



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

A Doctor of Philosophy Dissertation

**INVESTIGATION ON MODULARITY
AND DYNAMICS IN SIGNALING
NETWORKS**

School of Electrical Engineering

University of Ulsan

TRUONG CONG DOAN

November 2017

INVESTIGATION ON MODULARITY AND DYNAMICS IN SIGNALING NETWORKS

Under the supervision of

Prof. Kwon Yung-Keun

Submitted to

School of Electrical Engineering

University of Ulsan

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy



TRUONG CONG DOAN

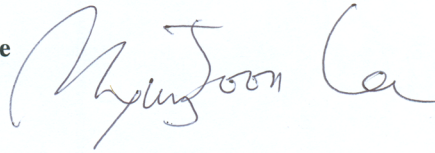
November 2017

The PhD dissertation presented by

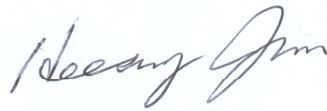
Truong Cong Doan

Approved by an academic committee:

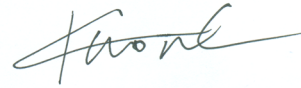
1. Chairman: **Prof. Myung-Joon Lee**
University of Ulsan



2. Member: **Prof. Hee-Sung Jun**
University of Ulsan



3. Member: **Prof. Yung-Keun Kwon, Advisor**
University of Ulsan



4. Member: **Prof. JiWon Seo**
Hanyang University



5. Member: **Dr. Dongil Seo**
Google Korea LLC



December, 2017

School of Electrical Engineering

University of Ulsan

TRUONG CONG DOAN

[ABSTRACT]

INVESTIGATION ON MODULARITY AND DYNAMICS IN SIGNALING NETWORKS

Although there have been many studies revealing that dynamic robustness of a biological network is related to its modularity characteristics, no proper tool exists to investigate the relation between network dynamics and modularity. Accordingly, I developed a novel Cytoscape app called MORO, which can conveniently analyze the relationship between network modularity and robustness. I employed an existing algorithm to analyze the modularity of directed graphs and a Boolean network model for robustness calculation. In particular, to ensure the robustness algorithm's applicability to large-scale networks, I implemented it as a parallel algorithm by using the OpenCL library. A batch-mode simulation function was also developed to verify whether an observed relationship between modularity and robustness is conserved in a large set of randomly structured networks. The app provides various visualization modes to better elucidate topological relations between modules, and tabular results of centrality and gene ontology enrichment analyses of modules. I tested the proposed app to analyze large signaling networks and showed an interesting relationship between network modularity and robustness. My app can be a promising tool which efficiently analyzes the relationship between modularity and robustness in large signaling networks.

Secondly, biological networks consisting of molecular components and interactions are represented by a graph model. There have been some studies based on that model to analyze a relationship between structural characteristics and dynamical behaviors in signaling network. However, little attention has been paid to changes of modularity and robustness in mutant networks. Therefore, I investigated the changes of modularity and robustness by edge-removal mutations in three signaling networks.

I first observed that both the modularity and robustness increased on average in the mutant network by the edge-removal mutations. However, the modularity change was negatively correlated with the robustness change. This implies that it is unlikely that both the modularity and the robustness values simultaneously increase by the edge-removal mutations. Another interesting finding is that the modularity change was positively correlated with the degree, the number of feedback loops, and the edge betweenness of the removed edges whereas the robustness change was negatively correlated with them. I note that these results were consistently observed in randomly structure networks. Additionally, I identified two groups of genes which are incident to the highly-modularity-increasing and the highly-robustness-decreasing edges with respect to the edge-removal mutations, respectively, and observed that they are likely to be central by forming a connected component of a considerably large size. The gene-ontology enrichment of each of these gene groups was significantly different from the rest of genes. Finally, I showed that the highly-robustness-decreasing edges can be promising edgetic drug-targets, which validates the usefulness of my analysis. Taken together, the analysis of changes of robustness and modularity against edge-removal mutations can be useful to unravel novel dynamical characteristics underlying in signaling networks.

ACKNOWLEDGEMENT

I would like to express my deep gratitude to my advisor, Prof. Kwon Yung-Keun. Prof. Kwon instructed and supported me a lot in my Ph.D. study and related research through his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I would be very happy if I have an opportunity to work with him again. Once again, I would like to say thanks to him because of everything he did for me.

I would like to acknowledge my committee members for their valuable comments and for their broad perspective in redefining the ideas in this dissertation.

I would like to say thanks to my friends and labmates. They helped me a lot to be familiar with the life in Korea, and shared interesting things in the life and research. With my Vietnamese friends, they also shared with me all sad and happy emotion. I also thank labmates who make a good environment in the lab.

To my parents, they always give me the greatest mental support to my study and also this dissertation. I would like to thank them for being the source of my life. All the support they have provided me over the years was the greatest gift anyone has ever given me.

Last but not least, I would like to thank my sweet family including my wife, son, and daughter. My wife cares for our lovely children thoughtfully during my PhD course and she also encourages me to try my best to do PhD successfully. Additionally, my son and daughter are biggest motivation for me to achieve my PhD's degree successfully. Thank all of you so much. My today's achievement is a small gift for you.

TRUONG CONG DOAN

Ulsan, Republic of Korea

November, 2017

VITA

Truong Cong Doan was born in Nghe an province on August 05, 1980 and has been living in Hanoi since 1998, Vietnam. He received the degree of bachelor in Applied Mathematics and Informatics (2002) from Hanoi University of Science, Vietnam. He worked for the Digitech Co., Ltd as a professional developer in Hanoi, Vietnam from 2002 to 2004. Then, He became an information technology lecturer at Tri Duc English & Informatics Center from 2004 to 2007. He also got a Master's degree in August 2007 from Le Quy Don Technical University in Hanoi, Vietnam. Then, he worked as a senior lecturer at Faculty of Information Technology in Hanoi Open University, Vietnam from 2008 to 2014. He began working full time towards his PhD at University of Ulsan, South of Korea under the guidance of Prof. Kwon Yung-Keun. Since then, he started to conduct researches in Complex Systems Computing lab, and focused on bioinformatics and parallel computing fields.

Publications

- [1] Truong C-D, Tran T-D, Kwon Y-K: **MORO: a Cytoscape app for relationship analysis between modularity and robustness in large-scale biological networks.** *BMC Systems Biology* 2016, **10**(Suppl 4):122. [SCI; 2.58]
- [2] Truong C-D, Kwon Y-K: **Investigation on changes of modularity and robustness by edge-removal mutations in signaling networks.** *BMC Systems Biology* 2017. [SCI; 2.58]
- [3] Truong C-D, Kwon Y-K: **The negative relationship between robustness and assortativity in signaling networks.** 201x. [Preparing to submit]

TABLE OF CONTENT

LIST OF FIGURES.....	12
LIST OF TABLES.....	15
CHAPTER 1. INTRODUCTION.....	16
1.1 Motivation.....	16
1.2 Research objectives	19
1.3 Dissertation outline.....	20
CHAPTER 2. BACKGROUND.....	22
2.1 Biological networks	22
2.1.1 Introduction.....	22
2.1.2 Datasets of signaling networks	23
2.2 Random network generation.....	23
2.3 Network modularity	25
2.4 Boolean network model	26
2.4.1 Introduction.....	26
2.4.2 Boolean network dynamics and in-/out-module robustness against initial states perturbation.....	28
2.4.3 Boolean network dynamics against update-rule mutation.....	29
2.5 Structural properties of network.....	31
2.5.1 Feedback loops.....	31
2.5.2 Centrality	31
2.6 Related works	32
2.6.1 Cytoscape Plugins	32
CHAPTER 3. MORO: A CYTOSCAPE APP FOR RELATIONSHIP ANALYSIS BETWEEN MODULARITY AND ROBUSTNESS IN LARGER-SCALE SIGNALING NETWORKS.....	35
3.1 Overview	35

3.2 Implementation.....	36
3.2.1 The Overall process of MORO App	36
3.2.2 Parallel computation of robustness	37
3.3 A batch-mode simulation on random Boolean networks.....	37
3.4 Visualization of relations between modules.....	37
3.5 Module centrality and GO analysis	38
3.6 Results	39
3.6.1 Analysis of modularity and robustness	39
3.6.2 Time performance analysis.....	42
3.6.3 Module centrality analysis.....	43
3.6.4 GO analysis.....	44
3.7 Conclusions	45
CHAPTER 4. INVESTIGATION ON CHANGES OF MODULARITY AND ROBUSTNESS BY EDGE-REMOVAL MUTATIONS IN SIGNALING NETWORKS	46
4.1 Overview	46
4.2 Change of modularity and robustness by edge-removal mutations.....	47
4.3 Software for statistical tests.....	47
4.4 Results	47
4.4.1 Relationship between changes of modularity and robustness by edge- removal mutations.....	47
4.4.2 Structural characteristics to affect the changes of the modularity and the robustness	49
4.4.3 Topological distribution of highly modularity-increasing and robustness- decreasing edges by removal mutations.....	52
4.4.4 Gene ontology analysis of a set of genes incident to highly-modularity- increasing or highly-robustness-decreasing edges.....	55
4.4.5 Edge-based drug discovery.....	56

4.5 Conclusions	58
CHAPTER 5. CONTRIBUTION SUMMARY AND FURTHER WORK.....	60
5.1 Contribution summary	60
5.1.1 MORO: a GPU-based software	60
5.1.2 Negative relationship between changes of modularity and robustness.....	61
5.2 Future Work.....	61
APPENDIX A	63
APPENDIX B.....	66
APPENDIX C	69
APPENDIX D	96
REFERENCES	98

LIST OF FIGURES

Figure 1.1 An illustrative example of modularity and robustness.	17
Figure 1.2 An illustrative example of edge-removal mutations.....	19
Figure 2.1 – Four kinds of biological networks.	22
Figure 2.2 An illustrative example of calculating modularity.	25
Figure 2.3 – An illustrative example of calculating the attractor similarity.	27
Figure 2.4 An illustrative example of calculating network dynamics against update-rule mutation.	30
Figure 2.5 – Cytoscape, an environment for data integration, network analysis and visualization.	33
Figure 3.1 – The overall process to analyze the relationship between the network robustness and modularity in MORO.	36
Figure 3.2 – User interface for a batch-mode simulation on RBNs.	38
Figure 3.3. Analysis results of the STKE network by MORO.	40
Figure 3.4 Changes of module centrality values against the module size in the STKE network.	43
Figure 4.1. Analysis of the changes of the modularity and the robustness by edge-removal mutations in T-LGL signaling network.....	48
Figure 4.2. Relationship of each of the changes of the modularity and the robustness with the edge-based structural properties in T-LGL signaling network.....	50
Figure 4.3. Topological distributions of High-MI/High-RD edges and their incident nodes in T-LGL signaling network.	51
Figure 4.4. Comparison of node-based centralities between High-MI-incident/High-RD-incident group and the rest of genes in the signaling networks.....	53
Figure 4.5. Edge-removal analysis for edgetic drug discovery in T-LGL signaling network.	57
Figure S3.1. Analysis results of the HSN network by MORO.	69
Figure S3.2. Correlations between the modularity and robustness of 6,400 random Boolean networks.....	71
Figure S3.3. Changes of module centrality values against the module size in the HSN network.	72

Figure S3.4. Changes of module centrality values against the module size in STKE-shuffled random networks.....	73
Figure S3.5. Changes of module centrality values against the module size in HSN-shuffled random networks.....	74
Figure S3.6. Correlation between module centrality values and in-/out-module robustness in the STKE network.....	75
Figure S3.7. Correlation between module centrality values and in/out-module robustness in the HSN network.....	76
Figure S4.1. Analysis of the changes of the modularity and the robustness by edge-removal mutations in STF signaling network.....	77
Figure S4.2. Analysis of the changes of the modularity and the robustness by edge-removal mutations in HIV-1 signaling network.....	78
Figure S4.3. Analysis of normal distributions of averages of modularity changes and robustness changes in T-LGL network.....	79
Figure S4.4. Analysis of normal distributions of averages of modularity changes and robustness changes in STF network.....	80
Figure S4.5. Analysis of normal distributions of averages of modularity changes and robustness changes in HIV-1 network.....	81
Figure S4.6. Analysis of outliers of averages of modularity changes and robustness changes in T-LGL network.....	82
Figure S4.7. Analysis of outliers of averages of modularity changes and robustness changes in STF network.....	83
Figure S4.8. Analysis of outliers of averages of modularity changes and robustness changes in HIV-1 network.....	84
Figure S4.9. Relationship between the changes of the modularity and the robustness in T-LGL signaling network.....	85
Figure S4.10. Relationship between the changes of the modularity and the robustness in random networks.....	86
Figure S4.11. Relationship of each of the changes of the modularity and the robustness with the structural properties in STF signaling network.....	87
Figure S4.12. Relationship of each of the changes of the modularity and the robustness with the structural properties in HIV-1 signaling network.....	88

Figure S4.13. Relationship of each of the changes of the modularity and the robustness with the structural properties in random networks shuffled from T-LGL network.	89
Figure S4.14. Relationship of each of the changes of the modularity and the robustness with the structural properties in random networks shuffled from STF network.	90
Figure S4.15. Relationship of each of the changes of the modularity and the robustness with the structural properties in random networks shuffled from HIV-1 network.	91
Figure S4.16. Topological distributions of High-MI/High-RD edges and their incident nodes in STF signaling network.	92
Figure S4.17. Topological distributions of High-MI/High-RD edges and their incident nodes in HIV-1 signaling network.	93
Figure S4.18. Edge-removal analysis for edgetic drug discovery in STF signaling network.	94
Figure S4.19. Edge-removal analysis for edgetic drug discovery in HIV-1 signaling network.	95

LIST OF TABLES

Table 3.1. Running time of MORO.....	42
Table 3.2. GO analysis in the HSN network.....	44
Table 4.1. Results of GO analysis between High-MI-incident/High-RD-incident group and the rest of genes in T-LGL signalling network.....	54
Table S4.1. GO analysis results between High-MI-incident/High-RD-incident group and the rest of genes in STF network.	96
Table S4.2. GO analysis results between High-MI-incident/High-RD-incident group and the rest of genes in HIV-1 network.....	97

CHAPTER 1. INTRODUCTION

1.1 Motivation

Network modularity represents the degree to which a network is divided into modules of separate community structures. A highly modularized network has dense connectivity between the nodes within each module but sparse connectivity between the nodes of different modules. Many plugins based on the Cytoscape platform (Shannon, et al., 2003) have been developed for modularity analysis in biological networks. For example, clusterMaker (Morris, et al., 2011) implemented several clustering algorithms such as k-means, k-medoid, SCPS, and AutoSOME to visualize a structure of modules within biological networks. GIANT (Cumbo, et al., 2014) was proposed to investigate topological or functional relationships in a metabolic network by performing a clustering analysis and a functional cartography of nodes. Another well-known plugin is NeMo (Rivera, et al., 2010), which can identify diverse network communities by means of a neighbor-sharing score based on a hierarchical agglomerative clustering method. These plugins have a limitation, though, in that they focus only on the structural analysis of a network and its visualization, without any consideration of dynamics analysis. This restricts their use to undirected networks such as protein–protein networks, or to analysis of directed networks that ignores the direction information.

Herein I note previous studies showing that dynamical behaviors, particularly robustness, of biological networks can be highly affected by their modularity characteristics. For instance, a recent study reported that a modular organization of cancer signaling networks is associated with the patient survivability, which suggests a relationship between modularity and network robustness (Takemoto and Kihara, 2013). Also, the robustness against state perturbations of a human signaling network was negatively correlated to network modularity (Tran and Kwon, 2013). Modular stabilizing in protein–protein interaction networks can be recombined to create highly robust chimeric proteins in evolution (Lin, et al., 2007). It has been also argued that modularity reduces robustness against mutation in metabolic networks (Holme, 2011). Because of the importance of network modularity and robustness, there is a pressing

need to develop a tool that can analyze both simultaneously. Figure 1.1 shows an illustration example of modularity and robustness in signalling network.

Another challenge is that of how to know the changes of modularity and robustness by structural modification in signaling network. Robustness and modularity are key properties to understand complex dynamics in large-scale biological networks. The former means the capability of a network to maintain

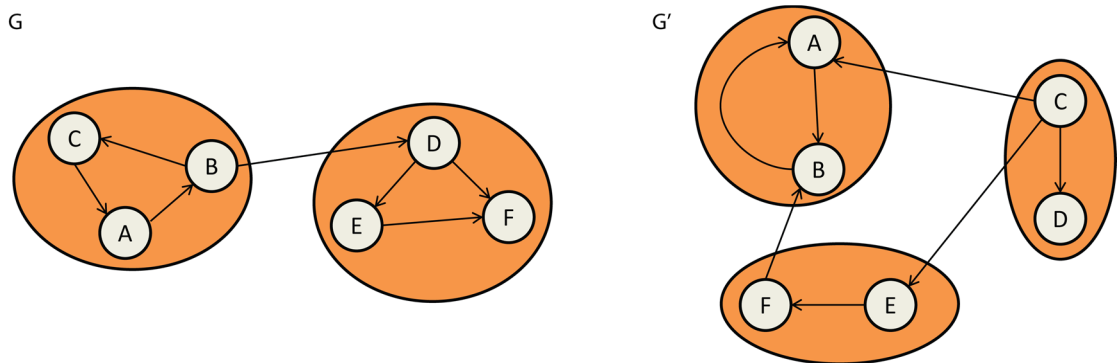


Figure 1.1 An illustrative example of modularity and robustness.

Given two networks G and G' . They have the same number of nodes (6) and edges (7). They also have 2 and 3 modules, respectively. Modules are colored and surrounded by solid line. However, modularity of G (0.35714) is higher than that of G' (0.22449) whereas its robustness of the former (0.23333) is smaller than that of the latter (0.78333). This observation raises the question of whether modularity is correlated with robustness in signaling network.

functioning against external and internal perturbations (Kitano, 2004), and the latter describes the divisibility of a network into clusters (Girvan and Newman, 2002). The robust dynamics (Ingolia, 2004; Little, et al., 1999; Yi, et al., 2000) and the modularized structures (Kreimer, et al., 2008; Lin, et al., 2007; von Dassow and Munro, 1999) have been ubiquitously observed through various biological examples. It is also notable that these properties can be changed by structural mutations because they are highly dependent on the network structure. For example, a few studies showed that the modularity is greatly changed by the removal of hubs (Han, et al., 2004) or by stabilizing events in protein–protein interaction networks. Some other studies also proved that the robustness is considerably changeable according to a variety of mutations (Kaneko, 2007; Le and Kwon, 2013; Paroni, et al., 2016; Trinh and Kwon, 2016). Additionally, there were some previous studies to investigate a relation between the robustness and the modularity. For example, it was shown that

the modularized structure of bone networks improves the robustness compared to a regular network of the same size (Viana, et al., 2009). Some other studies observed that both the robustness and the modularity characteristics could be emergently improved through a network evolution process (Hintze and Adami, 2008; Variano, et al., 2004). Moreover, there were some studies to explicitly examine linear correlations between the robustness and the modularity over differently structured networks (Holme, 2011; Tran and Kwon, 2013; Truong, et al., 2016). In metabolic networks, the robustness against the mutant concentrations of metabolites or the mutant expression of enzymes has increased or decreased, respectively, as the modularity increases (Holme, 2011). On the other hand, the robustness against a gene state perturbation was negatively correlated with the modularity in signaling networks (Tran and Kwon, 2013; Truong, et al., 2016). Although these previous studies found interesting relations between the robustness and the modularity, there are some issues needed to be investigated as follows. The first issue is that there is little known knowledge about changes of the modularity and the robustness. In particular, there was no intensive study about the relationship of the changes of the modularity and the robustness by structural mutations. I note that the previous studies (Holme, 2011; Tran and Kwon, 2013; Truong, et al., 2016) focused on the robustness and the modularity over networks with very different structures, whereas this study focuses on the changes of the robustness and the modularity over mutant networks with a slight structural modification. This means that the findings in the previous studies do not necessarily hold in my analysis. Another interesting issue is whether some well-known motifs are relevant to the changes of the modularity and the robustness or not. In fact, some previous studies have shown that network motifs such as feedback loops (FBLs) and feed-forward loops (FFLs) ubiquitously found in various biological networks can affect the robustness (Kim, et al., 2008; Le and Kwon, 2013). For instance, it was reported that more positive and less negative FBLs are observed in robust networks (Kwon and Cho, 2008). Another study showed that coherent coupling of FBLs is a design principle of a robust signaling network (Kwon and Cho, 2008). It was also reported that coherent FFLs strengthen the robustness against update-rule perturbations (Le and Kwon, 2013). To my best knowledge, even there was no reported motif which is relevant to the modularity property. Taken together, there is little known about motifs which indicate the changes of the modularity, the robustness, or both. The last issue is that there was no previous study to compare sets

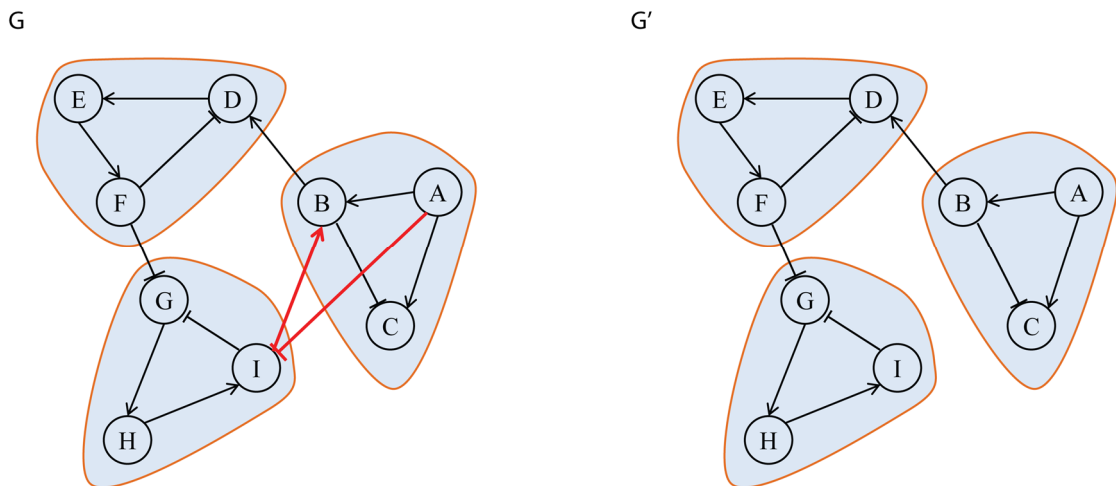


Figure 1.2 An illustrative example of edge-removal mutations.

