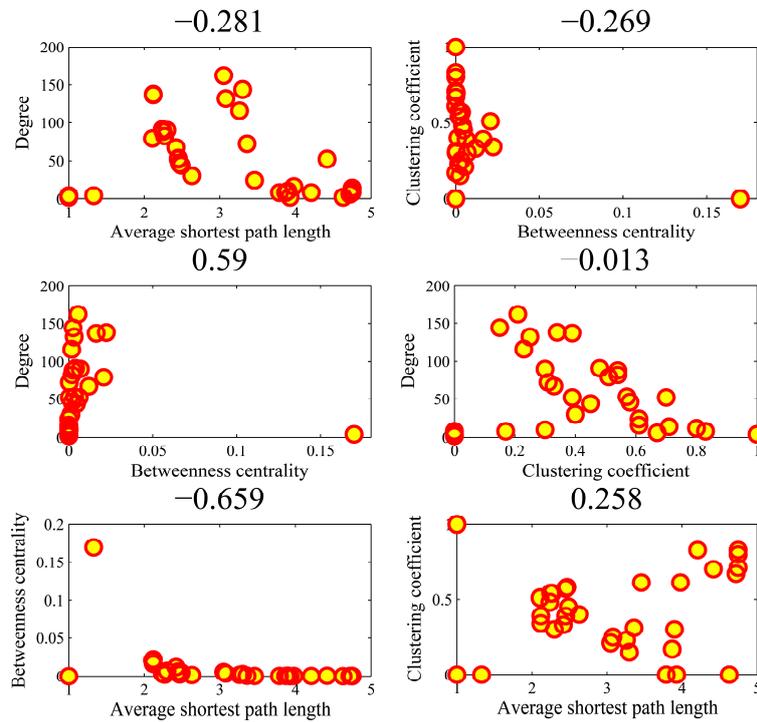


Figure 11. The Spearman rank correlations between the four parameters in the DM (diabetes mellitus) genes.

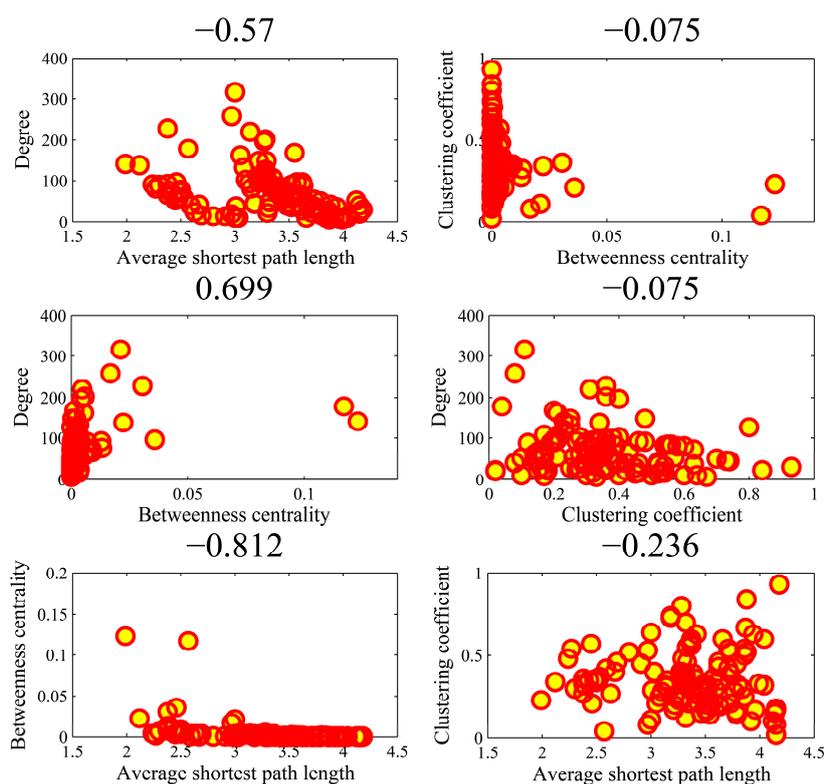


Figure 12. The Spearman rank correlations between the four parameters in the HCC (hepatocellular carcinoma) genes.

4. Conclusions

The outcome of complex diseases is contributed to from the dysfunctional interactions of multiple genes, molecules, and the environment. The complex mechanisms of disorder often bring tremendous difficulties in the prevention, diagnosis, and treatment of complex diseases. The network techniques provide powerful tools of organizing the functional relationship and structure via comprehensive interactome. In this philosophy, the network topological parameters are important to describe the network. In this work, we provide a study of investigating the network properties of disease genes in an interactome by integrating functional linkages among genes. The findings highlight the importance of location in the network for complex disease genes.

We firstly investigated the network parameters in the top-ten most popular disease genes. We found that the parameters of them are often different from those of the normal genes. This indicates the location of network is related to the dysfunctions of the complex diseases. We then confirm our findings in the disease gene sets of three major complex diseases (i.e., AD, DM, and HCC). Similar topological patterns in these disease genes were observed with those of the top-ten popular genes. We justified our results of topological importance of the disease genes in the current uncomplete interactome. The parameters were relatively quantified via the same background of interactome. When more gene interactions and context-specific networks become available, the renewed parameters will demonstrate the changing dynamics of network topology.