(a) The original network $G(V,A)$. (b) The mutant network $G'(V,A')$ by removal of $I \rightarrow B$ and $A \rightarrow I$. It was observed that both networks G and G' consist of three modules. Modularity and robustness values in G were 0.35799 and 0.88889, respectively, whereas those in G' were 0.48347 and 0.74444, respectively. Therefore, the changes of the modularity and the robustness were positive (0.12548) and negative (-0.14445), respectively.

of nodes or interactions which efficiently control the changes of the modularity and the robustness. This can be impressive because the result can be used to identify functionally important nodes or interactions such as drug targets. Thus, it is necessary to employ a Boolean network model and modularity measure to investigate the changes of modularity and robustness by edge-removal mutations in signaling networks. Figure 1.2 shows an illustrative example of edge-removal mutations in networks.

1.2 Research objectives

In the first study, I devised a novel Cytoscape app called MORO that can analyze a relationship between dynamical robustness and structural modularity in biological networks represented by directed graphs. In addition, to make it possible to analyze very large-scale networks, I implemented the robustness computation portion of the app as a parallel algorithm by using the OpenCL library. It was also designed to efficiently visualize how the detected modules are located relative to each other. Furthermore, it elucidates analysis results of centrality and gene ontology (GO)

enrichment of modules. Moreover, it provides a batch-mode simulation function to validate whether a result observed in a biological network is consistently conserved in many randomly organized networks. In this study, I tested my app in a case study investigating large-scale signaling networks and observed that modularity and robustness are negatively correlated, similar to previous findings (Tran and Kwon, 2013). It was verified by means of batch-mode simulation that these findings hold in random networks. Moreover, I found some GO terms which are differently enriched between the largest module and the rest of the modules, and it was shown that the module size is positively correlated with five centrality values. In summary, my app can efficiently analyze the relationship between modularity and robustness in large signaling networks

In the second work, I tried to investigate the changes of the modularity and the robustness by edge-removal mutations in signaling networks. Through intensive simulations using a Boolean network model (Graudenzi, et al., 2011; Kauffman, 2004), I first found that both the modularity and the robustness increased on average against edge-removal mutations, but the change of modularity is negatively correlated with the change of robustness. More intriguingly, the modularity change was positively correlated with the degree, the number of FBLs, and the edge betweenness of removed edges, whereas the robustness change was negatively correlated with them. Additionally, I found that these findings are consistently conserved in the random networks. Moreover, I identified two groups of genes which are incident to the highly-modularity-increasing and the highly-robustness-decreasing edges against the edge-removal mutations, respectively, and observed that they are likely to be central by forming a considerably large connected component. The gene-ontology enrichment of each of the gene groups was clearly different from the rest of genes. Finally, I found that the highly-robustness-decreasing edges can be promising edgetic drug-targets. Taken together, the analysis of the changes of the robustness and the modularity against the edge-removal mutations can be useful to reveal novel dynamical characteristics of signaling networks.

1.3 Dissertation outline

This thesis is organized into five chapters. Chapter 1 presents my motivation and also introduces new findings of this work. In Chapter 2, background for my work is

presented such as structural and dynamic properties of biological networks, random Boolean network model, the databases used, and related work to the issues that I addressed. In Chapter 3, and 4, I present more detail about the overview, results, which I dealt with. In particular, Chapter 3 introduces a software application, MORO, which employs an OpenCL library to perform network dynamics calculations and to examine in-/out- module robustness in parallel. Chapter 4 shows my new findings in investigating changes of modularity and robustness by edge-removal mutations in signaling network. Chapter 5 summarizes my main findings and also offered some future work.

CHAPTER 2. BACKGROUND

2.1 Biological networks

2.1.1 Introduction

Biological network is any network that exists in biological systems ranging from food webs to various biomedical networks in molecular biology. The variety of interactions between components in a cell such as genes, proteins and metabolites makes the biological networks complex and diverse. However, most studies have been recently made on molecular biological networks, including protein-protein interaction (PPI), gene regulatory and metabolic networks. In which, nodes in PPI networks are proteins that are connected to each other by physical interactions (Rual, et al., 2005; Stelzl, et al., 2005); gene regulatory networks whose genes or transcription factors are connected if the expression of one gene modulates expression

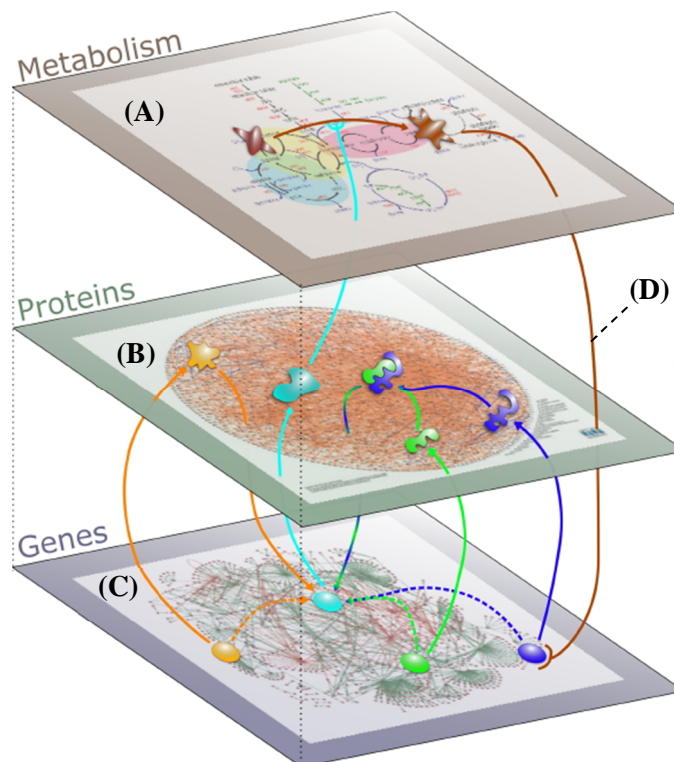


Figure 2.1 – Four kinds of biological networks.

(A) Metabolic network. (B) Protein network. (C) Gene regulatory network. (D) Signaling network – the network of communication between cellular components (i.e., gene, protein and metabolite). This demonstration is from CEA Sciences (<http://ceasciences.fr/>).

of another one by either activation or inhibition (Carninci, et al., 2005) and metabolic networks whose metabolic products and substrates that participate in one reaction (Jeong, et al., 2000). In addition to the three above networks, signaling network is a network of communication between the components that control and coordinate basic activities of cell (Jordan, et al., 2000). All of my studies were conducted on the signaling networks.

2.1.2 Datasets of signaling networks

In the first study, I tested MORO with two large-scale signalling networks, the canonical cell signaling network (STKE; <http://stke.sciencemag.org>) and the human signal transduction network (HSN; <http://www.bri.nrc.ca/wang>) which consist of 754 proteins and 1,624 interactions, and 5,443 genes and 37,663 interactions, respectively.

In the second work, to investigate real signaling networks, I used three datasets of signaling networks: a T-LGL survival network (T-LGL) (Saadatpour, et al., 2011) consisting of 60 genes and 142 interactions, a signal transduction network in fibroblasts (STF) (Hirabayashi, et al., 2004) consisting of 139 genes and 557 interactions, and a HIV-1 interaction network in T-cell (HIV-1) (Oyeyemi, et al., 2015) consisting of 138 genes and 368 interactions collected by manually curating signaling pathways from cellcollective (www.cellcollective.org) (Helikar, et al., 2012).

2.2 Random network generation

To validate that the findings in real signaling networks are general principles, I extensively simulated randomly structured networks generated by five models: Barabási-Albert (BA) model (Barabási and Albert, 1999), Erdős-Rényi (ER) model (Erdős and Rényi, 1959), an Erdős-Rényi variant model (Le and Kwon, 2011) and two shuffling models. Actually, all of them have been widely used to investigate biological networks (Kwon and Cho, 2008; Le and Kwon, 2013; Maslov and Sneppen, 2002; Shen-Orr, et al., 2002).

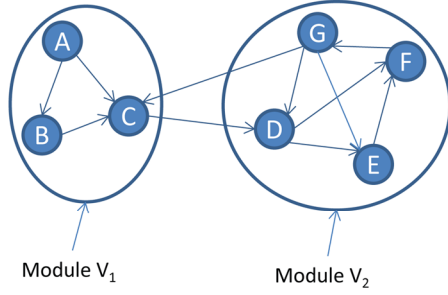
The BA model uses a preferential attachment scheme, which is a type of network growth model, as follows. The desirable number of nodes (N), the number of nodes of a seed network (e), and the number of interactions that should be added at each iteration (d) are given as parameters. A small seed network $G(V, A)$ is then created,

where $V = \{v_1, v_2, \dots, v_e\}$ and $A = \{(v_i, v_j) \mid i, j = 1, 2, \dots, e, i \neq j\}$, i.e., a complete network. At each iteration, a new node v is added to V . Then, d different interactions that individually connect v and $v' \in V \setminus \{v\}$ are newly added to A , where v' is determined with a probability proportional to the connectivity of v' (the connectivity of a node is defined as the number of interactions incident to the node), and both the direction and sign of the added interactions are specified uniformly at random. This iteration process is repeated until $|V|=N$.

In the ER model, the desirable number of nodes (N) and a probability (p) are given as parameters. The decision whether to create an interaction from an arbitrary node v to another arbitrary node v' is then independently determined with a probability p . I also use a variant of the ER model where the desirable numbers of nodes (N) and interactions (E) are given as parameters. An RBN is then generated in such a way that E different interactions are chosen uniformly at random out of $N \times (N-1)$ possible candidates.

Moreover, I implemented two shuffling techniques where a reference network should be given. The first shuffling technique creates random networks by shuffling the direction and the sign of every interaction from the reference network (Shuffle I). More specifically, each directed link denoted by (v_i, v_j, τ) where v_i , v_j , and τ denote a starting node, an ending node, and the sign of the link, respectively, is replaced by one of (v_i, v_j, τ) , $(v_i, v_j, -\tau)$, (v_j, v_i, τ) , and $(v_j, v_i, -\tau)$ uniformly at random (Le and Kwon, 2013). On the other hand, the other shuffling technique creates random networks by rewiring the edges of the reference network such that the in-degree and the out-degree of all nodes are conserved (Shuffle II) (Maslov and Sneppen, 2002; Maslov, et al., 2002). More specifically, a pair of directed links (v_a, v_b, τ_{ab}) and (v_c, v_d, τ_{cd}) such that there is no link from v_a to v_d and from v_c to v_b is randomly selected, and the pair is replaced by a new pair of links (v_a, v_d, τ_{ab}) and (v_c, v_b, τ_{cd}) . In the tool, the number of rewirings is set to the multiplication of the value of the "Shuffling intensity" parameter and the number of edges of the reference network. I note that the shuffling models generate random networks whose structure is more similar to the reference network than BA, ER, and ER-variant models because the degree distribution is conserved.

➤ Given a directed graph $G(V, E)$



$$M(P) = \sum_{i=1}^M \left(\frac{w_{V_i}}{\omega} - \frac{w_{V_i}^{in} w_{V_i}^{out}}{\omega^2} \right) \in [0, 1]$$

$$M(G) = \max_P M(P)$$

- $P = \{V_1, V_2\}$
- *Module V_1 :*
- $w_{V_1}: \{A \rightarrow B, A \rightarrow C, B \rightarrow C\}$
- $w_{V_1}^{out}:: \{C \rightarrow D\}$
- $w_{V_1}^{in}:: \{G \rightarrow C\}$

- *Module V_2 :*
- $w_{V_2}: \{D \rightarrow E, E \rightarrow F, F \rightarrow G, G \rightarrow D, D \rightarrow F, G \rightarrow E\}$
- $w_{V_2}^{out}:: \{G \rightarrow C\}$
- $w_{V_2}^{in}:: \{C \rightarrow D\}$

- $M(P) = \left(\frac{3}{11} - \frac{1*1}{11^2} \right) + \left(\frac{6}{11} - \frac{1*1}{11^2} \right) = \mathbf{0.80165}$

Figure 2.2 An illustrative example of calculating modularity.

Given graph G . It consists of two modules, V_1 and V_2 . The number of interactions whose starting and ending nodes are both included in module V_i , $\omega_{V_i}^{out}$ and $\omega_{V_i}^{in}$ are the numbers of interactions whose starting or ending nodes only, respectively, are included in module V_i , and ω is the total number of interactions in the network. Finally, modularity values can be obtained by applying the optimization algorithm.

2.3 Network modularity

Given a network represented by a directed graph $G(V, A)$ where V and A are the sets of nodes and interactions, respectively, I employ the modularity measure introduced in a previous study (Leicht and Newman, 2008). A partition $P = \{V_1, V_2, \dots, V_M\}$ of V is a set of nonempty disjoint subsets of V that covers V (i.e. $V_i \cap V_j = \emptyset$ for all $i, j \in \{1, 2, \dots, M\}$ and $i \neq j$, and $\cup_{i=1}^M V_i = V$). Then, the modularity of the partition $M(P)$ is defined as $M(P) = \sum_{i=1}^M \left(\frac{\omega_{V_i V_i}}{\omega} - \frac{\omega_{V_i}^{in} \omega_{V_i}^{out}}{\omega^2} \right)$, where $\omega_{V_i V_i}$ is the number of interactions whose starting and ending nodes are both included in module V_i , $\omega_{V_i}^{out}$ and $\omega_{V_i}^{in}$ are the numbers of interactions whose starting or ending nodes only, respectively, are included in module V_i , and ω is the total number of interactions in the network. Then, the modularity of the network is defined as $M(G) = \max_P M(P)$. However, it is difficult to obtain the optimal partition. In my studies, the modularity value of a network is averaged over 30 trials by using an optimization algorithm

proposed in a previous study (Noack, 2009). Figure 2.3 shows an example how to calculate modularity.

2.4 Boolean network model

2.4.1 Introduction

In order to analyze the network dynamics, I employed a Boolean network model that has been frequently used to investigate the complex dynamics of biological networks (Campbell and Albert, 2014; Steinway, et al., 2015). A Boolean network is represented by a directed graph $G(V, A)$, where V is a set of Boolean variables and A is a set of ordered pairs of the Boolean variables called directed links. Each $v_i \in V$ has a value of 1 (“on”) or 0 (“off”), which represents the possible state of the corresponding gene, and a state of a network G is defined as a vector of the states of all nodes. A directed link $e(v_j, v_k) \in A$ has a positive (“activating”) or negative (“inhibiting”) relationship from v_j to v_k . The value of each variable v_i at time $t + 1$ is determined by the values of k_i other variables $v_{i_1}, v_{i_2}, \dots, v_{i_{k_i}}$ with a link to v_i at time t by a Boolean function $f_i: \{0,1\}^{k_i} \rightarrow \{0,1\}$; all variables are synchronously updated. Hence, the update rule can be written as $v_i(t + 1) = f_i(v_{i_1}(t), v_{i_2}(t), \dots, v_{i_{k_i}}(t))$. Here, I employed a nested canalyzing function (NCF) model (Kauffman, et al., 2003) (see Supporting Text A1 in Appendix A for details) to represent an update rule f_i as follows:

$$v_i(t + 1) = \begin{cases} O_1 & \text{if } v_{i_1}(t) = I_1 \\ O_2 & \text{if } v_{i_1}(t) \neq I_1 \text{ and } v_{i_2}(t) = I_2 \\ O_3 & \text{if } v_{i_1}(t) \neq I_1 \text{ and } v_{i_2}(t) \neq I_2 \text{ and } v_{i_3}(t) = I_3 \\ & \vdots \\ O_{k_i} & \text{if } v_{i_1}(t) \neq I_1 \text{ and } \dots \text{ and } v_{i_{k_i-1}}(t) \neq I_{k_i-1} \text{ and } v_{i_{k_i}}(t) = I_{k_i} \\ 1 - O_{k_i} & \text{otherwise} \end{cases} \quad (1)$$

where I_m and O_m ($m = 1, 2, \dots, k_i$) are called canalyzing and canalyzed Boolean values, respectively. The NCF model can generate various canalyzing rules which are ubiquitously found in molecular interactions (Harris, et al., 2002). It was also shown that NCFs properly fit the experimental data obtained from a literature (Kauffman, et

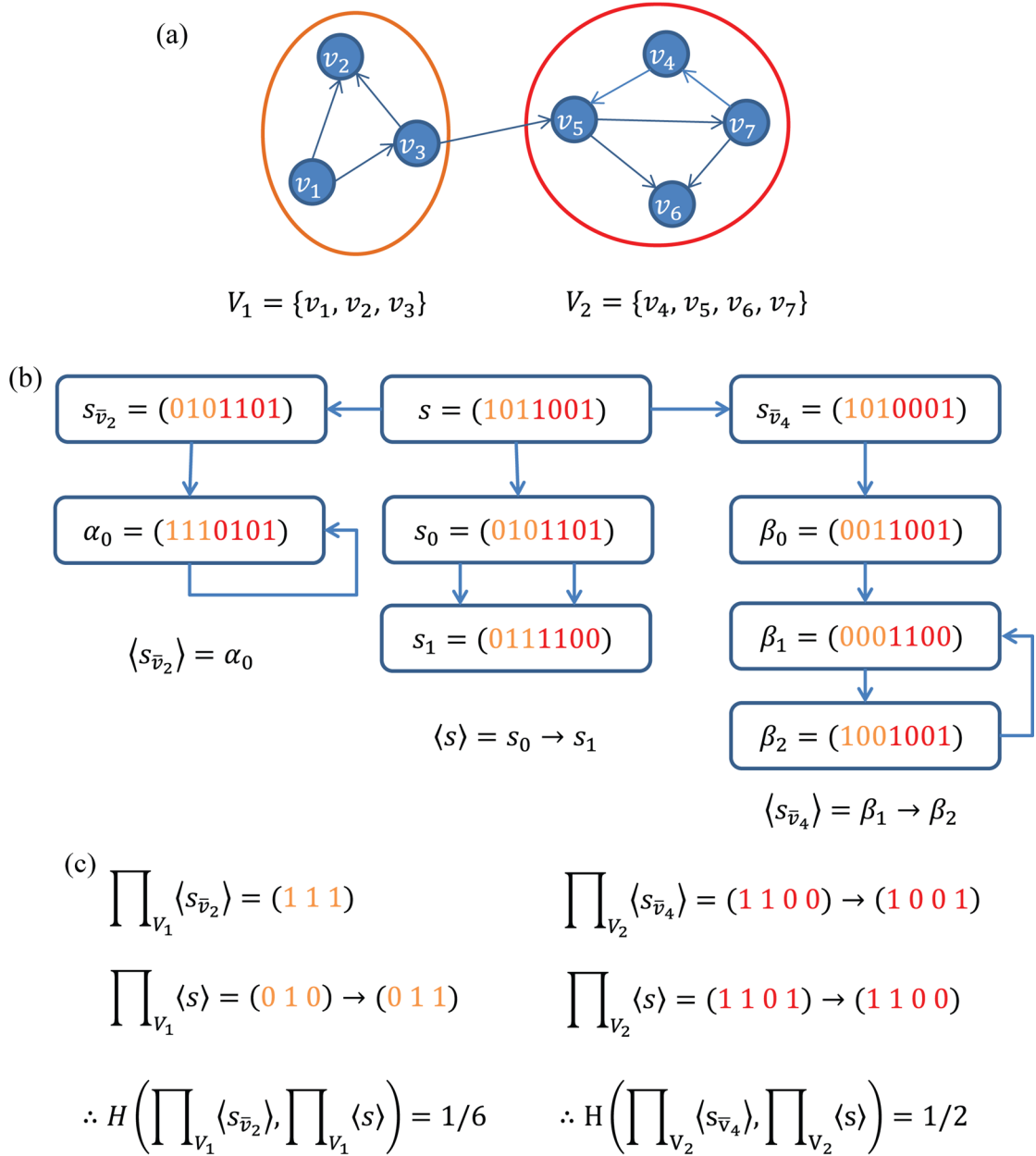


Figure 2.3 – An illustrative example of calculating the attractor similarity.

(a) A given network consists of two modules. (b) An example of analysis of attractor. (c) The result of calculating attractor similarity by using Hamming distance.

al., 2003), and many logical interaction rules inferred from gene expression data can be represented by NCFs (Harris, et al., 2002; Naldi, et al., 2010). For example, 133 out of 139 rules compiled from a dataset about a transcriptional regulatory network (Harris, et al., 2002) and 39 out of 42 rules inferred from a dataset about signaling pathways (Naldi, et al., 2010) were NCFs. These imply that NCFs-embedded random networks can describe the network dynamics considerably similar to that of real

biological networks. In this study, each NCF is randomized by specifying all I_m s and O_m s between 0 and 1 uniformly at random.

2.4.2 Boolean network dynamics and in-/out-module robustness against initial states perturbation

In this model, each state $s(t) = (v_1(t), v_2(t), \dots, v_N(t))$ at time t transits to the next state $s(t+1)$ according to the set of update rules $F = \{f_1, f_2, \dots, f_N\}$, i.e., $s(t+1) = F(s(t))$, where I randomly choose either a logical conjunction or disjunction for f_i with a uniform probability distribution. For instance, if a Boolean variable v has a positive relationship from v_1 , a negative relationship from v_2 and a positive relationship from v_3 , then the conjunction and disjunction update rules are $v(t+1) = v_1(t) \wedge \bar{v}_2(t) \wedge v_3(t)$ and $v(t+1) = v_1(t) \vee \bar{v}_2(t) \vee v_3(t)$, respectively. In the case of the conjunction, the value of v at time $t+1$ is 1 only if the values of v_1, v_2 and v_3 at time t are 1, 0 and 1, respectively. A state of G is defined as a vector of values v_1 through v_N . A state trajectory starts from an initial state $s(0)$ and eventually converges to either a fixed-point or limit-cycle attractor. Because these attractors can represent diverse biological network behaviors such as multistability, homeostasis, and oscillation, a change in the converging attractor can be interpreted as a loss of robustness. I denote the attractor converged to starting from an initial state $s(0)$ by $\langle s \rangle$. The network is considered to be robust against mutation at v_i if $\langle s \rangle$ is equal to $\langle s_{\bar{v}_i} \rangle$, where $\bar{v}_i (= \neg v_i)$ indicates the state perturbation of s subjected to v_i . This concept to measure robustness has been widely used (Ciliberti, et al., 2007; Kitano, 2004; Kwon and Cho, 2008). More specifically, the robustness of a network $\gamma(G)$ is defined as follows:

$$\gamma(G) = \frac{1}{N \cdot |S|} \sum_{s \in S} \sum_{i=1}^N I(\langle s \rangle = \langle s_{\bar{v}_i} \rangle),$$

where S is the set of whole states (i.e. $S = 2^N$), and $I(\cdot)$ is an indicator function. Because $|S|$ is a very large number, I used a sample subset $\tilde{S} \subseteq S$ with $|\tilde{S}| = 2N$ instead of S to calculate $\gamma(G)$. Given a partition $P = \{V_1, V_2, \dots, V_M\}$, I employed the in-module and out-module robustness of a module V_i , $\gamma_{in}(V_i)$ and $\gamma_{out}(V_i)$, respectively, defined in (Tran and Kwon, 2013) as follows:

$$\gamma_{in}(V_i) = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \sum_{v \in V_i} \frac{H(\Pi_{V_i}\langle s \rangle, \Pi_{V_i}\langle s_{\bar{v}} \rangle)}{|V_i|}$$

and

$$\gamma_{out}(V_i) = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \sum_{v \in V_i} \frac{H(\Pi_{V \setminus V_i}\langle s \rangle, \Pi_{V \setminus V_i}\langle s_{\bar{v}} \rangle)}{|V_i|},$$

where $\Pi_{V_i}\langle s \rangle$ represents a projection operator to extract the partial attractor of a given subset $V_i \subseteq V$ from an attractor $\langle s \rangle$, and $H(\langle s \rangle, \langle s' \rangle)$ denotes a similarity measure between two attractors $\langle s \rangle$ and $\langle s' \rangle$. More particularly, given $\langle s \rangle = s_0 \rightarrow s_1 \rightarrow \dots \rightarrow s_{l-1}$ and $\langle s' \rangle = s'_0 \rightarrow s'_1 \rightarrow \dots \rightarrow s'_{l'-1}$ ($1 \leq l \leq l'$ is assumed without loss of generality), $H(\langle s \rangle, \langle s' \rangle)$ is defined as follows:

$$H(\langle s \rangle, \langle s' \rangle) = \frac{1}{l'} \sum_{j=0}^{l-1} \left(1 - \frac{h(s_j, s'_j)}{N} \right)$$

where h is the Hamming distance (i.e. the number of different bits between two binary sequences). Then, the in-module and out-module robustness of a network, $\gamma_{in}(V_i)$ and $\gamma_{out}(V_i)$, respectively, are defined as follows:

$$\gamma_{in}(G) = \frac{1}{M} \sum_{i=1}^M \gamma_{in}(V_i)$$

and

$$\gamma_{out}(G) = \frac{1}{M} \sum_{i=1}^M \gamma_{out}(V_i)$$

Figure 2.3 shows an example of calculating the attractor similarity. As shown in the figure, in-/out-module robustness of network can be calculated after all nodes have initial states mutation.

2.4.3 Boolean network dynamics against update-rule mutation

Let $G(V, A)$ a Boolean network with a list of update-rules $F = \{f_1, f_2, \dots, f_N\}$. Every initial state converges to an attractor which can describe diverse network dynamics such as multi-stability, homeostasis, and oscillation (Bhalla, et al., 2002; Pomerening, et al., 2003). Let $\alpha(s, G, F)$ the attractor which the initial state s converged. The network is considered as robust against a perturbation at v_i if the

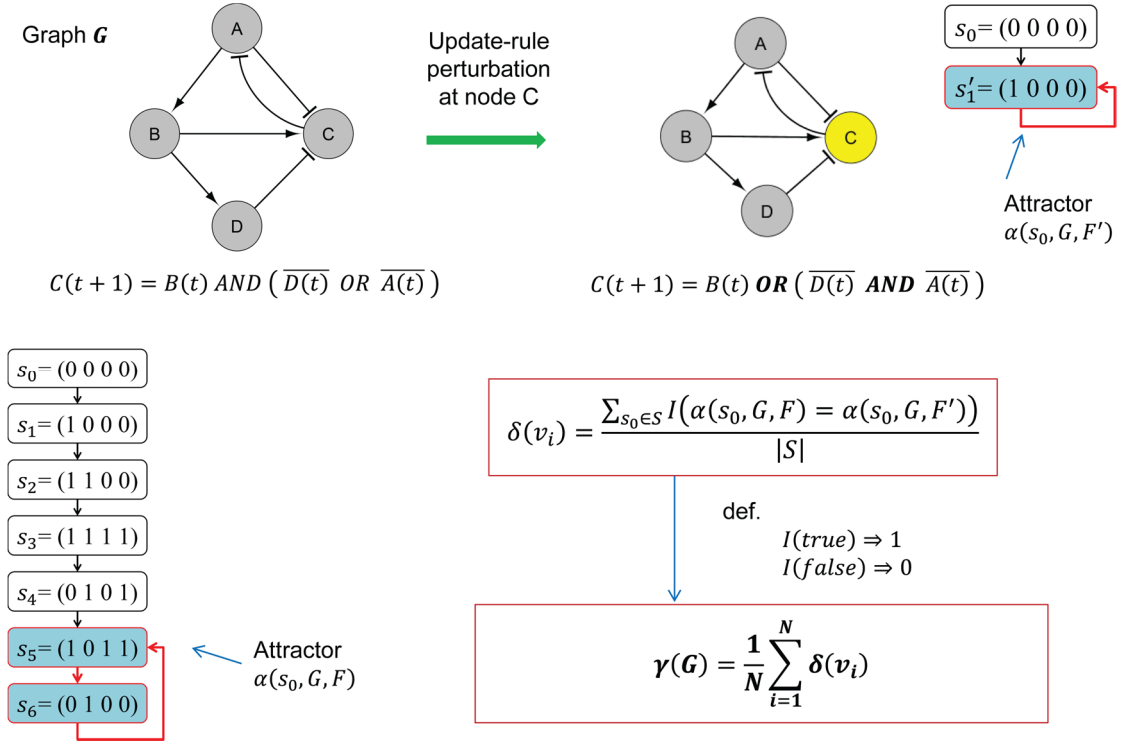


Figure 2.4 An illustrative example of calculating network dynamics against update-rule mutation.

The original network G (left), the same network with a update-rule mutation on node C (right). The arrows and bar-headed lines represent positive and negative interactions, respectively. The network state is represented by a vector of values of four Boolean variables in the sequence of (A B C D). For a same initial state $s = (0\ 0\ 0\ 0)$, the mutated networks converges to a different fixed-point attractors compared to the limit-cycle attractor of the original network.

attractor is conserved and I herein considered an update-rule mutation which describes a scenario that F is changed to $F'_i = \{f_1, \dots, f'_i, \dots, f_N\}$, where f'_i means that every canalyzing and canalyzed values were flipped (i.e., all I_m and O_m are changed into $1 - I_m$ and $1 - O_m$, respectively). This update-rule mutation may represent a deleterious change in the function of a protein or gene (Ng and Henikoff, 2003), and have been used in a previous study (Le and Kwon, 2013). Then the network robustness $\gamma(G)$ is defined as follows:

$$\gamma(G) = \frac{1}{N|S|} \sum_{s \in S} \sum_{i=1}^N I(\alpha(s, G, F) = \alpha(s, G, F'_i)),$$

where S is a set of initial states (i.e. $S = 2^N$), and $I(\cdot)$ is a function which outputs 1 or 0 if the condition is met or not, respectively. Because $|S|$ is a very large number, I

used a sample subset $\tilde{S} \subseteq S$ with $|\tilde{S}| = 2N$ instead of S to calculate $\gamma(G)$. Figure 2.4 shows the example how to calculate network robustness against update-rule mutation.

2.5 Structural properties of network

2.5.1 Feedback loops

Feedback loops (FBL), a circular chain of relationships, plays an important role in the dynamic behaviors of cellular signaling networks (Ananthasubramaniam and Herzog, 2014; Kwon, et al., 2007). Given a network, an FBL is a closed simple cycle in which all nodes except the starting and ending nodes, are not revisited. More specifically, $v_0 \rightarrow v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_{L-1} \rightarrow v_L$ is an FBL of length $L (\geq 1)$ if there are links from v_{i-1} to v_i ($i = 1, 2, \dots, L$) with $v_0 = v_L$ and $v_j \neq v_k$ for $j, k \in \{0, 1, \dots, L-1\}$ and $j \neq k$. The number of FBLs of a network element c (a node or an edge) denoted as $NuFBL(c)$ is the number of different FBLs involving c .

2.5.2 Centrality

Previous studies have shown that the structural centrality properties of genes/interactions in biological networks can be strongly related to their importance: the more central a node/edge is, the more functionally important it may be. A brief introduction of the most well-known structural centrality measures such as degree, betweenness, stress, closeness, and eigenvector follows.

- *Degree (DEG)* of a node denotes the number of neighbor nodes that are linked with it in a network. This is a local structural measure which considers only the immediate neighborhoods. Based on this notion, DEG of an edge is similarly defined as the sum of the degrees of both end nodes of the edge.
- *Betweenness (BEW)* quantifies the ability of a protein to monitor communication between other proteins through shortest paths (Freeman, 1977). More specifically, it is defined as follows:

$$BEW(v) = \sum_{u \neq w \in V \setminus \{v\}} \frac{\sigma_{uw}(v)}{\sigma_{uw}} \quad (5)$$

where σ_{uw} denotes the number of shortest paths between u and w , and $\sigma_{uw}(v)$ denotes the one such that v is passed through.

- *Edge Betweenness (EBEW)* is defined as the relative number of shortest paths between pairs of nodes that run along an edge (Girvan and Newman, 2002), similar to Betweenness of a node (Freeman, 1977). EBEW has been used as an important edge centrality measure of a network in previous studies (de Reus, et al., 2014; Zhang, et al., 2014).
- *Stress (STR)* is based on the enumeration of shortest paths (Shimbel, 1953), and is similar to Betweenness; however, instead of summing up the relative number of shortest paths for each pair of proteins, stress counts the absolute number of shortest paths. This gives an approximation of the amount of ‘work’ or ‘stress’ the protein has to sustain in the network:

$$STR(v) = \sum_{u \neq w \in V \setminus \{v\}} \sigma_{uw}(v). \quad (6)$$

- *Closeness (CLO)* uses the sum of the minimal distances from a protein to all other proteins. The closeness measure is defined as the reciprocal of this sum:

$$CLO(v) = \frac{1}{\sum_{u \in V \setminus \{v\}} dist(v, u)} \quad (7)$$

where $dist(v, u)$ denotes the length of the shortest paths between v and u .

- *Eigenvector (EIG)* is defined as the principal eigenvector of the adjacency matrix, D , of the network. It simulates a mechanism in which each node affects all of its neighbors simultaneously. Given the adjacency matrix D , the eigenvector (e) and eigenvalue (λ) are obtained via the equation $\lambda e = De$. Let e_1 be the eigenvector corresponding to the largest (principal) eigenvalue. Then, the eigenvector-based centrality of a protein can be denoted by the i th component of e_1 :

$$EIG(v) = e_1(v). \quad (8)$$

2.6 Related works

2.6.1 Cytoscape Plugins

There have been many softwares and tools introduced for bioinformatics (Brazas, et al., 2011). A set of them mainly focused on visualization (Suderman and Hallett, 2007) or modeling (Alves, et al., 2006) biological networks. Among them, Cytoscape, a free open-source software platform, is a state-of-the-art tool, which was offered at the first time for integrated models of biomolecular interaction networks (Shannon, et al., 2003). It was extended with new features of data integration and network visualization (Smoot, et al., 2011) or Cytoscape Web (Lopes, et al., 2010). One of the interesting features is extendibility by adding novel plugins which are usually implemented by developers for particular tasks. An existence of a large amount of plugins makes Cytoscape become a very powerful tool, which is not only for data integration and network visualization, but also for data analysis. Interoperation of

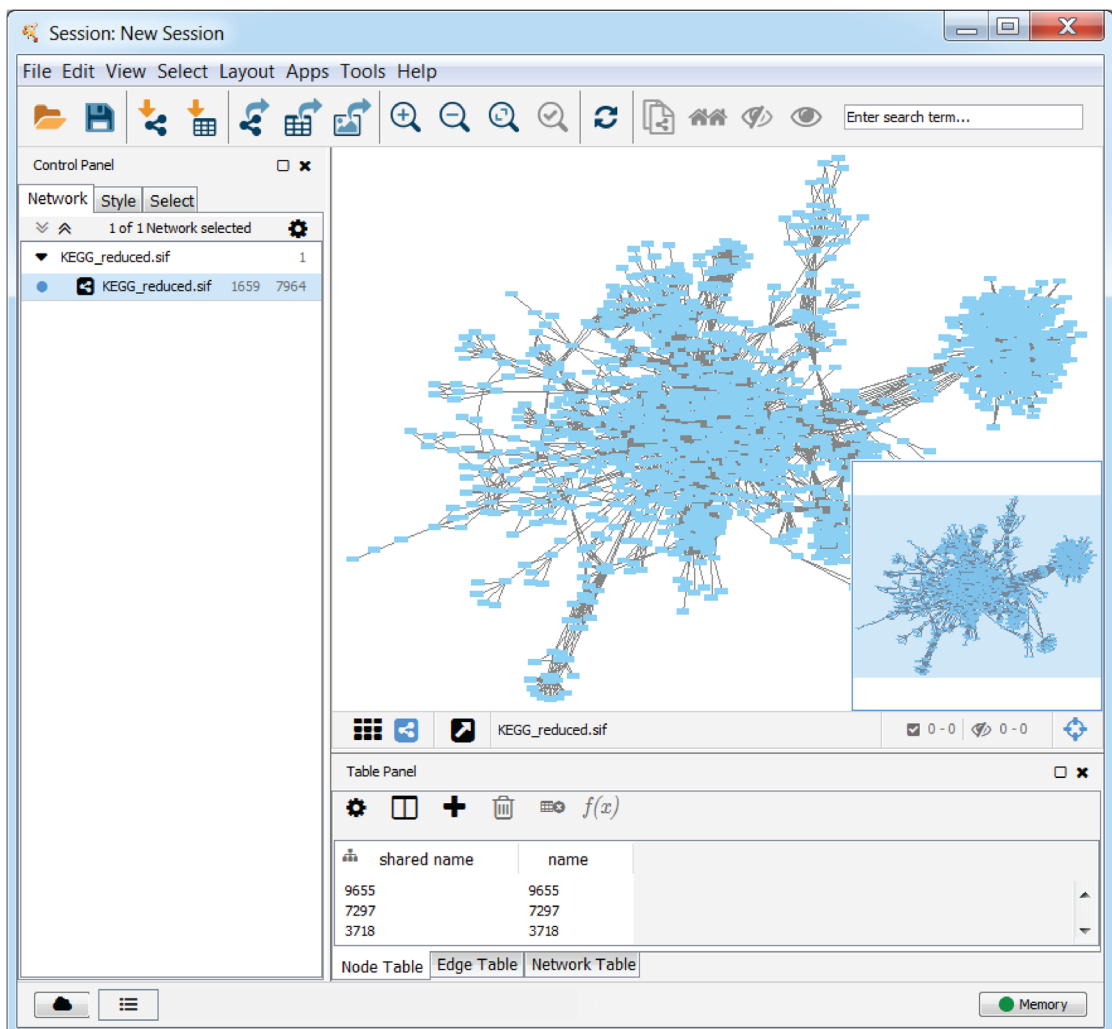


Figure 2.5 – Cytoscape, an environment for data integration, network analysis and visualization.

these plugins also makes Cytoscape become an entire solution for some

bioinformatics problems (Cline, et al., 2007). Figure 2.5 shows the interface of Cytoscape.

CHAPTER 3. MORO: A CYTOSCAPE APP FOR RELATIONSHIP ANALYSIS BETWEEN MODULARITY AND ROBUSTNESS IN LARGER-SCALE SIGNALING NETWORKS

3.1 Overview

Many plugins based on the Cytoscape platform have been developed for modularity analysis in biological networks such as clusterMaker (Morris, et al., 2011), Moduland (Szalay-Bekő, et al., 2012), NCMine (Tadaka and Kinoshita, 2016), PEPPER (Winterhalter, et al., 2014), GIANT (Cumbo, et al., 2014), and NeMo (Rivera, et al., 2010). However, they focus only on the structural analysis of a network and its visualization. These plugins have a limitation, though, in that they focus only on the structural analysis of a network and its visualization, without any consideration of dynamics analysis. This restricts their use to undirected networks such as protein–protein networks, or to analysis of directed networks that ignores the direction information.

Previous studies showing that dynamical behaviors, particularly robustness, of biological networks can be highly affected by their modularity characteristics. For instance, a recent study reported that a modular organization of cancer signaling networks is associated with the patient survivability, which suggests a relationship between modularity and network robustness (Takemoto and Kihara, 2013). Also, the robustness against state perturbations of a human signaling network was negatively correlated to network modularity (Tran and Kwon, 2013). Therefore, I devised a novel Cytoscape app called MORO that can analyze a relationship between dynamical robustness and structural modularity in biological networks represented by directed graphs. In addition, to make it possible to analyze very large-scale networks, I implemented the robustness computation portion of the app as a parallel algorithm by using the OpenCL library. It was also designed to efficiently visualize how the detected modules are located relative to each other. Furthermore, it elucidates analysis results of centrality and gene ontology (GO) enrichment of modules. Moreover, it provides a batch-mode simulation function to validate whether a result observed in a

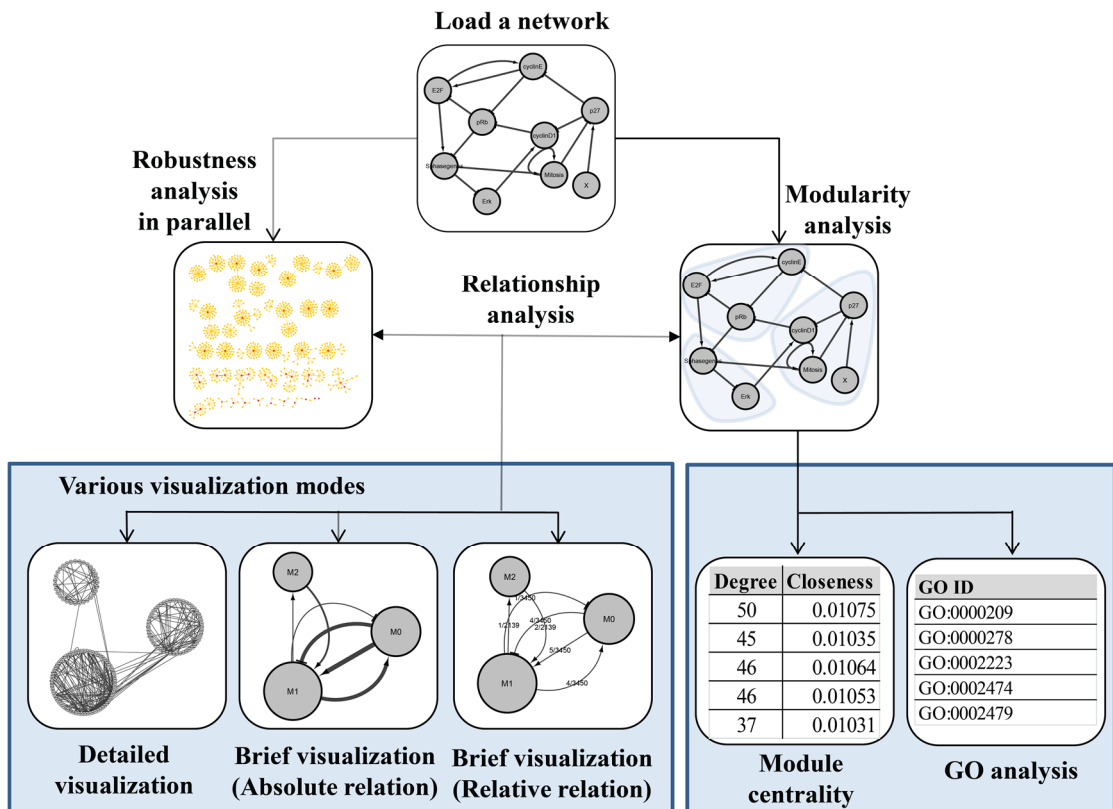


Figure 3.1 – The overall process to analyze the relationship between the network robustness and modularity in MORO.

After a directed network is loaded for analysis, the network modularity and robustness are calculated. In particular, the time consuming part is processed in parallel by using multi-core CPU or GPU. The analysis result can be checked by three types of visualization modes and a summary table. The centrality values and GO analysis of modules are additionally provided.

biological network is consistently conserved in many randomly organized networks. Moreover, I found some GO terms which are differently enriched between the largest module and the rest of the modules, and it was shown that the module size is positively correlated with five centrality values.

3.2 Implementation

3.2.1 The Overall process of MORO App

Figure 3.1 illustrates the main process of my app. First, a directed network is loaded for analysis. Next, the app computes the modularity and robustness of the network. In particular, the robustness algorithm was implemented in parallel computation by using the OpenCL library. The results can be visualized in three modes: a detailed visualization mode, a brief visualization with absolute relations, and

a brief visualization with relative relations. Also, the results can be summarized in tables that include centrality and gene ontology analyses. Details of this process are given in the following subsections.

3.2.2 Parallel computation of robustness

In my app, I employed a Boolean network model to compute robustness. In particular, I further calculated in-module and out-module robustness which represent how much the module subject to a perturbation and the groups of other modules, respectively, are robust against the perturbation. Unfortunately, it is very time-consuming to compute robustness. To reduce the running time, I implemented the robustness calculation part of the app as a parallel algorithm by using the OpenCL library (see Appendix B for the pseudo-code).

3.3 A batch-mode simulation on random Boolean networks

I developed a function for a batch-mode simulation on random Boolean networks (RBNs) to examine if a finding in biological networks holds in RBNs or not similarly in a previous study (Campbell and Albert, 2014; Kwon and Cho, 2008; Kwon and Cho, 2008; Kwon, et al., 2007; Le and Kwon, 2011; Le and Kwon, 2013; Trinh and Kwon, 2015; Trinh, et al., 2014). The batch-mode simulation requires two steps for configuring parameters. The first step is to select an RBN generation model from among five models: Barabási-Albert (BA) model (Barabási and Albert, 1999), Erdős-Rényi (ER) model (Erdős and Rényi, 1959), an Erdős-Rényi variant model (Le and Kwon, 2011) and two shuffling models (Le and Kwon, 2013; Maslov and Sneppen, 2002; Maslov, et al., 2002), and the second step is to set the number of considered initial-states and the type of update-rule schemes (see the subsection “Robustness dynamics in a Boolean network model” for details). Once computations of modularity and robustness are completed, all results are saved in a resulting file, “net_based_result.txt” which describes modularity and robustness results of each RBN (see Supporting Text A2 in Appendix A for details).