To further check the importance of network localization for disease genes, we studied the topological features in their nearest neighbors. Their network topological patterns were also different from those of disease genes. In order to further confirm the meaningfulness of network parameters, we identified pairs of correlations between parameters. The analyses demonstrated most of them had no significant correlation to each other. Thus, the characteristics of independent network property can be used to partially describe the network topology.

Our results and findings imply the locations of disease genes in interactome play crucial roles in disease onset and progression. The network topology indicates the pathogenesis of complex diseases. From the network perspective, it is possible to deepen our understanding of the functionality

transmission between genes and decipher novel mechanisms of complex diseases. Our analysis reveals the importance of investigating the topological structures of disease genes from the network perspective in systems biomedicine.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Supplementary material S1: Disease gene lists for Alzheimer's disease, diabetes mellitus, and hepatocellular carcinoma.

Author Contributions: Conceptualization, Z.L.; Methodology, Z.L. and X.Z.; Software, X.Z.; Formal Analysis, X.Z. and Z.L.; Data Curation, X.Z. and Z.P.; Writing-Original Draft Preparation, X.Z. and Z.L.; Writing-Review & Editing, Z.L.; Visualization, X.Z.; Supervision, Z.L.

Funding: This work was partially supported by the National Natural Science Foundation of China. (Nos. 61572287 and 61533011), the Innovation Method Fund of China (Ministry of Science and Technology of China, No.SQ2018IMC600001), the Shandong Provincial Key Research and Development Program (Nos. 2018GSF118043 and 2017CXGC1502), and the Fundamental Research Funds of Shandong University (No. 2016JC007). The paper was also supported by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, Ministry of Education of China.

Acknowledgments: Thanks are due to the editor and three anonymous reviewers for their constructive comments and suggestions, which greatly improved our paper. We thank Yanqiu Wang for her assistance.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Liu, Z.P.; Wang, Y.; Zhang, X.S.; Chen, L. Network-based analysis of complex diseases. *IET Syst. Biol.* **2012**, *6*, 22–33, doi:10.1049/iet-syb.2010.0052.
- Liu, Z.P. Identifying network-based biomarkers of complex diseases from high-throughput data. *Biomark. Med.* **2016**, *10*, 633–650, doi:10.2217/bmm-2015-0035.
- Barabasi, A.L.; Gulbahce, N.; Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **2011**, *12*, 56–68, doi:10.1038/nrg2918.
- Barabasi, A.L.; Oltvai, Z.N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **2004**, *5*, 101–113, doi:10.1038/nrg1272nrg1272.
- Ideker, T.; Dutkowsky, J.; Hood, L. Boosting signal-to-noise in complex biology: Prior knowledge is power. *Cell* **2011**, *144*, 860–863, doi:10.1016/j.cell.2011.03.007.
- Li, M.J.; Wang, P.; Liu, X.; Lim, E.L.; Wang, Z.; Yeager, M.; Wong, M.P.; Sham, P.C.; Chanock, S.J.; Wang, J. GWASdb: A database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.* **2011**, *40*, D1047–D1054, doi:10.1093/nar/gkr1182.
- Consortium, W.T.C.C. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature* **2007**, *447*, 661–678, doi:10.1038/nature05911.
- Welter, D.; MacArthur, J.; Morales, J.; Burdett, T.; Hall, P.; Junkins, H.; Klemm, A.; Flicek, P.; Manolio, T.; Hindorf, L.; et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **2014**, *42*, D1001–D1006, doi:10.1093/nar/gkt1229.
- Dolgin, E. The most popular genes in the human genome. *Nature* **2017**, *551*, 427–431, doi:10.1038/d41586-017-07291-9.
- Barabasi, A.L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512.
- Vidal, M.; Cusick, M.E.; Barabasi, A.L. Interactome networks and human disease. *Cell* **2011**, *144*, 986–998, doi:10.1016/j.cell.2011.02.016.
- Chen, S.; Fragoza, R.; Klei, L.; Liu, Y.; Wang, J.; Roeder, K.; Devlin, B.; Yu, H. An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. *Nat. Genet.* **2018**, *50*, 1032–1040, doi:10.1038/s41588-018-0130-z.
- Khurana, V.; Peng, J.; Chung, C.Y.; Auluck, P.K.; Fanning, S.; Tardiff, D.F.; Bartels, T.; Koeva, M.; Eichhorn, S.W.; Benyamini, H.; et al. Genome-scale networks link neurodegenerative disease genes to α -synuclein through specific molecular pathways. *Cell Syst.* **2017**, *4*, 157–170.e114, doi:10.1016/j.cels.2016.12.011.
- Newman, M.E.J. The structure and function of complex networks. *Siam Rev.* **2003**, *45*, 167–256, doi:10.1137/S003614450342480.
- Kang, T.; Moore, R.; Li, Y.; Sontag, E.; Bleris, L. Discriminating direct and indirect connectivities in biological networks. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12893–12898, doi:10.1073/pnas.1507168112.