3.4 Visualization of relations between modules

My app provides three types of visualizations to show the relationship between modules. The first type is a detailed visualization mode in which all nodes and interactions of the loaded network are shown and the nodes are grouped into modules

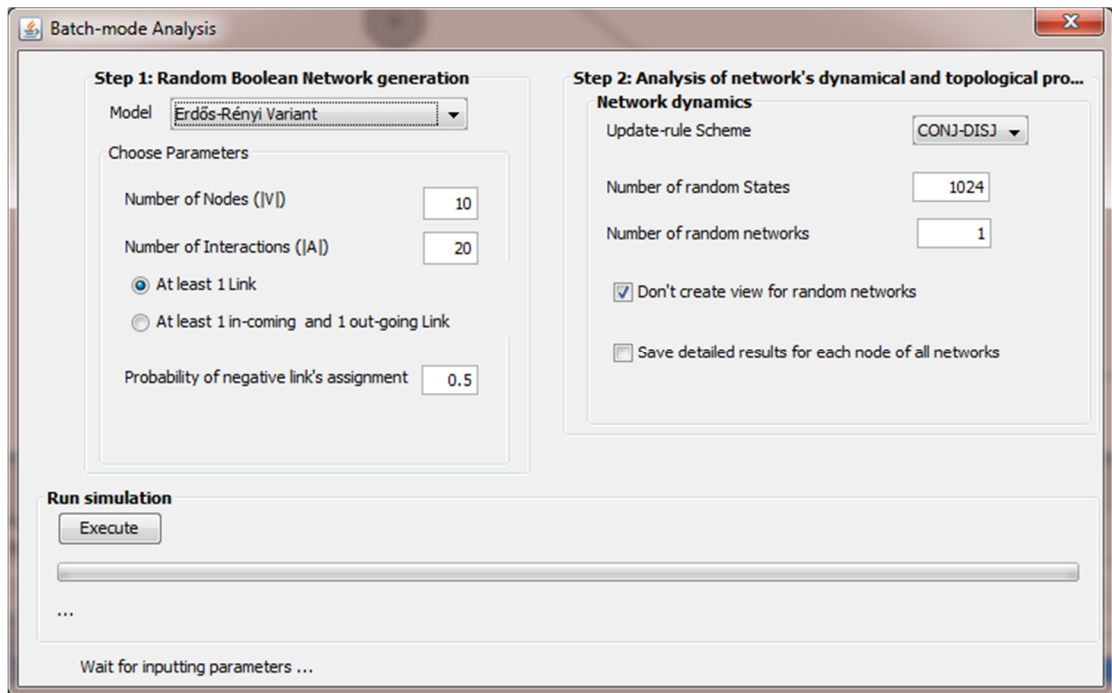


Figure 3.2 – User interface for a batch-mode simulation on RBNs.

There are two steps for configuring parameters of the batch-mode simulation: selecting an RBN generation model, setting the number of considered initial-states and the type of update-rule schemes.

placed by using the Cytoscape group attributes layout. The second type is a brief visualization mode with absolute relations, in which a group node corresponds to a detected module and the weight of a link between group nodes denotes the number of interactions between a pair of modules. The last mode is the same as the second mode except that the weight of a link denotes the ratio of the number of interactions between a pair of modules to the maximal possible number of interactions between them, that is $w/(n_1n_2)$, where w is the number of actual interactions between the pair of modules, and n_1 and n_2 are the numbers of nodes included in each of the modules.

3.5 Module centrality and GO analysis

Many previous studies have shown that the centrality properties of genes/proteins in biological networks are strongly related to their functional roles in a topological or dynamical sense. To extend this concept to module-based centrality analysis, I implemented a function to examine five centrality measures including degree (Jeong, et al., 2001), closeness (Wuchty and Stadler, 2003), betweenness (Freeman, 1977), stress (Shimbel, 1953) and eigenvector (Bonacich, 1987) of modules (See centrality

section of CHAPTER 2 for more detail). Besides, I developed a GO analysis function to compare the functional difference between two groups of modules. To this end, I first identify two groups of genes by selecting some modules of interest. Then, Entrez gene id is mapped to UniProtKB by utilizing the web service at <http://www.uniprot.org/> (Consortium, 2015), and all relevant GO terms are extracted by using the web service at <http://www.ebi.ac.uk/QuickGO/> (Binns, et al., 2009). Finally, GO terms which are most differently enriched between the two gene groups are listed in a table or exported into a text file.

3.6 Results

3.6.1 Analysis of modularity and robustness

The analysis and visualization results of the STKE and HSN networks are shown in Figure 3.3 and Figure S3.1 in Appendix C, respectively. In particular, Figure 3.3(a) and Figure S3.1(a) (in Appendix C) explain various summarized results including the number of modules, modularity, robustness, in-/out-module robustness, and centrality values. Specifically, the number of modules were 16 and 22, the modularity values were 0.72825 and 0.54534, and the robustness values were 0.67721 and 0.75400 in the STKE and HSN networks, respectively. By selecting the visualization option, I can observe the relation between the detected modules in three different modes: a detailed mode (Figure 3.3(b) and Figure S3.1(b) in Appendix C), a brief mode with absolute relations (Figure 3.3(c) and Figure S3.1(c) in Appendix C), and a brief mode with relative relations (Figure 3.3(d) and Figure S3.1(d) in Appendix C). In the detailed mode, each module is represented by a circular group of genes and all interactions between the genes are presented in the network. In other words, the visualized network is actually same with the first given network except that the genes belonging to a same module are located close to each other. On the other hand, each module is represented by a single node and a relation between modules is represented by a directed link in both of the brief modes. The only difference between the two brief modes is that the weight of a link means the number of interactions between a pair of modules in the brief mode with absolute relations, whereas it means the ratio of the number of interactions between a pair of modules to the maximal possible number of interactions between them. By properly specifying the appearance ratio parameter which is defined the ratio of the number of interactions to be visible over the total

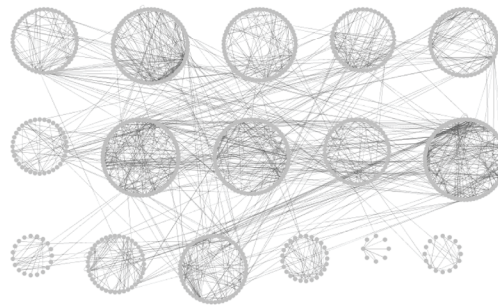
number of interactions between modules, I can retrieve more reduced information

(a) Results of each module && whole network

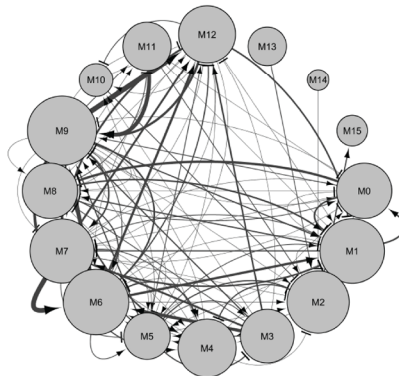
Module's information												
Module ID	#Nodes	#Edges	In-Rob(s)	In-Rob(r)	Out-Rob(s)	Out-Rob(r)	Degree	Closeness	Betweenness	Stress	Eigenvector	
0	50	70	0.84297	0.81922	0.02125	0.0225	17	0.02564	2.93507	22	0.17244	
1	69	168	0.80038	0.78884	0.01925	0.02038	21	0.02439	8.27477	67	0.13733	
2	56	71	0.83664	0.81864	0.02372	0.02511	16	0.025	13.30151	71	0.17266	
3	46	67	0.80022	0.79082	0.03795	0.04018	15	0.02326	19.72936	85	0.07772	
4	30	39	0.84453	0.83281	0.00885	0.00937	10	0.025	0.75945	11	0.22676	
5	19	24	0.90214	0.83717	0	0	18	0.02632	2.2728	15	0.30039	
6	70	121	0.84453	0.83772	0.02277	0.02411	32	0.03125	24.82232	155	0.40832	
7	53	133	0.80027	0.79496	0.03508	0.03715	29	0.02941	21.32408	130	0.38299	
8	72	131	0.81283	0.79384	0.02214	0.02344	21	0.02941	6.50281	45	0.35966	
9	61	106	0.83867	0.83088	0.03136	0.03302	17	0.03564	0.76434	6	0.16433	

Network's information									
#Nodes	#Edges	Robustness(s)	Robustness(r)	Modularity	#Modules	In-Rob(s)	In-Rob(r)	Out-Rob(s)	Out-Rob(r)
754	1624	0.67721	0.77521	0.72825	16	0.84001	0.82517	0.02257	0.0239

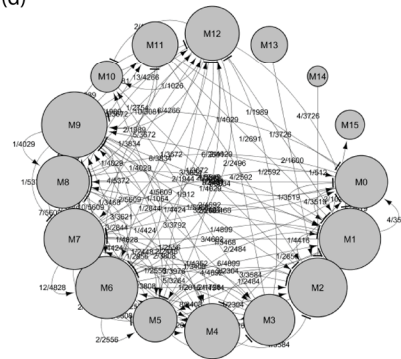
(b)



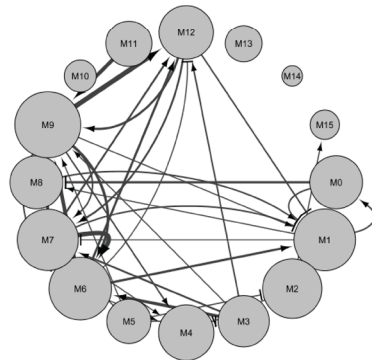
(c)



(d)



(e)



(f)

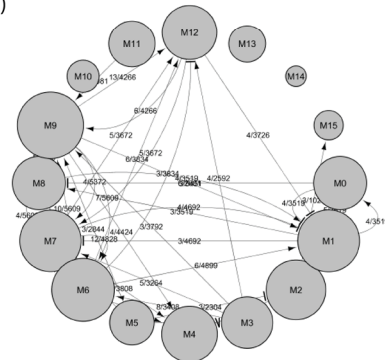


Figure 3.3. Analysis results of the STKE network by MORO.

(a) A summary table. (b) A total of 16 modules each of which is represented by a circular list of genes. (c)-(d) Results of the brief visualization mode with absolute and relative relations, respectively. (e)-(f) The reduced visualization results.

about the brief relations between modules. For example about the STKE network, Figures 3.3(e) and 3.3(f) shows the visualization results reduced from Figures 3.3(c) and 3.3(d), respectively, by specifying the appearance ratio to 0.3. Then, I can

identified which module is strongly interacting with or isolated from other modules (see Figure S3.1(e) and Figure S3.1(f) in Appendix C for the result of the HSN network).

To validate effectiveness of my app, I also conducted the same case study about large-scale signaling networks as in a previous study (Tran and Kwon, 2013) which reported that the network modularity tends to be negatively correlated to the robustness against state perturbations. To reproduce such a negative relationship between network modularity and the robustness in this study, I generated 6,400 random Boolean networks and computed the robustness and the modularity of each network by using MORO. I note that this extensive simulation could be conducted in a practical time by the parallel implementation of main functions in MORO. As a result, I could observe the same negative relationship between the modularity and the robustness, consistent to the result in (Tran and Kwon, 2013) (see Figure S3.2(a) in Appendix C). In addition, I observed that the results of STKE and HSN are very close to the trend line of the random Boolean networks. Moreover, I could also observe that the in-module robustness is clearly negatively correlated with the network modularity (Figure S3.2(b) in Appendix C), whereas the out-module robustness is not (Figure S3.2(c) in Appendix C). In addition, the in-module robustness was positively correlated with the network robustness (Figure S3.2(d) in Appendix C), whereas the out-module robustness was not (Figure S3.2(e) in Appendix C). As explained in the previous study, I could also conclude that the negative relationship between network robustness and modularity is mainly caused by the relationship between in-module robustness and network modularity through intensive simulations using my app. Figure 3.3 shows the analysis results of the STKE network by MORO. In particular, subfigure (a) is a summary table which shows modularity and robustness results in module and network levels are listed in the upper and the lower tables, respectively. Subfigure (b) is the result of the detailed visualization mode. I found a total of 16 modules each of which is represented by a circular list of genes. Two subfigures (c and d) are results of the brief visualization mode with absolute and relative relations, respectively. Each module is represented by a single group node whose radius is proportional to the number of nodes belonging to the module. The weight of a link denotes the number of interactions between the corresponding pair of modules and the ratio of the number of interactions between a pair of modules to the maximal possible

HSN network

Number of considered initial-states (S)	Single CPU (A)	Multi-core CPU (B)	Speedup (A/B)	GPU (C)	Speedup (A/C)
50	467:00:15	00:10:13	2744	00:09:58	2925
100	934:52:07	00:20:01	5488	00:19:16	5850
150	1401:47:01	00:30:39	8232	00:28:75	8775
200	1869:03:03	00:40:52	10976	00:38:38	11700
1000	9345:16:06	03:24:33	54880	03:11:01	58500

STKE network

Number of considered initial-states (S)	Single CPU (A)	Multi-core CPU (B)	Speedup (A/B)	GPU (C)	Speedup (A/C)
50	01:22:50	00:00:06	825	00:00:13	380
100	02:45:00	00:00:10	990	00:00:24	412
150	04:07:15	00:00:14	1060	00:00:35	424
200	05:30:00	00:00:18	1100	00:00:46	430
1000	27:30:00	00:01:27	1137	00:03:40	450

Table 3.1. Running time of MORO.

A total of running time to compute robustness and modularity is compared among single-CPU, multi-core CPU, and GPU options. Time is formatted as *hh:mm:ss*.

number of interactions between them in subfigures (c and d), respectively. Subfigures (e and f) are the reduced visualization results. They are subnetworks induced from subfigures (c and d), respectively, by removing all links except about 30% of links with the highest weight values (This is performed by specifying the appearance ratio parameter in MORO).

3.6.2 Time performance analysis

To show the computational cost of MORO, I examined the running time in calculating robustness and modularity in the HSN and STKE networks. I tested the app on a system with an NVIDIA GeForce GTX 680 GPU with 1,536 processor at 1GHz, seven-core Intel(R) Core i7-4770K CPU 3.50 GHz, and 16 GB of memory. Table 3.1 shows the result. In case of the HSN network, the speedup by the GPU-based parallel computation over the single-CPU was slightly greater than that by multi-core CPU, and both speedups were proportional to the number of considered initial states. On the other hand, it is interesting that the speedup by multi-core CPU was greater than that by GPU, and both were not proportional to the number of initial states in case of the STKE network. I infer that the analysis of the STKE network was terminated before the parallel computation power is fully utilized due to the relatively small size of the network. Taken together, I can efficiently analyze the relation

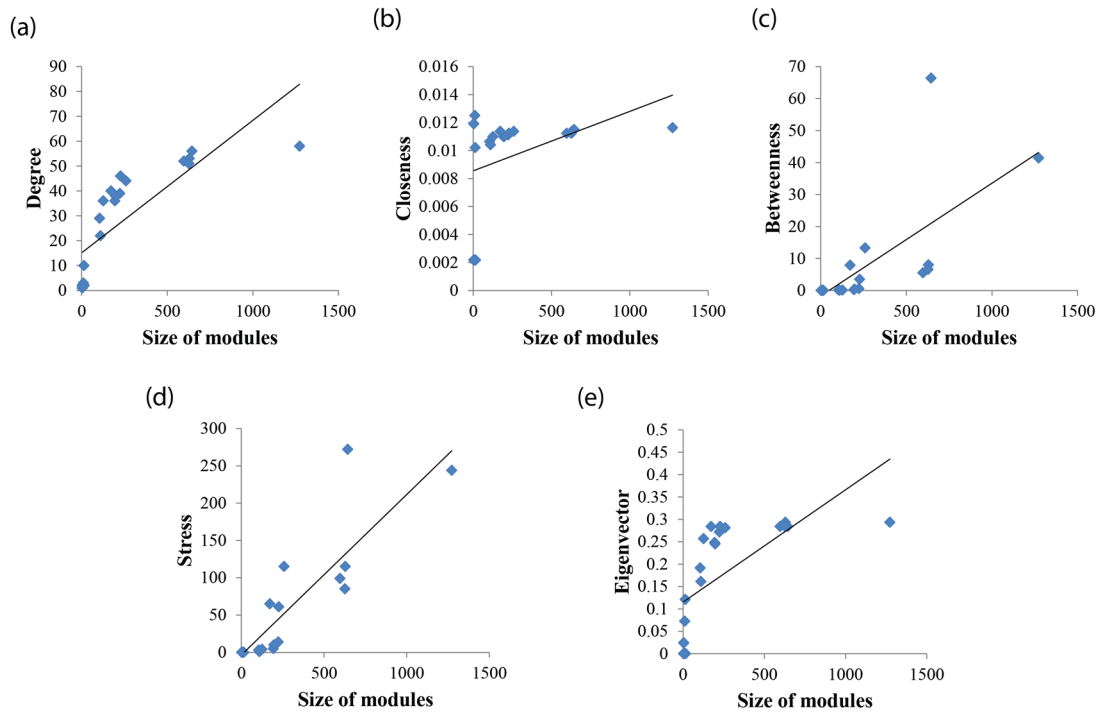


Figure 3.4 Changes of module centrality values against the module size in the STKE network.

(a)-(e) Results with respect to degree, closeness, betweenness, stress, and eigenvalue. The module size which is defined as the number of nodes belonging to the module showed positive relationships with all module centrality measures. The correlation coefficients in (a)-(e) were 0.75339, 0.564168, 0.599553, 0.657316, and 0.511411, respectively, with all p-value $< 10^{-4}$.

between robustness and modularity in large-scale networks by parallel computation with two options, multi-core CPU and GPU.

3.6.3 Module centrality analysis

After I obtain the modular structure of a network, I can analyse the centrality of modules based on the brief module visualization result. Specifically, I consider a module network where a node and a link represent a module and the set of interactions between a pair of modules, respectively. Then, I can examine five well-known centrality values such as degree, closeness, betweenness, stress, and eigenvector in the module network. In this case study, I examined the change of the centrality values against the module size, which is defined by the number genes belonging to a module, in the STKE (Figure 3.4) and HSN (Figure S3.3 in Appendix C) networks. It is interesting that all centrality measures or all except closeness

Category	GO Terms	The largest module		The rest of genes		P-value
		No. of genes	%	No. of genes	%	
Cellular component	Cytoplasm	161	20.33	767	16.49	0.00794
	Nucleus	230	29.04	531	11.42	0
	Protein complex	26	3.28	83	1.79	0.0054
Molecular Function	Protein binding	249	31.44	1115	23.97	0.00001
	Metal ion binding	85	10.73	351	7.55	0.00227
	Nucleotide binding	58	7.32	234	5.03	0.00814
	DNA binding	127	16.04	150	3.23	0
Biological Process	Gene expression	40	5.05	139	2.99	0.00263
	Viral process	38	4.80	132	2.84	0.00338
	Regulation of transcription, DNA-templated	115	14.52	177	3.81	0

Table 3.2. GO analysis in the HSN network.

GO terms which are significantly enriched between the largest module and the rest of modules are listed. All *P*-values are calculated by using a z-test.

showed the positive relations with the module size in the STKE and HSN networks, respectively. In other words, the module was likely to be more central as the module size gets larger. To investigate if this property is reserved in random networks, I generated two groups of 100 random networks by shuffling interactions of the STKE and HSN networks while preserving a degree distribution, and examined the change of the centrality values against the module size (see Figures S3.4 and S3.5 in Appendix C). Similar to the result in the signaling networks, the module size was positively correlated with the centrality values in the random networks. This suggests that the hub modules might play an important role in the community network (Estrada and Rodríguez-Velázquez, 2005; Kim and Anderson, 2012; Li, et al., 2015). Additionally, I examined the relationship between the in-/out-module robustness and the module centrality values in the STKE and HSN networks (see Figures S3.6 and S3.7 in Appendix C). Unlike the relation with the module size, the in-/out-module robustness was not significantly correlated with the centrality values. In other words, the centrality of modules cannot indicate the in-/out-module robustness in the signaling network.

3.6.4 GO analysis

It is possible to analyze GO enrichment (Consortium, 2008) by using MORO. To show this function, I first specified two groups of genes, which consist of the genes in the largest module (1,042 genes) and the rest of genes (4,401 genes), respectively, in

the HSN network. Table 3.2 shows all GO terms which were more enriched in the largest module than in the others: cytoplasm, nucleus, and protein complex in cellular component terms; protein, metal ion, nucleotide, and DNA bindings in molecular function terms; gene expression, viral process, and regulation of DNA-templated transcription in biological processes terms. As a result, MORO can provide the useful information about GO analysis between any two groups of modules.

3.7 Conclusions

Many recent reports have reported that robust behavior against mutations might be correlated to the modularity of a signaling network. Motivated by these results, I developed a novel Cytoscape app called MORO, which can analyze the relationship between network robustness and modularity. I implemented it in parallel by using the OpenCL library to allow application to very-large-scale networks. In addition, my app can provide information about topological relations between modules by means of various visualization modes and centrality analysis. MORO includes also five centrality measures which can examine how centrally each module is positioned in terms of relations among the modules. Moreover, it can conveniently analyze the gene ontology enrichment of modules only if Entrez id of gene is given. A batch-mode simulation function was also included to allow verification of whether a finding is a design principle of random networks.

CHAPTER 4. INVESTIGATION ON CHANGES OF MODULARITY AND ROBUSTNESS BY EDGE-REMOVAL MUTATIONS IN SIGNALING NETWORKS

4.1 Overview

Robustness and modularity are key properties to understand complex dynamics in large-scale biological networks. It is also notable that these properties can be changed by structural mutations because they are highly dependent on the network structure. Although these previous studies found interesting relations between the robustness and the modularity, there are some issues needed to be investigated as follows. The first issue is that there is little known knowledge about changes of the modularity and the robustness. Another interesting issue is that there is little known about motifs which indicate the changes of the modularity, the robustness, or both. The last issue is that there was no previous study to compare sets of nodes or interactions which efficiently control the changes of the modularity and the robustness. This can be impressive because the result can be used to identify functionally important nodes or interactions such as drug targets.

In this work, I tried to investigate the changes of the modularity and the robustness by edge-removal mutations in signaling networks. I first found that both the modularity and the robustness increased on average against edge-removal mutations, but the change of modularity is negatively correlated with the change of robustness. More intriguingly, the modularity change was positively correlated with the degree, the number of FBLs, and the edge betweenness of removed edges, whereas the robustness change was negatively correlated with them. Additionally, I found that these findings are consistently conserved in the random networks. Moreover, I identified two groups of genes which are incident to the highly-modularity-increasing and the highly-robustness-decreasing edges against the edge-removal mutations, respectively, and observed that they are likely to be central by forming a considerably large connected component. The gene-ontology enrichment of each of the gene groups was clearly different from the rest of genes. Finally, I found that the highly-robustness-decreasing edges can be promising edgetic drug-targets.

4.2 Change of modularity and robustness by edge-removal mutations

This study focuses on how the modularity and the robustness of a network are changed by edge-removal mutations. Let m_1 and r_1 be the modularity and the robustness of the wild-type network, respectively. Given a removal rate parameter n (%), the mutant network is constructed by simultaneously removing approximately n percent of a total number of edges from the wild-type network. Then let m_2 and r_2 be the modularity and the robustness of the mutant network. I defined the changes of the modularity and the robustness by the edge-removal mutations as $(m_2 - m_1)$ and $(r_2 - r_1)$, respectively. An illustrative example of the notion about the changes of modularity and robustness by edge-removal mutations is shown in Figure 1.2.

4.3 Software for statistical tests

In this study, IBM SPSS statistics (SPSS, 2012) was used to conduct all statistical tests.

4.4 Results

4.4.1 Relationship between changes of modularity and robustness by edge-removal mutations

I first investigated the changes of the modularity and the robustness by edge-removal mutations in three real networks T-LGL, STF, and HIV-1 (see section 4.2 for more detail), and the results are shown in Figure 4.1 (T-LGL) and Figures S4.1 and S4.2 (STF and HIV-1, respectively) in Appendix C. In this study, I computed the average changes of the modularity and the robustness values over 5,000 trials of edge-removal mutations. In addition, I varied the removal rate, which denotes the percentage of the number of removed edges over the total number of edges, from 1% to 5%. I first tested whether the average changes are significantly positive using one-sample t-test. I note that the average changes were normally distributed, as assessed by Kolmogorov-Smirnov's test (see Figures S4.3-S4.5 in Appendix C for details) and there were no or very few significant outliers, as assessed by a boxplot inspection (see Figures S4.6-S4.8 in Appendix C for details). As shown in Figure 4.1(a), I observed that both average changes were positive for all removal rates, which means that the modularity and the robustness values were increased by edge-removal (All P-values < 0.0001; see Figure S4.1(a) and Figure S4.2(a) in Appendix C for the results of STF

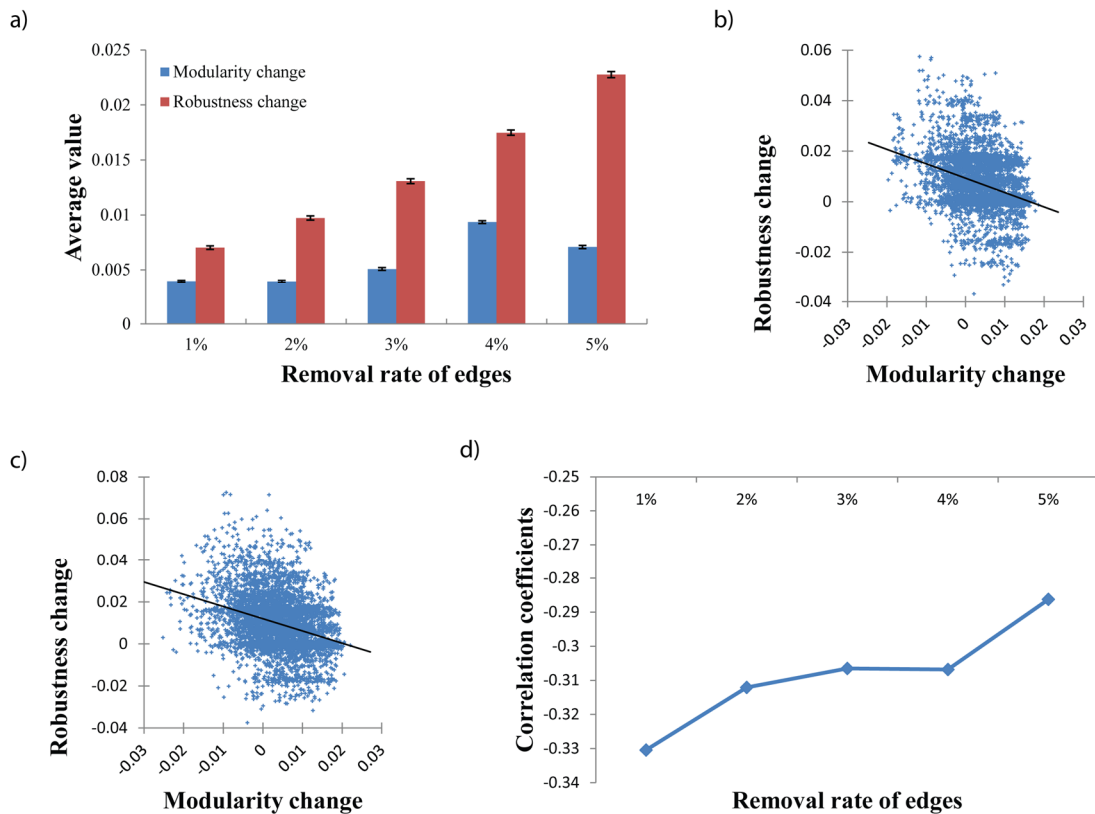


Figure 4.1. Analysis of the changes of the modularity and the robustness by edge-removal mutations in T-LGL signaling network.

The removal rate of edges was varied from 1% to 5% (More specifically, the numbers of removed edges were 2, 3, 4, 5, and 7, respectively, among a total of 142 edges). For each removal rate, 5,000 trials of edge-removal were examined. See Appendix C: Fig. S4.1 and S4.2 for the analysis results of STF and HIV-1 signaling networks. **(a)** Results of average changes of the modularity and the robustness against the removal rate of edges. Y-axis value and error bar represents the average and the standard deviation divided by the square root of the sample size (5000), respectively. Both average values were significantly larger than zero (All P-values <0.0001 using one-sample t-test). The one-sample t-test was valid because the average values were normally distributed (see Appendix C: Fig. S4.3) and there were no or very few significant outliers (see Appendix C: Fig. S4.6). **(b)-(c)** Relationship between the changes of the modularity and the robustness in the case that the removal rate is 1% and 2%, respectively. A significant negative relationship was observed (Correlation coefficients were -0.33042 and -0.31208 in (b) and (c), respectively, with all P-values <0.0001). This relationship was consistently observed for larger removal rates (see Appendix C: Fig. S4.9). **(d)** A trend of correlation coefficients between the changes of the modularity and the robustness against the removal rate of edges.

and HIV-1 networks, respectively).

In addition, the increase of the robustness was positively related to the removal rate. To examine the relationship between the changes of modularity and robustness values, I scattered them in the cases that the removal rate is 1% (Fig. 4.2(b)) and 2% (Fig. 4.2(c)). Intriguingly, there was a negative correlation between the modularity change and the robustness change, and this was consistently observed in the cases of larger removal rates (see Appendix C: Fig. S4.9) and the other networks (see Appendix C: Fig. S4.1(b)-(c)) and Fig. S4.2(b)-(c)). Figure 4.1(d) shows the trend of the correlation coefficient values between the changes of modularity and robustness values against the removal rate, and I observed that they were significantly negative irrespective of the removal rates (see Appendix C: Fig. S4.1(d) and S4.2(d) for the results of STF and HIV-1 networks, respectively). Actually, the negative relationship between the modularity and the robustness in signaling networks was observed in previous studies (Tran and Kwon, 2013; Truong, et al., 2016). However, it should be noted that the previous finding does not imply any relation between the changes of the modularity and the robustness by edge-removal mutations. To further examine if the negative relationship I found is a general property in randomly structured networks, I generated three sets of 100 random networks shuffled from T-LGL, STF, and HIV-1 (see Chapter 2 for more detail), and could observe consistent (see Appendix C: Fig. S4.10). This implies that such the negative relation between the changes of the modularity and the robustness can be regarded as a general principle conserved in randomly structured networks.

4.4.2 Structural characteristics to affect the changes of the modularity and the robustness

I showed that the changes of the modularity and the robustness are correlated when a network is subject to edge-removal mutations. To reveal structural characteristics to affect the changes of the modularity and the robustness, I investigated the correlations of each of the changes of the modularity and the robustness with each of three edge- based structural properties, DEG, NuFBL and EBEW (see Chapter 2 for the definitions) in T-LGL signaling network (Fig. 4.2; see Appendix C: Fig. S4.11 and S4.12 for the results of STF and HIV-1, respectively). In Fig. 4.2, average DEG, NuFBL, and EBEW values of the removed edges over 5000 trials with 1% of the removal rate were examined. Intriguingly, I found that the change of the modularity is positively correlated with the average DEG, EBEW and

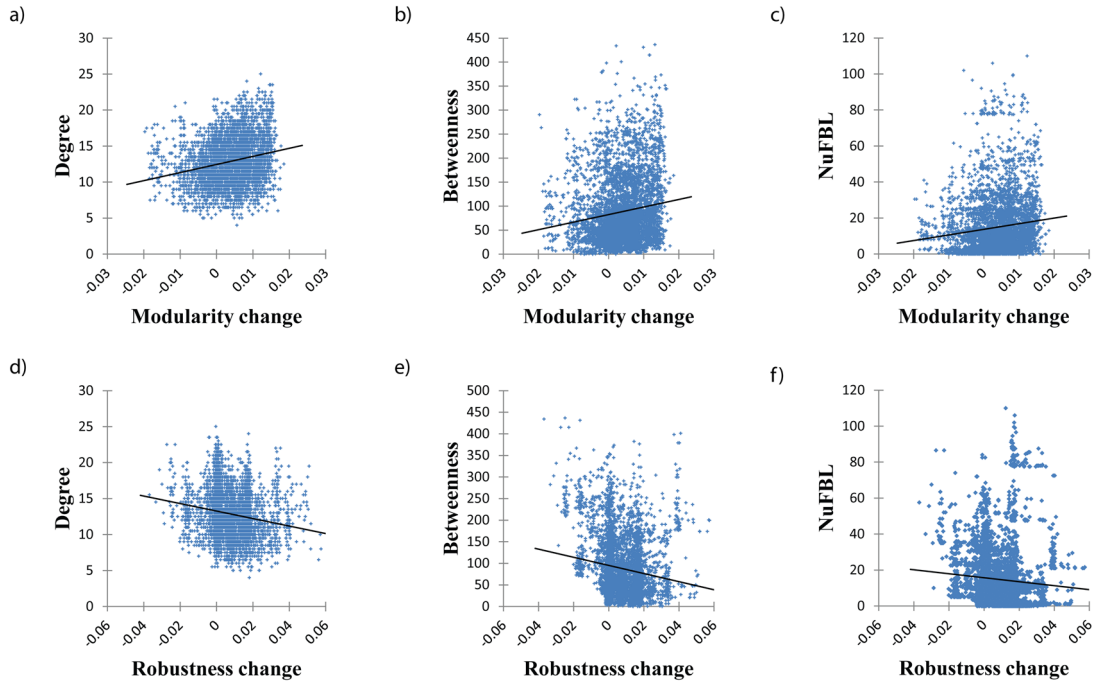


Figure 4.2. Relationship of each of the changes of the modularity and the robustness with the edge-based structural properties in T-LGL signaling network.

The removal rate was set to 1%, and a total of 5,000 trials of removals were examined. **(a)-(c)** Relations of the change of modularity with edge-based degree, EBEW, and NuFBL, respectively. The change of modularity was significantly positively correlated with all structural properties (Correlation coefficients were 0.22443, 0.14564, and 0.12888, respectively, with all P-values<0.0001). **(d)-(f)** Relations of the change of robustness with edge-based degree, EBEW, and NuFBL, respectively. The change of robustness was significantly negatively correlated with all structural properties (Correlation coefficients were -0.18050, -0.15030, and -0.07933, respectively, with all P-values<0.0001).

NuFBL of removed edges (The correlation coefficients in Fig. 4.2(a)-(c) were 0.24708, 0.13786, and 0.11720, respectively, with all p-values<0.001). That is to say, removing edges with a higher degree, EBEW, or NuFBL is more likely to increase the network modularity. These results can be relevant to previous results. For example, the edges with high betweenness values are most likely to lie between subgraphs (Yoon, et al., 2006), and thus removing those edges could make a network more separately or more modularized. I also found that the change of the robustness is negatively correlated with the average DEG, EBEW and NuFBL of the removed edges (The correlation coefficients in Fig. 4.2(d)-(f) were -0.21738, -0.14694, and -0.10537, respectively, with all p-values<0.0001). In other words, removing edges with

compared with some previously known results regarding node-based mutations. For example, some studies reported that a node involving more FBLs is likely to be sensitive against node-based mutations. To show that these results hold in random networks, I generated three sets of 100 random Boolean networks each of which was shuffled from T-LGL, STF, and HIV-1 networks, respectively. Through extensive simulations with the removal rate of 1%, I could observe consistent results (see Appendix C: Fig. S4.13-S4.15, All P-values<0.0001 using t-test). In other words, the degree, the edge-betweenness and the number of FBLs were positively correlated with the change of the modularity whereas they were negatively correlated in the random networks. It means that those structural characteristics might be a vital factor in controlling both the changes of the modularity and the robustness.

4.4.3 Topological distribution of highly modularity-increasing and robustness-decreasing edges by removal mutations

In the previous subsection, it was shown that the change of the modularity is positively correlated with the degree, the edge betweenness, and the number of involved FBLs with respect to the removed edges whereas the change of robustness is negatively correlated with them. From these results, I hypothesized that the edges whose removal will increase the modularity or decrease the robustness tend to be centrally located in signaling networks. To validate this hypothesis, I first specified “Highly-modularity-increasing” (High-MI) and “Highly-robustness-decreasing” (High-RD) sets of edges as follows: I examined the changes of the modularity and the robustness over 5000 trials of edge-removal mutations with 1% removal rate, and collected top- K set of edges among them in an increasing (resp. decreasing) order of the change of the modularity (resp. the robustness). Considering the distributions of the change of the modularity (resp. robustness), K was chosen to 20, 20, and 18 (resp., 31, 18, and 16) for T-LGL, STF, and HIV-1 networks, respectively. Then High-MI (resp. High-RD) denotes the union of the edges each of which was included in the modularity-increasing (resp. robustness-decreasing) top- K edges. Accordingly, I identified High-MI (High-RD) groups consisting of 22, 79, and 42 edges (resp. 30, 69, and 33 edges) in T-LGL, STF, and HIV-1 networks, respectively. Furthermore, I defined High-MI-incident (High-RD-incident) group which is a set of genes incident to an edge in the High-MI (resp. High-RD) edge group, and found the number of genes in the High-MI-incident (resp. High-RD-incident) were 29, 81, and 59 (resp.

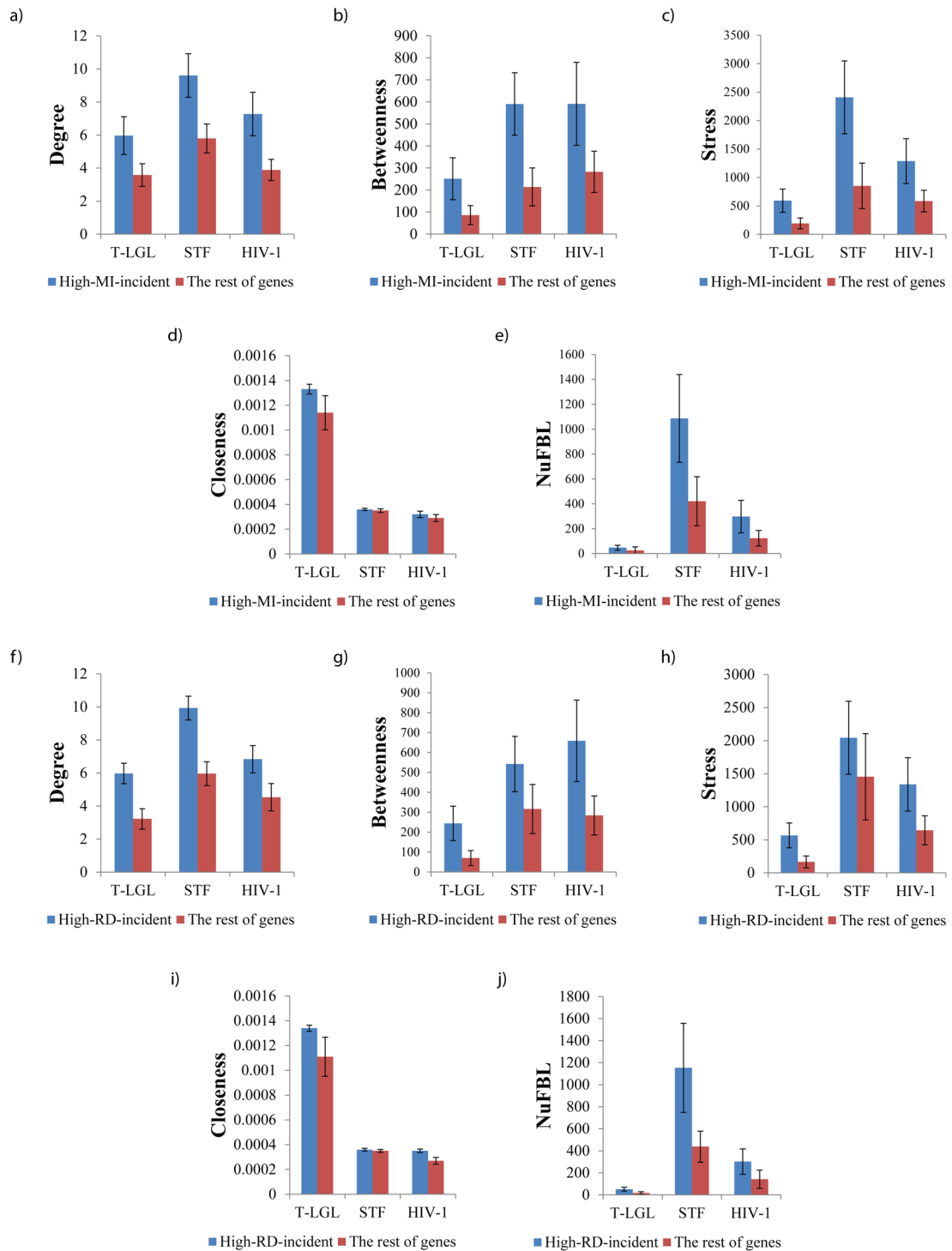


Figure 4.4. Comparison of node-based centralities between High-MI-incident/High-RD-incident group and the rest of genes in the signaling networks.

(a)-(e) Results of node-based degree, betweenness, stress, closeness, and NuFBL, respectively, with respect to High-MI-incident group. (f)-(j) Results of node-based degree, betweenness, stress, closeness, and NuFBL, respectively, with respect to High-RD-incident group.

33, 72, and 48) in T-LGL, STF, and HIV-1 networks, respectively. The topological

distributions of High-MI and High-RD edge sets in T-LGL, STF, and HIV-1 networks

Type of GO analysis	GO term	High-MI-incident (%)	The rest of genes (%)	P-value	
Modularity change	Non-membrane spanning protein tyrosine kinase activity	75.00	25.00	5.4E-6	
	Positive regulation of peptidase activity	57.62	48.01	3.4E-12	
	Positive regulation of adaptive immune response	75.00	25.00	34.0E-6	
	Regulation of immunoglobulin mediated immune response	75.00	25.00	2.9E-6	
	Phosphatase binding	85.71	14.29	72.0E-9	
	Protein phosphatase binding	83.33	16.67	260.0E-9	
	Cytokine receptor binding	64.29	35.71	2.1E-15	
	Death receptor binding	60.00	40.00	960.0E-12	
	GO term	High-RD-incident (%)	The rest of genes (%)	P-value	
Robustness change	Positive regulation of cysteine-type endopeptidase activity	63.62	42.41	24.0E-12	
	Positive regulation of peptidase activity	67.22	38.41	3.4E-12	
	Response to nicotine	60.00	40.00	200.0E-9	
	Regulation of adaptive immune response	83.33	16.67	540.0E-9	
	Positive regulation of apoptotic signaling pathway	67.22	38.41	13.0E-12	
	Regulation of NIK/NF-kappaB signaling	75.00	25.00	5.4E-6	
	Necroptotic signaling pathway	79.25	26.42	380.0E-9	
	Extrinsic apoptotic signaling pathway in absence of ligand	66.84	40.11	370.0E-12	

Table 4.1. Results of GO analysis between High-MI-incident/High-RD-incident group and the rest of genes in T-LGL signalling network.

All P-values using Bonferroni test (see Appendix D: Table S4.1 and S4.2 for the results of STF and HIV-1 signaling networks).

are shown in Figure 4.3 and Fig. S4.16-S4.17 in Appendix C, respectively. As expected, it was observed that the edges in High-MI and High-RD groups are likely to be located at the centre of the signaling network. In order to more clarify this

observation, I compared node-based centrality values between each set of High-MI-incident and High-RD-incident groups and the set of rest genes. Specifically, I computed average degree, node-based betweenness (Freeman, 1977), stress (Shimbel, 1953), closeness (Wuchty and Stadler, 2003), and the number of involved FBLs (Kwon and Cho, 2008) for each group of nodes (Fig. 4.4). As depicted in the figure, I found that genes of High-MI-incident and High-RD-incident groups showed higher degree, node-based betweenness, stress, closeness, and the number of FBLs than the rest of genes (Only three cases among 30 comparisons did not show significant differences.) In other words, the genes incident to the interactions whose greatly increase the modularity or decrease the robustness tends to be central in the signaling network. Additionally, I visualized the connectedness of edges of High-MI and High-RD groups by projecting them into a subnetwork from T-LGL (see Fig. 4.3(c) and (d) in Appendix C, respectively), STF (see Fig. S4.16(c) and (d) in Appendix C, respectively), and HIV-1 (see Fig. S4.17(c) and (d) in Appendix C, respectively) networks. As shown in the figures, every subnetwork forms a single connected component. This implies that the highly modularity-increasing or robustness-decreasing edges with respect to edge-removal mutations are closely located in signaling networks. Figure 4.4 shows comparison of node-based centralities between High-MI-incident/High-RD-incident group and the rest of genes in the signaling networks. Subfigures from (a) to (e) show results of node-based degree, betweenness, stress, closeness, and NuFBL, respectively, with respect to High-MI-incident group. In each subfigure, “The rest of genes” means non “High-MI-incident” genes. Subfigures from (f) to (j) show results of node-based degree, betweenness, stress, closeness, and NuFBL, respectively, with respect to High-RD-incident group. In each subfigure, “The rest of genes” means non “High-RD-incident” genes. For all subfigures, Y-Axis value and error bar represents the average and 95% confidence interval, respectively. Genes belonging to High-MI-incident and High-RD-incident group showed higher degree, node-based betweenness, stress, closeness, and the number of involved feedback loops than the rest of genes (All P-values<0.05 except stress of STF in (c), closeness of HIV-1 in (d), and NuFBL of T-LGL in (j)).