16. Von Mering, C.; Jensen, L.J.; Kuhn, M.; Chaffron, S.; Doerks, T.; Kruger, B.; Snel, B.; Bork, P. STRING 7--recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* **2007**, *35*, D358–D362, doi:10.1093/nar/gkl825.
17. Jeong, H.; Mason, S.P.; Barabasi, A.L.; Oltvai, Z.N. Lethality and centrality in protein networks. *Nature* **2001**, *411*, 41–42, doi:10.1038/35075138.
18. Yu, H.; Greenbaum, D.; Xin Lu, H.; Zhu, X.; Gerstein, M. Genomic analysis of essentiality within protein networks. *Trends Genet.* **2004**, *20*, 227–231, doi:10.1016/j.tig.2004.04.008.
19. Zotenko, E.; Mestre, J.; O’Leary, D.P.; Przytycka, T.M. Why do hubs in the yeast protein interaction network tend to be essential: Reexamining the connection between the network topology and essentiality. *PLoS Comput. Biol.* **2008**, *4*, e1000140, doi:10.1371/journal.pcbi.1000140.
20. Santolini, M.; Barabasi, A.L. Predicting perturbation patterns from the topology of biological networks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E6375–E6383, doi:10.1073/pnas.1720589115.
21. Goh, K.I.; Cusick, M.E.; Valle, D.; Childs, B.; Vidal, M.; Barabasi, A.L. The human disease network. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8685–8690, doi:10.1073/pnas.0701361104.
22. Lin, L.; Yang, T.; Fang, L.; Yang, J.; Yang, F.; Zhao, J. Gene gravity-like algorithm for disease gene prediction based on phenotype-specific network. *BMC Syst. Biol.* **2017**, *11*, 121, doi:10.1186/s12918-017-0519-9.
23. Stumpf, M.P.; Thorne, T.; de Silva, E.; Stewart, R.; An, H.J.; Lappe, M.; Wiuf, C. Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6959–6964, doi:10.1073/pnas.0708078105.
24. Feldman, I.; Rzhetsky, A.; Vitkup, D. Network properties of genes harboring inherited disease mutations. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4323–4328, doi:10.1073/pnas.0701722105.
25. Cowen, L.; Ideker, T.; Raphael, B.J.; Sharan, R. Network propagation: A universal amplifier of genetic associations. *Nat. Rev. Genet.* **2017**, *18*, 551–562, doi:10.1038/nrg.2017.38.
26. Huang, J.K.; Carlin, D.E.; Yu, M.K.; Zhang, W.; Kreisberg, J.F.; Tamayo, P.; Ideker, T. Systematic evaluation of molecular networks for discovery of disease genes. *Cell Syst.* **2018**, *6*, 484–495.e5, doi:10.1016/j.cels.2018.03.001.
27. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30.
28. Bader, G.D.; Betel, D.; Hogue, C.W. BIND: The Biomolecular Interaction Network Database. *Nucleic Acids Res.* **2003**, *31*, 248–250.
29. Chatr-aryamontri, A.; Ceol, A.; Palazzi, L.M.; Nardelli, G.; Schneider, M.V.; Castagnoli, L.; Cesareni, G. MINT: The Molecular INteraction database. *Nucleic Acids Res.* **2007**, *35*, D572–D574, doi:10.1093/nar/gkl950.
30. Chatr-Aryamontri, A.; Oughtred, R.; Boucher, L.; Rust, J.; Chang, C.; Kolas, N.K.; O’Donnell, L.; Oster, S.; Theesfeld, C.; Sellam, A.; et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **2017**, *45*, D369–D379, doi:10.1093/nar/gkw1102.
31. Aranda, B.; Achuthan, P.; Alam-Faruque, Y.; Armean, I.; Bridge, A.; Derow, C.; Feuermann, M.; Ghanbarian, A.T.; Kerrien, S.; Khadake, J.; et al. The IntAct molecular interaction database in 2010. *Nucleic Acids Res.* **2010**, *38*, D525–D531, doi:10.1093/nar/gkp878.
32. Xenarios, I.; Salwinski, L.; Duan, X.J.; Higney, P.; Kim, S.M.; Eisenberg, D. DIP, the Database of Interacting Proteins: A research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **2002**, *30*, 303–305.
33. Keshava Prasad, T.S.; Goel, R.; Kandasamy, K.; Keerthikumar, S.; Kumar, S.; Mathivanan, S.; Telikicherla, D.; Raju, R.; Shafreen, B.; Venugopal, A.; et al. Human Protein Reference Database—2009 update. *Nucleic Acids Res.* **2009**, *37*, D767–D772, doi:10.1093/nar/gkn892.
34. Martin, T.; Zhang, X.; Newman, M.E. Localization and centrality in networks. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2014**, *90*, 052808.
35. Albert, R.; Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97.
36. Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **2005**, *118*, 4947–4957, doi:10.1242/jcs.02714.
37. Liu, Z.P. Quantifying gene regulatory relationships with association measures: A comparative study. *Front. Genet.* **2017**, *8*, 96, doi:10.3389/fgene.2017.00096.