4.4.4 Gene ontology analysis of a set of genes incident to highly-modularity-increasing or highly-robustness-decreasing edges

I conducted Gene Ontology (GO) enrichment analysis (The Gene Ontology Consortium, 2008) using ClueGO tool (Bindea, et al., 2009) to investigate the locational and functional characteristics of sets of High-MI-incident and High-RD-incident genes. The results are shown in Table 4.1 (see Appendix D: Table S1-S2). Some GO terms such as protein tyrosine kinase and peptidase activity are more highly observed in High-MI-incident and High-RD-incident groups. The former is an enzyme which transfers a phosphate group from adenosine triphosphate to a protein in a cell, and the latter is catalysis of the hydrolysis of a peptide bond. In addition, High-MI-incident and High-RD-incident gene groups showed a greater fraction of response function terms. Regulation of adaptive immune response is any process that modulates the frequency, rate, or extent of an adaptive immune response regarding to robustness change. Furthermore, High-MI-incident group showed a greater portion of vital binding functions. For example, a protein phosphatase is an enzyme that removes a phosphate group from the phosphorylated amino acid residue of its substrate protein, and its binding function is interacting selectively and non-covalently with any protein phosphatase. On the other hand, High-RD-incident group showed a greater fraction related to signaling pathway. For instance, necroptosis is a programmed form of necrosis, or inflammatory cell death, and its signaling pathway is a series of molecular signals which triggers the necroptotic death of a cell. Taken together, significantly different functions between High-MI-incident/High-RD-incident groups of genes and the rest of genes can be characterized.

4.4.5 Edge-based drug discovery

I performed a case study to show an application for edge-based drug discovery. For every interaction in High-RD group, I examined the inclusion frequency of the interaction in top- K edge sets ranked by a decreasing order of the robustness change among 5000 trials of edge-removal mutations with 1% removal rate. I found that ($JAK \rightarrow STAT3$), ($IP3R1 \rightarrow Ca$), and ($gp41 \rightarrow CD28$) showed the highest frequency in the T-LGL, STF, and HIV-1 networks, respectively. I hypothesized that these edges can be candidates of edgetic drug-targets, because they most frequently caused the highest decreasing robustness through removal mutations. To validate this, I surveyed some recent experimental studies. Regarding ($JAK \rightarrow STAT3$) interaction of T-LGL network (Fig. 4.5), it was shown that the interaction is associated with oncogenesis, proliferation, survival, metastasis, angiogenesis, and immune evasion in

gastrointestinal cancers (Bournazou and Bromberg, 2013; Nikolaou, et al., 2013). For example, a colorectal cancer might be developed by dysregulation of the interleukin (IL)-6-mediated $JAK \rightarrow STAT3$ pathway, and therefore strategies targeting the IL-6/JAK/STAT3 pathway have emerged as attractive options to treat colorectal cancer (Wang and Sun, 2014). Next, the ($IP3R1 \rightarrow Ca$) interaction of STF network (see

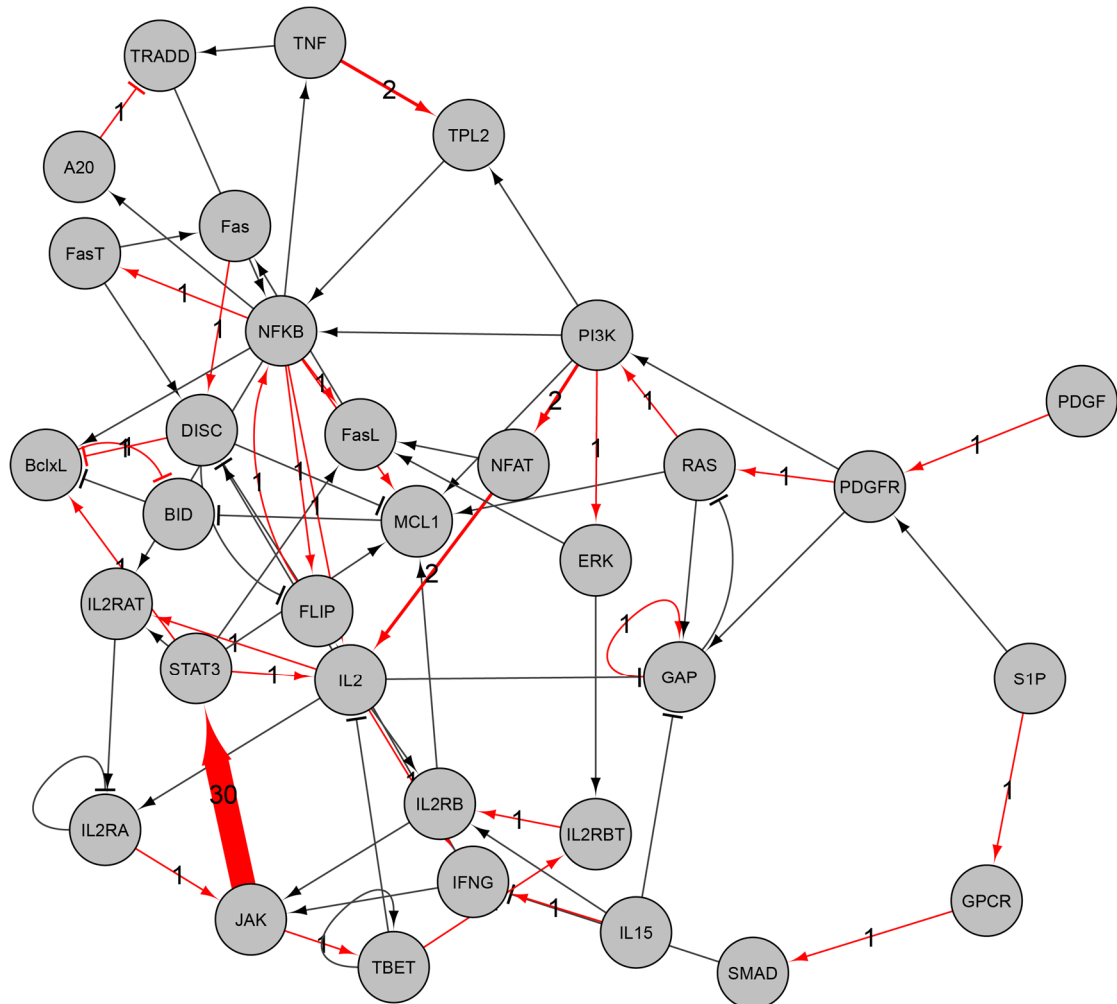


Figure 4.5. Edge-removal analysis for edgetic drug discovery in T-LGL signaling network.

The arrows and bar-headed lines represent positive and negative interactions, respectively. Line thickness is proportional to the inclusion frequency of the interaction in top-K edge sets ranked in a decreasing order of the robustness change among 5000 trials of edge-removal mutations with 1% removal rate. The interaction ($JAK \rightarrow STAT3$) was observed 30 times in top-K edge sets (K was chosen to 30). (see Appendix C: Fig. S4.18 and S4.19 for the results of STF and HIV-1 networks.)

Appendix C: Fig. S4.18) played an important role of dynamical relationship between IP3R1 and PI3K, which are the most influential components associated with drug

resistance (Puniya, et al., 2016). Systemic analysis of these components and their upstream components has resulted in identifying novel combinations of drug targets. In HIV-1 network, (*gp41* → *CD28*) was found to be the highest frequency interaction (see Appendix C: Fig. S4.19), but there was no relevant experimental study to support it. However, I could find biological evidence related to the second highest frequency interaction, (*PI3K* → *PIP3*). It is included in PI3K/Akt/mTOR pathway (Slomovitz and Coleman, 2012), which is known to be frequently activated in ovarian cancer. Therefore, inhibitors targeting this pathway can be evaluated as treatment strategies for ovarian cancer, either mono-therapy or in combination with cytotoxic agents (Mabuchi, et al., 2015). Another interesting point was the feedback loops involved with those interactions. I found a large number of feedback loops were related to (*JAK* → *STAT3*), (*IP3R1* → *Ca*), and (*PI3K* → *PIP3*) in T-LGL, STF, HIV-1, respectively (The numbers were 70, 286, and 872, respectively). Considering that the number of involved FBLs was shown to be associated with the functional importance of a node or an interaction, it implies that the found interactions can be promising drug-targets.

4.5 Conclusions

There have been many computational studies about the network robustness and modularity, whereas there are few studies on investigating the modularity change and the robustness change. Through extensive simulations, I found that both the modularity and the robustness increased on average in mutant networks by edge-removal mutations in this study. However, it was interesting that the changes of the modularity and the robustness were negatively correlated. Another interesting finding is that the changes of the modularity and the robustness are positively and negatively, respectively, correlated with each of the degree, the number of FBLs, and the edge betweenness of removed edges. These results were consistently observed in randomly structure networks. Additionally, I identified two sets of genes which are incident to the highly-modularity-increasing and the highly-robustness-decreasing edges, respectively, and observed that they are likely to be central by forming a large connected component. These two gene sets were enriched with different GO terms and the investigation on the reason why such GO terms are related to modularity and robustness will be a future study. Finally, I found that the highly-robustness-

decreasing edge can be considered for promising edge-based drug-targets. Taken together, my results in this study can be useful to unravel novel dynamical characteristics of signaling networks.

CHAPTER 5. CONTRIBUTION SUMMARY AND FURTHER WORK

5.1 Contribution summary

Biological networks consisting of molecular components and interactions are represented by a graph model. There have been some studies based on that model to analyze a relationship between structural characteristics and dynamical behaviors in signaling network. Inspired from that, in this study, I firstly developed a novel software tool, MORO, which can analyze the relationship between network robustness and modularity of large-scale biological networks in parallel. Then, I investigated changes of modularity and robustness by edges-removal mutations in signaling networks.

5.1.1 MORO: a GPU-based software

Although there have been many studies revealing that dynamic robustness of a biological network is related to its modularity characteristics, no proper tool exists to investigate the relation between network dynamics and modularity. Accordingly, I developed a novel Cytoscape app called MORO, which can conveniently analyze the relationship between network modularity and robustness. I employed an existing algorithm to analyze the modularity of directed graphs and a Boolean network model for robustness calculation. In particular, to ensure the robustness algorithm's applicability to large-scale networks, I implemented it as a parallel algorithm by using the OpenCL library. A batch-mode simulation function was also developed to verify whether an observed relationship between modularity and robustness is conserved in a large set of randomly structured networks. The app provides various visualization modes to better elucidate topological relations between modules, and tabular results of centrality and gene ontology enrichment analyses of modules. I tested the proposed app to analyze large signaling networks and showed an interesting relationship between network modularity and robustness. My app can be a promising tool which efficiently analyzes the relationship between modularity and robustness in large signaling networks. I note that this MORO's version does not contain a function which can investigate changes of modularity and robustness in signalling networks.

5.1.2 Negative relationship between changes of modularity and robustness

There have been some studies based on that model to analyze a relationship between structural characteristics and dynamical behaviors in signaling network. However, little attention has been paid to changes of modularity and robustness in mutant networks. In this study, I investigated the changes of modularity and robustness by edge-removal mutations in three signaling networks. I first observed that both the modularity and robustness increased on average in the mutant network by the edge-removal mutations. However, the modularity change was negatively correlated with the robustness change. This implies that it is unlikely that both the modularity and the robustness values simultaneously increase by the edge-removal mutations. Another interesting finding is that the modularity change was positively correlated with the degree, the number of feedback loops, and the edge betweenness of the removed edges whereas the robustness change was negatively correlated with them. I note that these results were consistently observed in randomly structure networks. Additionally, I identified two groups of genes which are incident to the highly-modularity-increasing and the highly-robustness-decreasing edges with respect to the edge-removal mutations, respectively, and observed that they are likely to be central by forming a connected component of a considerably large size. The gene-ontology enrichment of each of these gene groups was significantly different from the rest of genes. Finally, I showed that the highly-robustness-decreasing edges can be promising edgetic drug-targets, which validates the usefulness of my analysis.

5.2 Future Work

For MORO App, there was only one kind of mutations provided in the current version. In the future, MORO will be extended to a variety of mutation types such as a node-based knockout, edge perturbation. Multiple mutations analysis could be further developed instead of single mutation at the current version. Moreover, other kinds of random network generations based on other topological properties such as small-world, hierarchical can be additionally inserted into MORO. Additionally, MORO will be extended to analyze publicly-available signaling networks represented by ordinary differential equations by devising a conversion method from continuous models to Boolean networks. Finally, I will upgrade the current version of MORO to

a bigger tool. The new app will provide a function which can investigate changes of modularity and robustness in signaling networks.

Regards to the changes of modularity and robustness by edges-removal mutations, the investigation is only based on an edges-removal mutation, further investigation into various kinds of edgetic mutations will be needed for more specific applications. For example, edge-addition, edge-switching and edge-attenuation mutations could be further developed to intensely analyze the network dynamics and to find out more potential edgetic drug-targets in drug discovery. Moreover, I will consider two sets of genes which are incident to the highly-modularity-decreasing and the highly-robustness-increasing edges, respectively, and observe that where are they located in networks. Additionally, these two gene sets were enriched with different GO terms and the investigation on the reason why such GO terms are related to modularity and robustness will be a future study. Finally, the highly-modularity-increasing edge can be considered for promising edge-based drug-targets.

Furthermore, the network element prioritization should not be limited on genes/proteins network but on other kinds such as protein complex networks where a node is a protein complex, or a pathway-based network where a node is a pathway. In general, it can be extended to deal with hyper networks which are constructed with super nodes (i.e., protein complexes, pathways) and links between such super nodes can be a factor which reflects functional relation between them. Besides biological networks, it is also interesting to extend the prioritization issues in various networks like social networks, computer networks, transportation networks, and so on. For example, the node/edge-based sensitivity analyses could be applied to online social networks such as Facebook, Twitter for identifying prominent individuals or connections.

APPENDIX A

Supporting Text A1: Nested Canalizing Functions

Given a Boolean network $G(V, A)$, the value of each variable v_i at time $t + 1$ is determined by the values of k_i other variables $v_{i_1}, v_{i_2}, \dots, v_{i_{k_i}}$ with a link to v_i at time t by the Boolean function $f_i: v_i(t + 1) = f_i(v_{i_1}(t), v_{i_2}(t), \dots, v_{i_{k_i}}(t))$. The rule f_i is called *canalizing* on the input variable v_{i_m} if there exist Boolean values, I_m and O_m , such that

$$v_{i_m}(t) = I_m \rightarrow v_i(t + 1) = O_m.$$

Then, I_m and O_m are called the canalizing and canalized values for the output variable v_i , respectively.

The notion of *nested canalizing functions (NCFs)* was introduced in Kauffman, et al. (2003), and they are a natural subset of canalizing rules. It was inspired by the question of what happens in the noncanalizing case: When a rule is not canalized by the value of the first input variable, is it canalized by one of the remaining input variables? This consecutive canalization test can be repeated for all inputs and therefore an NCF to update v_i can be represented as follows:

$$v_i(t + 1) = \begin{cases} O_1 & \text{if } v_{i_1}(t) = I_1 \\ O_2 & \text{if } v_{i_1}(t) \neq I_1 \text{ and } v_{i_2}(t) = I_2 \\ O_3 & \text{if } v_{i_1}(t) \neq I_1 \text{ and } v_{i_2}(t) \neq I_2 \text{ and } v_{i_3}(t) = I_3 \\ & \vdots \\ O_{k_i} & \text{if } v_{i_1}(t) \neq I_1 \text{ and } \dots \text{ and } v_{i_{k_i-1}}(t) \neq I_{k_i-1} \text{ and } v_{i_{k_i}}(t) = I_{k_i} \\ O_{default} & \text{if } v_{i_1}(t) \neq I_1 \text{ and } \dots \text{ and } v_{i_{k_i}}(t) \neq I_{k_i} \end{cases},$$

where all I_m and O_m ($m = 1, 2, \dots, k_i$) denote the canalizing and canalized values, respectively, and $O_{default} \neq O_{k_i}$. In addition, as in the previous studies (Kauffman, et al., 2003; Kauffman, et al., 2004), I independently and randomly specified O_1, \dots, O_{k_i} values with the probabilities

$$P(O_m = 1) = \frac{\exp(-2^{-m}\theta)}{1 + \exp(-2^{-m}\theta)}$$

where θ is a constant. On the other hand, the value of I_m is deterministically specified by the value of O_m and the sign of the interaction from v_{i_m} to v_i ($m = 1, \dots, k_i$) as the following table.

O_m	Sign of the interaction from v_{i_m} to v_i	I_m
1	Positive ($v_{i_m} \rightarrow v_i$)	1
1	Negative ($v_{i_m} \dashv v_i$)	0
0	Positive ($v_{i_m} \rightarrow v_i$)	0
0	Negative ($v_{i_m} \dashv v_i$)	1

Supporting Text A2: Output file by the batch-mode simulation on RBNs

After the batch-mode simulation is completed, a resultant file “net_based_result.txt” is created, which lists the network-based results. As shown in the figure below, the file consists of 11 results with respect to robustness, modularity and in-/out-module robustness. Each row lists the result of a single RBN.

Column	Name	Description
1	Network ID	The unique identification number of an RBN
2	No.Nodes	The number of nodes of an RBN
3	No.Edges	The number of edges of an RBN
4	sRobustness	The robustness against initial-state perturbation of an RBN
5	rRobustness	The robustness against update-rule perturbation of an RBN
6	No.Modules	The number of modules of an RBN
7	Modularity	The modularity value of an RBN
8	sInModuleR	The in-module robustness against initial-state perturbation of an RBN
9	rInModuleR	The in-module robustness against update-rule perturbation of an RBN
10	sOutModuleR	The out-module robustness against initial-state perturbation of an RBN
11	rOutModuleR	The out-module robustness against update-rule perturbation of an RBN

(Column description in “net_based_result.txt”)

Network ID	No.Nodes	No.Edges	sRobustness	rRobustness	No.Modules	Modularity	sInModuleR	rInModuleR	sOutModuleR	rOutModuleR
0	10	9	0.40000000	0.81484375	2	0.38888889	1.00000000	1.00000000	1.00000000	1.00000000
1	10	10	0.70000000	0.88125000	4	0.34000000	1.00000000	1.00000000	1.00000000	1.00000000
2	10	11	0.73906250	0.90000000	2	0.39256198	0.89166665	1.00000000	0.90312505	1.00000000
3	10	12	0.80000000	0.90000000	3	0.30902778	1.00000000	1.00000000	1.00000000	1.00000000
4	10	13	0.80000000	0.70000000	3	0.20118343	1.00000000	1.00000000	1.00000000	1.00000000
5	10	14	0.70000000	0.65625000	4	0.23469388	1.00000000	0.97705078	1.00000000	0.97607422
6	10	15	0.66250000	0.70625000	2	0.21333333	0.94835079	0.98524308	0.94531256	0.99218750
7	10	16	0.70000000	0.67421875	2	0.12500000	1.00000000	0.94372392	1.00000000	0.94443357
8	10	17	0.70000000	0.90000000	3	0.14359862	1.00000000	1.00000000	1.00000000	1.00000000
9	10	18	0.95625000	0.60390625	3	0.20370370	1.00000000	0.94189453	1.00000000	0.94091797
10	10	19	0.50000000	0.64531250	2	0.12188366	1.00000000	1.00000000	1.00000000	1.00000000
11	10	20	0.90000000	0.60468750	4	0.21125000	1.00000000	0.88037109	1.00000000	0.87719727
12	10	21	1.00000000	0.40000000	2	0.25396825	1.00000000	0.74945736	1.00000000	0.74207890
13	10	22	1.00000000	0.80000000	2	0.13119835	1.00000000	1.00000000	1.00000000	1.00000000
14	10	23	1.00000000	0.60000000	2	0.15122873	1.00000000	0.93437499	1.00000000	0.93500000
15	10	24	0.91718750	0.40859375	3	0.15277778	0.94357634	0.70572895	0.95703125	0.61111122
16	10	25	0.79609375	0.30000000	3	0.16080000	0.87532550	0.93164062	0.87011719	0.93554688

(Example of “net_based_result.txt”)

APPENDIX B

B1. Parallel robustness computation based on the OpenCL library

I extended the implementation in a previous study (Trinh, et al., 2014) so as to compute in-/out-module robustness. The following figure gives pseudocode describing two important functions, *parallel_computing_attractors_for_all_states* and *parallel_computing_attractors_for_all_rules*, which can compute the attractors in parallel for all initial states (S) and every update rule (F), respectively, given a Boolean network. In computing the attractors, I used an array ATT in which each element $ATT[s, f]$ represents an attractor of a network $G(V, A)$ corresponding to the initial state s and the sequence of update rules f . The algorithm iteratively computes state transitions until it arrives at a state that has already been visited. In the figure, dashed blocks denote kernel codes that are executed *in parallel* on CPUs or GPUs. In other words, the new MORO app computes them in parallel by distributing the tested cases to processing elements in the OpenCL device.

<pre> function <i>ATT</i> parallel_computing_attractors_for_all_states(<i>V, A, f, S</i>) // <i>V, A</i>: A set of nodes $V = \{v_1, v_2, \dots, v_N\}$ and a set of links <i>A</i> of a network (Here, $V[i]$ represents $v_i \in V$.) // <i>f</i>: A sequence of update rules (Here, $f = f_1, f_2 \dots f_N$ and f_i represents the update rule with respect to $v_i \in V$.) // <i>S</i>: A collection of initial states considered for the robustness investigation (Here, $S[i]$ represents i^{th} initial state in <i>S</i>.) // <i>ATT</i>: The resulting collection of attractors each of which is represented by a sequence of states. <i>ATT</i>[0.. $2^{ V }-1$] \leftarrow <i>NULL</i>; // Every element of <i>ATT</i> is initialized by <i>NULL</i>. <i>nth</i>[0.. $2^{ V }-1$] \leftarrow 0; // Every element of <i>nth</i> is initialized by 0. for <i>i</i> \leftarrow 1 to S // for every state do <i>s</i> \leftarrow <i>S</i>[<i>i</i>]; if (<i>ATT</i>[<i>s, f</i>] \neq <i>NULL</i>) continue; endif <i>traj</i> \leftarrow <i>NULL</i>; <i>count</i> \leftarrow 0; while (<i>TRUE</i>) do <i>count</i>++; <i>traj</i> \leftarrow <i>traj</i> \oplus <i>s</i>; // Here \oplus represents the string concatenation operation <i>nth</i>[<i>s</i>] \leftarrow <i>count</i>; <i>s'</i> \leftarrow update_states (<i>V, A, f, s</i>); // This computes the next state. if (<i>nth</i>[<i>s'</i>] \neq 0) do if (<i>ATT</i>[<i>s', f</i>] = <i>NULL</i>) do <i>att</i> \leftarrow <i>traj</i>_{<i>nth</i>[<i>s'</i>].<i>count</i>}; // Given a string $t = t_1 t_2 \dots t_T$, $t_{i..j}$ represents $t_{i+1} \dots t_{j-1} t_j$ which is a substring of <i>t</i>. else <i>att</i> \leftarrow <i>ATT</i>[<i>s', f</i>]; endif for <i>j</i> \leftarrow 1 to <i>count</i> do <i>ATT</i>[<i>traj</i>_{<i>j</i>}, <i>f</i>] = <i>att</i>; endfor break; else <i>s</i> \leftarrow <i>s'</i>; endif endwhile endfor return <i>ATT</i>; end </pre>	<pre> function [<i>ATT</i>] parallel_computing_attractors_for_all_rules (<i>V, A, F, s</i>) // <i>V, A</i>: A set of nodes $V = \{v_1, v_2, \dots, v_N\}$ and a set of links <i>A</i> of a network (Here, $V[i]$ represents $v_i \in V$.) // <i>F</i>: A collection of sequences of update rules (Here, <i>F</i>[<i>i</i>] represents i^{th} sequence of update rules in <i>F</i>) // <i>s</i>: An initial state considered for the robustness investigation // <i>ATT</i>: The resulting collection of attractors each of which is represented by a sequence of states. <i>ATT</i>[0.. $2^{ V }-1$] \leftarrow <i>NULL</i>; // Every element in <i>ATT</i> is initialized by <i>NULL</i> for <i>i</i> \leftarrow 1 to F // for every rule do <i>nth</i>[0.. $2^{ V }-1$] \leftarrow 0; // Every element in <i>nth</i> is initialized by 0. <i>traj</i> \leftarrow <i>NULL</i>; <i>count</i> = 0; while (<i>TRUE</i>) do <i>count</i>++; <i>traj</i> \leftarrow <i>traj</i> \oplus <i>s</i>; // \oplus represents the string concatenation operation. <i>nth</i>[<i>s</i>] \leftarrow <i>count</i>; <i>s'</i> \leftarrow update_states (<i>V, A, F</i>[<i>i</i>], <i>s</i>); // This computes the next state if (<i>nth</i>[<i>s'</i>] \neq 0) do <i>att</i> = <i>traj</i>_{<i>nth</i>[<i>s'</i>].<i>count</i>}; // Given a string $t = t_1 t_2 \dots t_T$, $t_{i..j}$ represents $t_{i+1} \dots t_{j-1} t_j$ which is a substring of <i>t</i>. <i>ATT</i>[<i>s, F</i>[<i>i</i>]] = <i>att</i>; break; else <i>s</i> \leftarrow <i>s'</i>; endif endwhile endfor return <i>ATT</i>; end </pre>
---	---

By using *parallel_computing_attractors_for_all_states* and *parallel_computing_attractors_for_all_rules*, I can compute the in-/out-module robustness of a network \mathbf{G} against initial-state perturbation and update rule perturbation ($\gamma_{in}(\mathbf{G})$ and $\gamma_{out}(\mathbf{G})$, respectively) as shown in the following pseudocode.


```

function [ $\gamma_{in}$   $\gamma_{out}$ ] in-/out-module robustness_initial_state ( $V, A, f, S, M$ )
//  $V, A$ : A set of nodes  $V=\{v_1, v_2, \dots, v_N\}$  and a set of links  $A$  of a network
// (Here,  $V[i]$  represents  $v_i \in V$ )
//  $f$ : A sequence of update rules (Here,  $f = f_1 f_2 \dots f_N$  and  $f_i$  represents the
// update rule with respect to  $v_i \in V$ )
//  $S$ : A collection of initial states considered for the robustness
// investigation (Here,  $S[i]$  represents  $i^{th}$  initial state in  $S$ .)
//  $M$ : A set of modules  $M=\{m_1, m_2, \dots, m_M\}$  of a network after using
// module detection algorithm. In particular, each node will be
// belonged to each module. In other words, each module contains a
// number of nodes of a network.
//  $\gamma_{in}$ : The resulting in-module robustness against initial-state
// perturbations
//  $\gamma_{out}$ : The resulting out-module robustness against initial-state
// perturbations
// Step 1: Examine the original attractors.
 $ATT \leftarrow$  parallel_computing_attractors_for_all_states ( $V, A, f, S$ );
// Step 2: Examine the changed attractors by initial-state perturbations.
for  $i \leftarrow 1$  to  $|S|$ 
     $S'[1..|V|] \leftarrow NULL$ ; // Every element of  $S'$  is initialized by
     $NULL$ .
    for  $j \leftarrow 1$  to  $|V|$ 
         $s \leftarrow S[i]$ ;
         $s_j \leftarrow 1 - s_j$ ; //  $s_j$  denotes the value of  $v_j$  in  $s$ , and then the resultant  $s$ 
        // denotes an initial-state perturbation at a node  $v_j \in V$ .
         $S'[j] \leftarrow s$ ;
    endfor
     $ATT' \leftarrow$  parallel_computing_attractors_for_all_states( $V, A, f, S'$ );
// Step 3: using the Hamming distance measure to examine the similarity
// between four partial attractors
// ( $\langle s_{in} \rangle$  and  $\langle s'_{in} \rangle$ ), ( $\langle s_{out} \rangle$  and  $\langle s'_{out} \rangle$ ) extracted from original attractors and
// attractor by perturbation based on the modular information of each node.
     $\gamma_{in} \leftarrow 0$ ;
     $\gamma_{out} \leftarrow 0$ ;
     $t_{in} \leftarrow 0$ ;  $t'_{in} \leftarrow 0$ ;
     $temp_{out} \leftarrow 0$ ;  $temp'_{out} \leftarrow 0$ ;
    for  $k \leftarrow 1$  to  $|M|$ 
        for  $k \leftarrow 1$  to  $|V_k|$ 
             $t_{in} += H(\langle s_{in} \rangle, \langle s'_{in} \rangle)$ ;
             $temp_{out} += H(\langle s_{out} \rangle, \langle s'_{out} \rangle)$ ;
        endfor
         $t'_{in} += t_{in} / |V_k|$ ;
         $temp'_{out} += temp_{out} / |V_k|$ ;
    endfor
     $\gamma_{in} += t'_{in} / |M|$ ;
     $\gamma_{out} = temp'_{out} / |M|$ ;
endfor
 $\gamma_{in} \leftarrow \gamma_{in} / |S|$ ; // As a result,  $\gamma_{in}$  represents the in-module robustness of the
// given network.
 $\gamma_{out} \leftarrow \gamma_{out} / |S|$ ; // As a result,  $\gamma_{out}$  represents the out-module robustness of the
// given network.
return  $\gamma_{in}$   $\gamma_{out}$ ;
end

```

```

function [ $\gamma_{in}$   $\gamma_{out}$ ] in-/out-module robustness_update_rule ( $V, A, f, S, M$ )
//  $V, A$ : A set of nodes  $V=\{v_1, v_2, \dots, v_N\}$  and a set of links  $A$  of a network
// (Here,  $V[i]$  represents  $v_i \in V$ )
//  $f$ : A sequence of update rules (Here,  $f = f_1 f_2 \dots f_N$  and  $f_i$  represents the
// update rule with respect to  $v_i \in V$ )
//  $S$ : A collection of initial states considered for the robustness
// investigation (Here,  $S[i]$  represents  $i^{th}$  initial state in  $S$ .)
//  $M$ : A set of modules  $M=\{m_1, m_2, \dots, m_M\}$  of a network after using
// module detection algorithm. In particular, each node will be
// belonged to each module. In other words, each module contains a
// number of nodes of a network.
//  $\gamma_{in}$ : The resulting in-module robustness against update-rule
// perturbations
//  $\gamma_{out}$ : The resulting out-module robustness against update-rule
// perturbations
// Step 1: Examine the original attractors.
 $ATT \leftarrow$  parallel_computing_attractors_for_all_states ( $V, A, f, S$ );
// Step 2: Examine the changed attractors by update-rule perturbations.
for  $i \leftarrow 1$  to  $|S|$ 
     $F[1..|V|] \leftarrow NULL$ ; // Every element of  $F$  is initialized by
     $NULL$ .
    for  $j \leftarrow 1$  to  $|V|$ 
         $f' \leftarrow f$ ;
        if ( $f_j = AND$ )  $f'_j \leftarrow OR$ ;
        else  $f'_j \leftarrow AND$ ; //  $f$  means an update-rule perturbation at a node
         $v_j \in V$ .
    endif
     $F[j] \leftarrow f'$ ;
endfor
     $ATT' \leftarrow$  parallel_computing_attractors_for_all_rules ( $V, A, F, S[i]$ );
// Step 3: using the Hamming distance measure to examine the similarity
// between four partial attractors
// ( $\langle s_{in} \rangle$  and  $\langle s'_{in} \rangle$ ), ( $\langle s_{out} \rangle$  and  $\langle s'_{out} \rangle$ ) extracted from original attractors
// and attractor by perturbation based on the modular information of each
// node.
     $\gamma_{in} \leftarrow 0$ ;
     $\gamma_{out} \leftarrow 0$ ;
     $t_{in} \leftarrow 0$ ;  $t'_{in} \leftarrow 0$ ;
     $temp_{out} \leftarrow 0$ ;  $temp'_{out} \leftarrow 0$ ;
    for  $k \leftarrow 1$  to  $|M|$ 
        for  $k \leftarrow 1$  to  $|V_k|$ 
             $t_{in} += H(\langle s_{in} \rangle, \langle s'_{in} \rangle)$ ;
             $temp_{out} += H(\langle s_{out} \rangle, \langle s'_{out} \rangle)$ ;
        endfor
         $t'_{in} += t_{in} / |V_k|$ ;
         $temp'_{out} += temp_{out} / |V_k|$ ;
    endfor
     $\gamma_{in} += t'_{in} / |M|$ ;
     $\gamma_{out} = temp'_{out} / |M|$ ;
endfor
 $\gamma_{in} \leftarrow \gamma_{in} / |S|$ ; // As a result,  $\gamma_{in}$  represents the in-module robustness of the
// given network.
 $\gamma_{out} \leftarrow \gamma_{out} / |S|$ ; // As a result,  $\gamma_{out}$  represents the out-module robustness of
// the given network.
return  $\gamma_{in}$   $\gamma_{out}$ ;
end

```

APPENDIX C

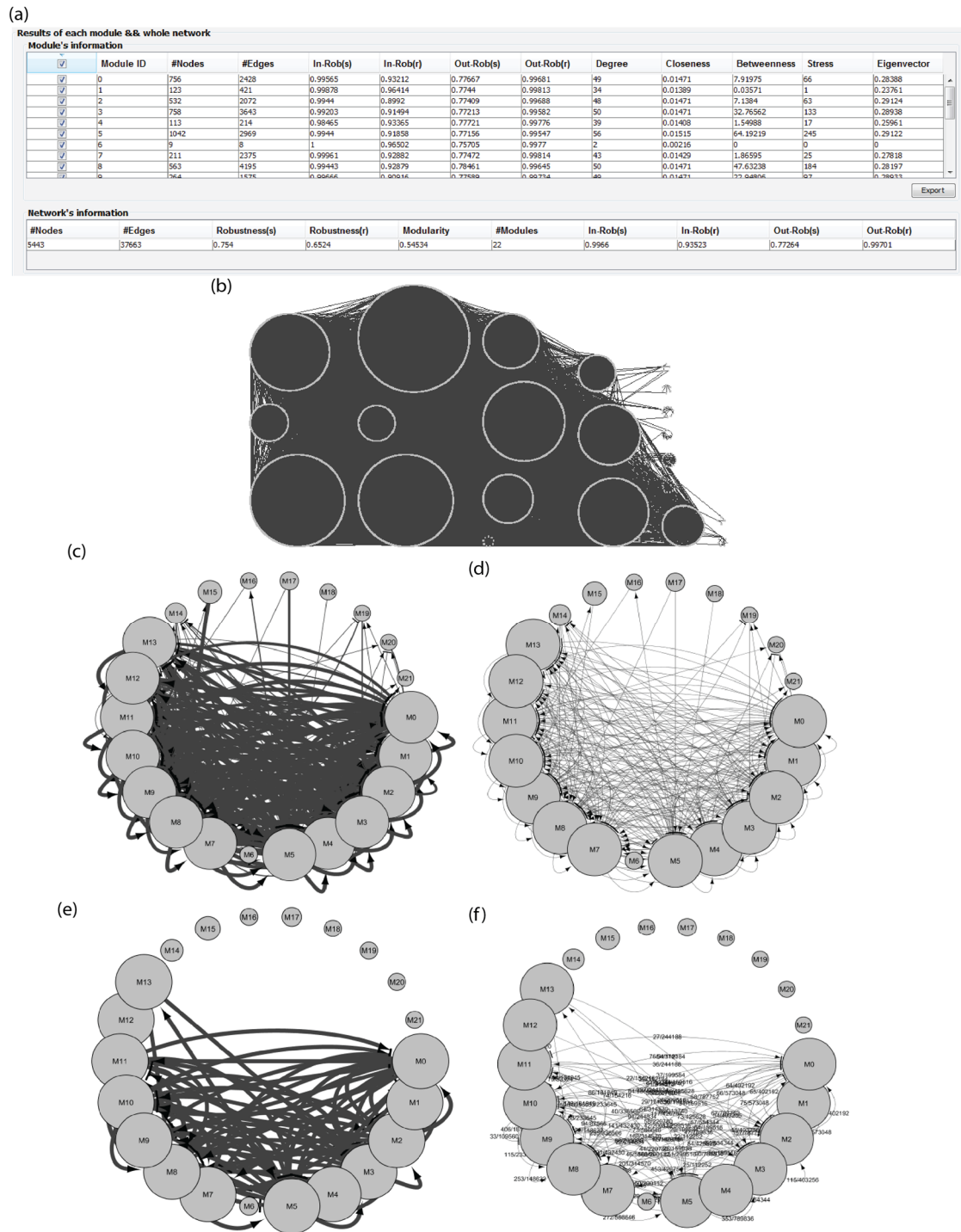


Figure S3.1. Analysis results of the HSN network by MORO.

(see caption of this figure in the next page)

Figure S3.1. Analysis results of the HSN network by MORO.

(a) A summary table. Modularity and robustness results in module and network levels are listed in the upper and the lower tables, respectively. (b) Result of the detailed visualization mode. I found a total of 22 modules each of which is represented by a circular list of genes. (c)-(d) Results of the brief visualization mode with absolute and relative relations, respectively. Each module is represented by a single group node whose radius is proportional to the number of nodes belonging to the module. The weight of a link denotes the number of interactions between the corresponding pair of modules and the ratio of the number of interactions between a pair of modules to the maximal possible number of interactions between them in (c) and (d), respectively. (e)-(f) The reduced visualization results. They are subnetworks induced from (c) and (d), respectively, by removing all links except about 30% of links with the highest weight values (This is performed by specifying the appearance ratio parameter in MORO).

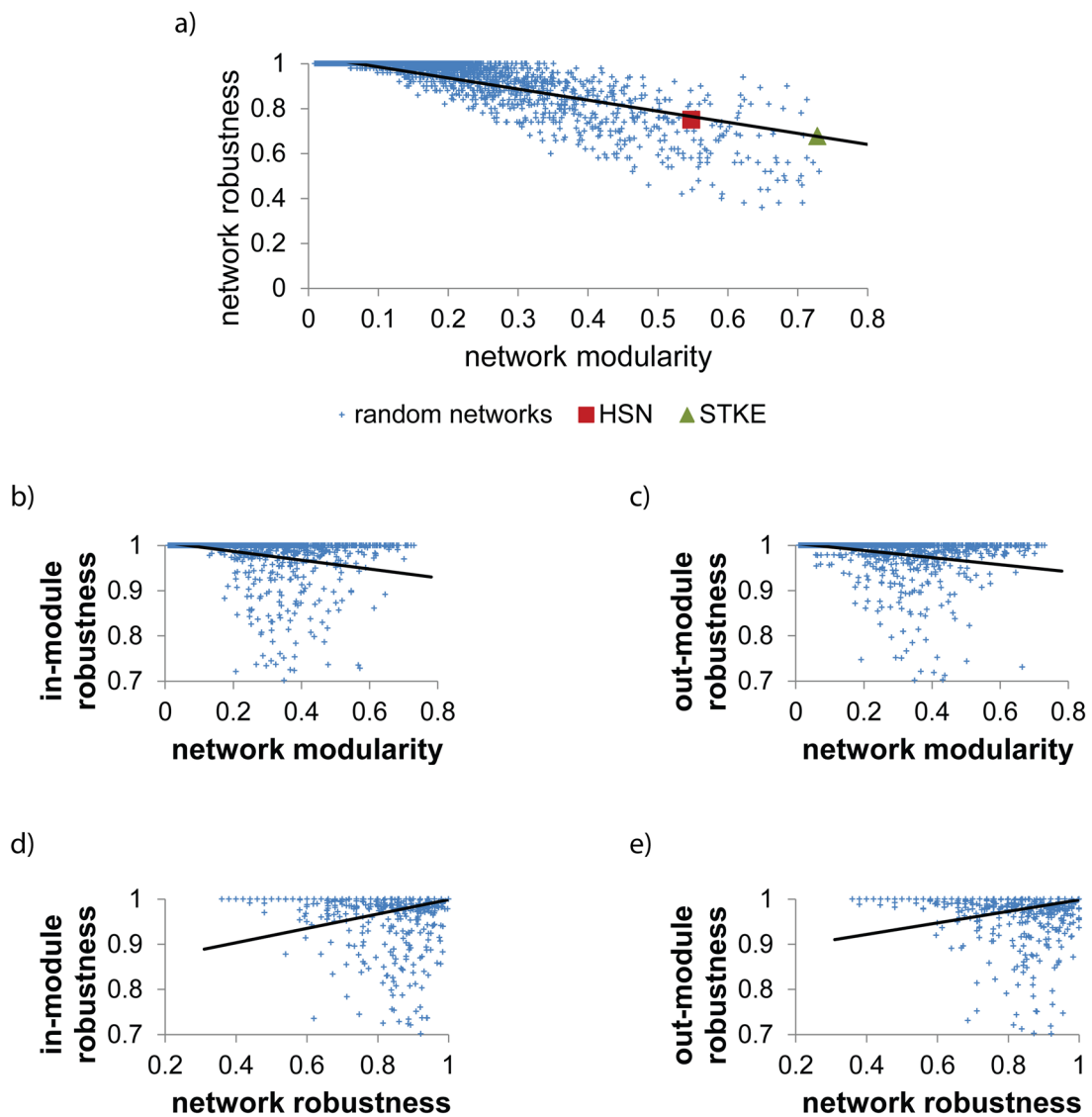


Figure S3.2. Correlations between the modularity and robustness of 6,400 random Boolean networks

Where the number of nodes is 50 and the number of interactions is in the range of [49, 2031].

(a) Relationship between network modularity and robustness: the correlation coefficient was negative (correlation coefficient = -0.80303 with p -value $< 10^{-4}$). The results for HSN and STKE, denoted by the rectangular and triangular points, respectively, were very close to the linear regression line. (b) Relationship of the network modularity to the in-module robustness (correlation coefficient = -0.30383 with p -value $< 10^{-4}$). (c) Relationship between network modularity and out-module robustness (not significant). (d) Relationship of the network robustness to the in-module robustness (correlation coefficient = 0.27801 with p -value $< 10^{-4}$). (e) Relationship between network robustness and out-module robustness (not significant).

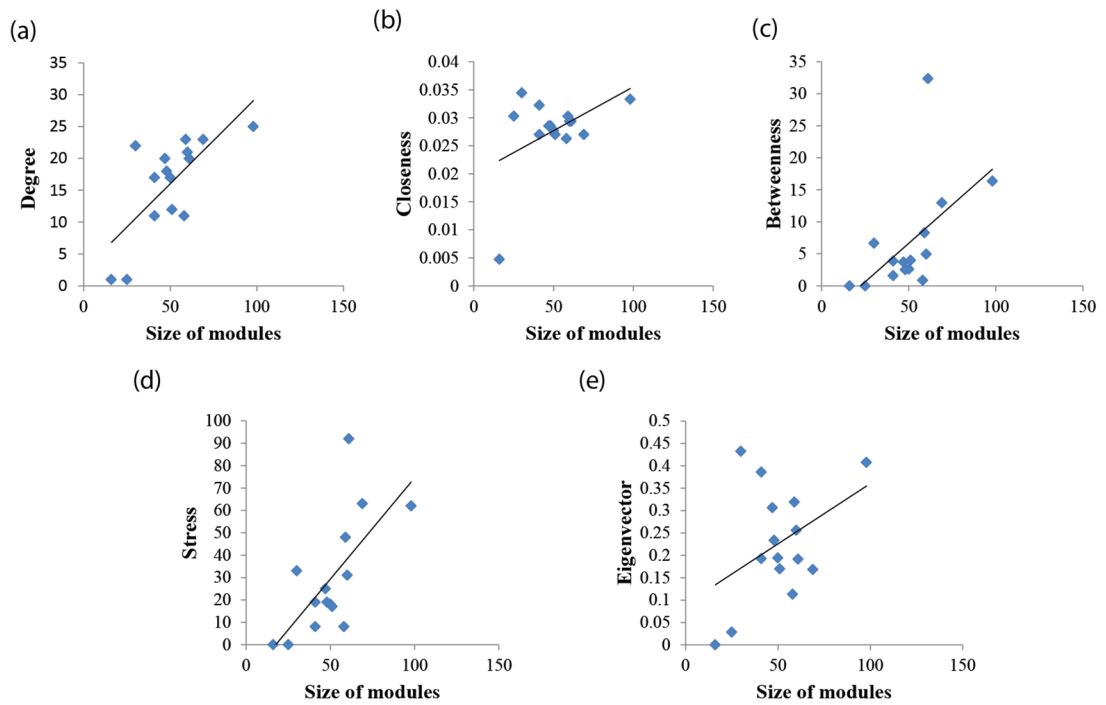


Figure S3.3. Changes of module centrality values against the module size in the HSN network.

(a)-(e) Results with respect to degree, closeness, betweenness, stress, and eigenvector. The module size which is defined as the number of nodes belonging to the module showed positive relationships with all module centrality measures except closeness. The correlation coefficients in (a), (c), (d), and (e) were 0.79367, 0.599553, 0.70063, and 0.870837, respectively, with all p-value $< 10^{-4}$.

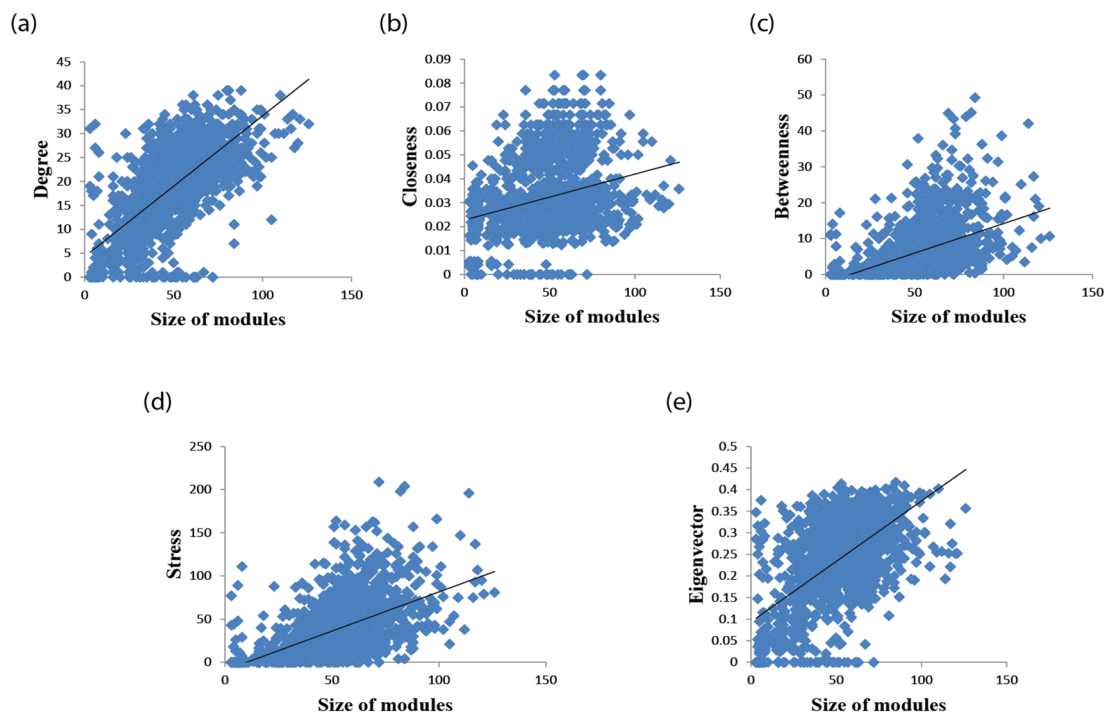


Figure S3.4. Changes of module centrality values against the module size in STKE-shuffled random networks.

I generated 100 random networks by shuffling interactions of the STKE network while preserving a degree distribution. (a)-(e) Results with respect to degree, closeness, betweenness, stress, and eigenvector. The module size which is defined as the number of nodes belonging to the module showed positive relationships with all module centrality measures. The correlation coefficients in (a), (b), (c), (d), and (e) were 0.70475, 0.26639, 0.50143, 0.57625, and 0.58761, respectively, with all p-value $< 10^{-4}$.

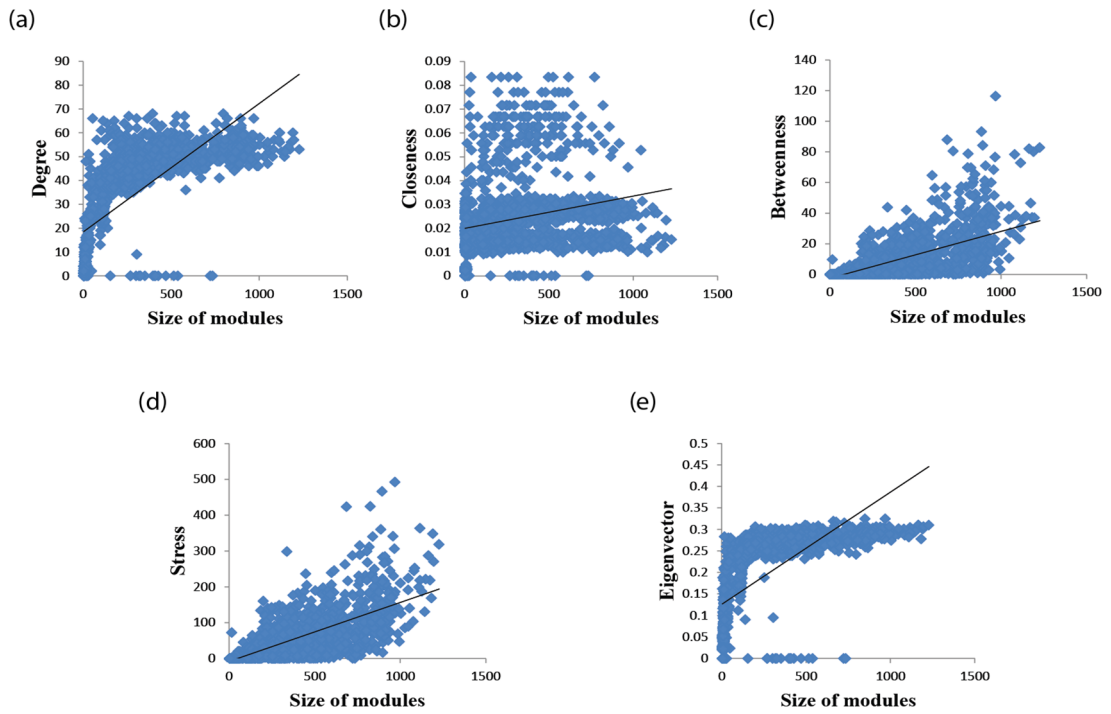


Figure S3.5. Changes of module centrality values against the module size in HSN-shuffled random networks.

I generated 100 random networks by shuffling interactions of the HSN network while preserving a degree distribution. (a)-(e) Results with respect to degree, closeness, betweenness, stress, and eigenvector. The module size which is defined as the number of nodes belonging to the module showed positive relationships with all module centrality measures. The correlation coefficients in (a), (b), (c), (d), and (e) were 0.73027, 0.24344, 0.66850, 0.74306, and 0.67059, respectively, with all p-value $< 10^{-4}$.

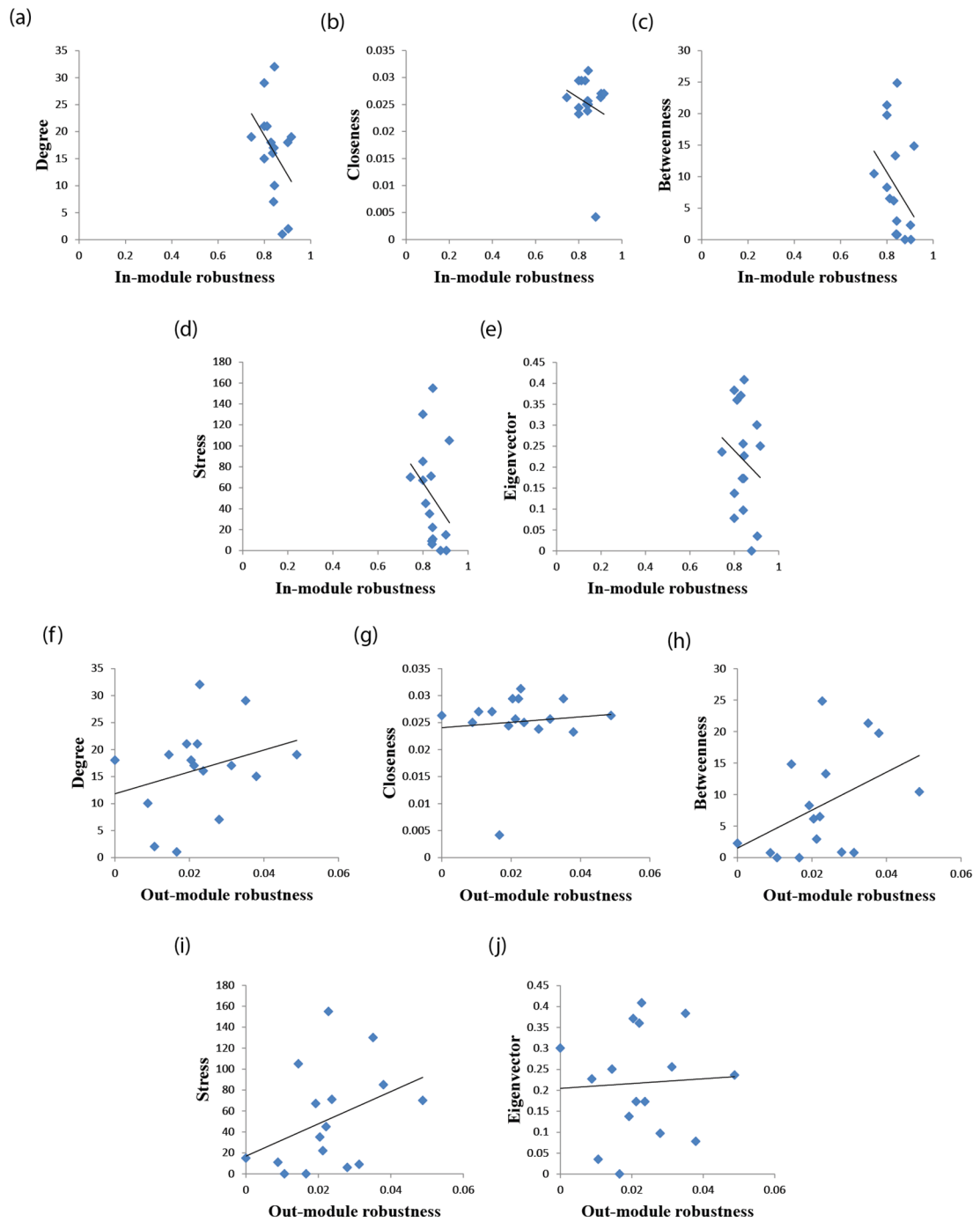


Figure S3.6. Correlation between module centrality values and in-/out-module robustness in the STKE network.

(a)-(e) Correlations of in-module robustness with degree, closeness, betweenness, stress, and eigenvector, respectively. There was no significant relation (all p-values > 0.13279). (f)-(j) Correlations of out-module robustness with degree, closeness, betweenness, stress, and eigenvector, respectively. There was no significant relation (all p-values > 0.09143).

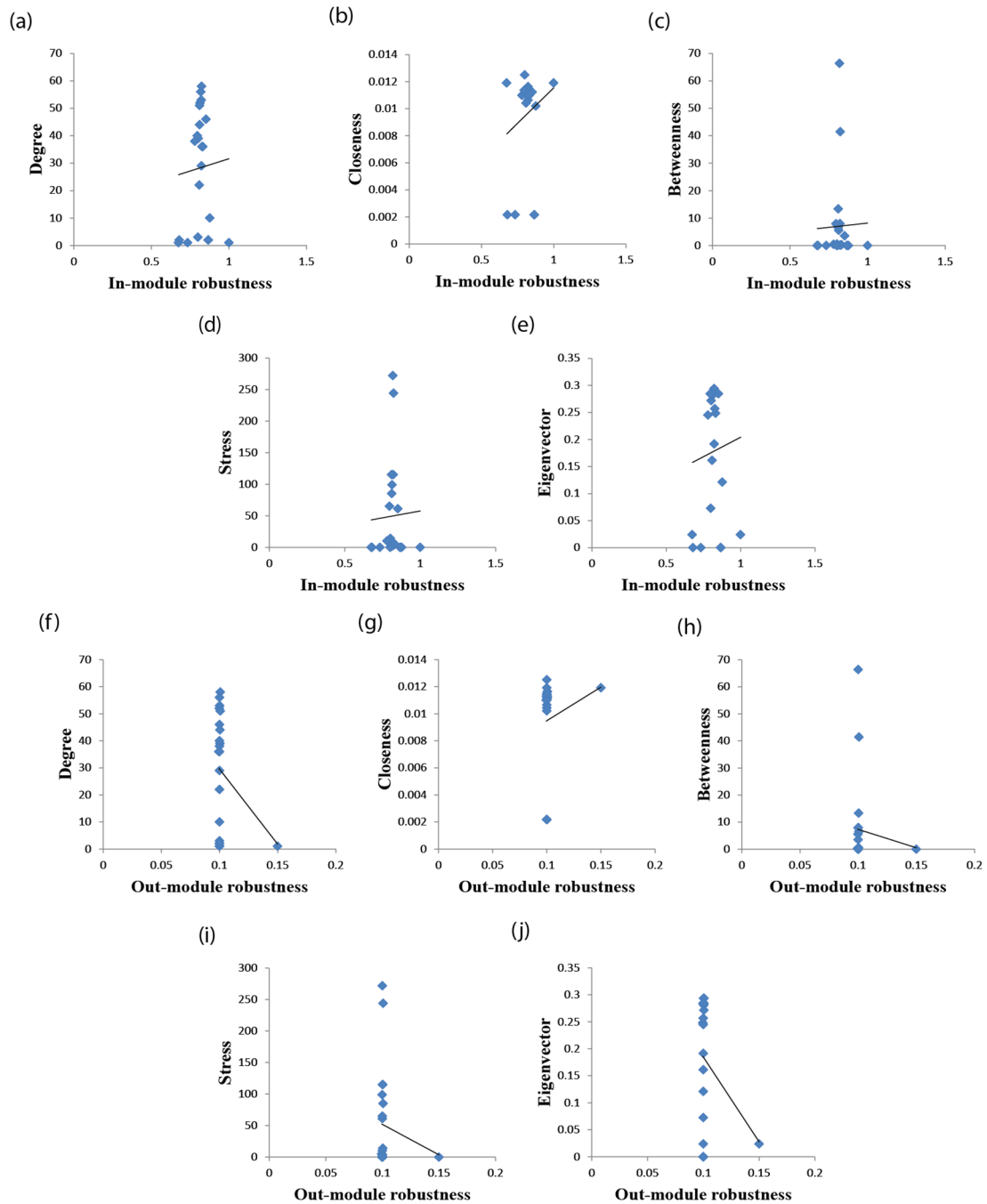


Figure S3.7. Correlation between module centrality values and in/out-module robustness in the HSN network.

(a)-(e) Correlations of in-module robustness with degree, closeness, betweenness, stress, and eigenvector, respectively. There was no significant relation (all p-values > 0.39269). (f)-(j) Correlations of out-module robustness with degree, closeness, betweenness, stress, and eigenvector, respectively. There was no significant relation (all p-values > 0.21193).

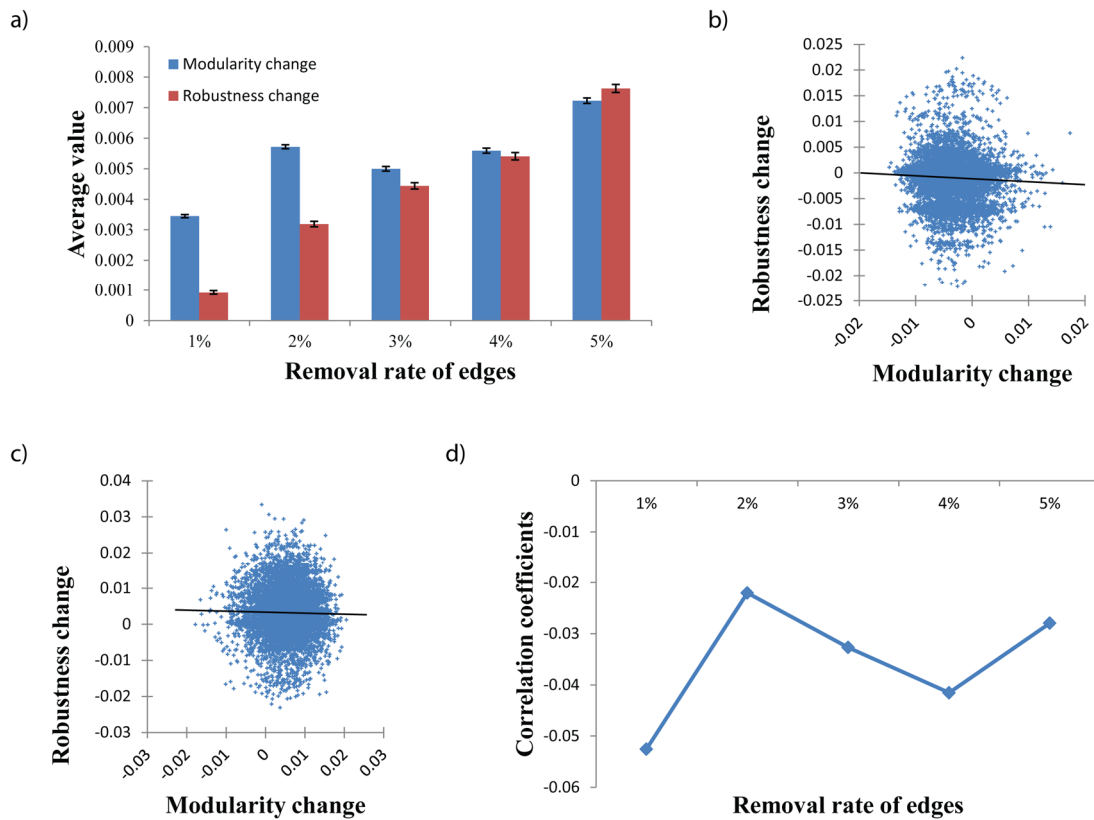


Figure S4.1. Analysis of the changes of the modularity and the robustness by edge-removal mutations in STF signaling network.

The removal rate of edges was varied from 1% to 5% (More specifically, the numbers of removed edges were 5, 11, 16, 22, and 27, respectively, among a total of 557 edges). For each removal rate, 5,000 trials of edge-removal were examined. **(a)** Results of average changes of the modularity and the robustness against the removal rate of edges. Y-axis value and error bar represent the average and the standard deviation divided by the square root of the sample size (5000), respectively. Both average values were significantly larger than zero (All P-values <0.0001 , using one-sample t-test). The one-sample t-test was valid because the average values were normally distributed (see Appendix C: Fig. S4.4) and there were very few outliers (see Appendix C: Fig. S4.7). **(b)-(c)** Relationship between the changes of the modularity and the robustness in the case that the removal rate is 1% and 2%, respectively. A significant negative relationship was observed (Correlation coefficients were -0.05254 and -0.022068, respectively, with all P-values <0.0001). This relationship was consistently observed for larger removal rates (Correlation coefficients when the removal rate of edge is 3%, 4%, and 5% were -0.03272 and, -0.04156, and -0.02795, respectively, with all P-values <0.0001). **(d)** A trend of correlation coefficients between the changes of the modularity and the robustness against the removal rate of edges.

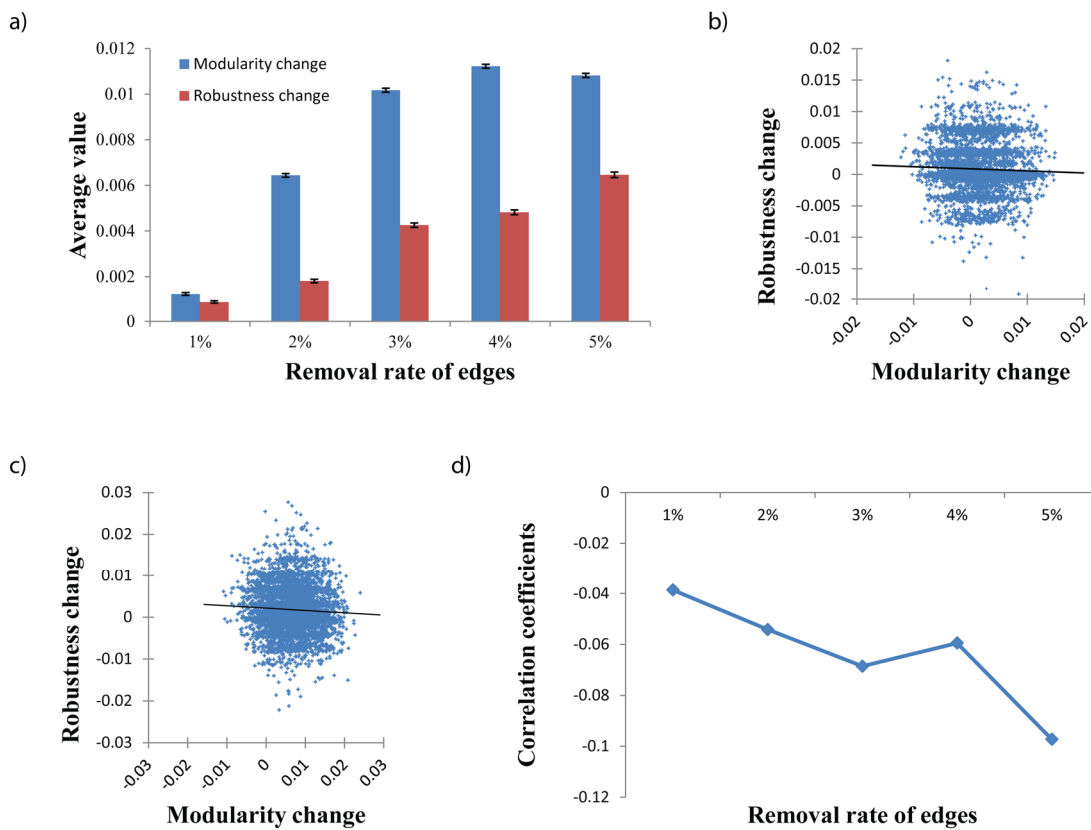


Figure S4.2. Analysis of the changes of the modularity and the robustness by edge-removal mutations in HIV-1 signaling network.

The removal rate of edges was varied from 1% to 5% (More specifically, the numbers of removed edges were 3, 7, 11, 14, and 18, respectively, among a total of 368 edges). For each removal rate, 5,000 trials of edge-removal were examined. **(a)** Results of average changes of the modularity and the robustness against the removal rate of edges. Y-axis value and error bar represent the average and the standard deviation divided by the square root of the sample size (5000), respectively. Both average values were significantly larger than zero (All P-values <0.0001 using one-sample t-test). The one-sample t-test was valid because the average values were normally distributed (see Appendix C: Fig. S4.5) and there were very few outliers (see Appendix C: Fig. S4.8). **(b)-(c)** Relationship between the changes of the modularity and the robustness in the case that the removal rate is 1% and 2%, respectively. A significant negative relationship was observed (Correlation coefficients were -0.03867 and -0.05417, respectively with all P-values <0.0001). This relationship was consistently observed for larger removal rates (Correlation coefficients when the edge-removal rate is 3%, 4%, and 5% were -0.06862 and -0.05948, and -0.09733, respectively, with all P-values <0.0001). **(d)** A trend of correlation coefficients between the changes of the modularity and the robustness against the removal rate of edges.

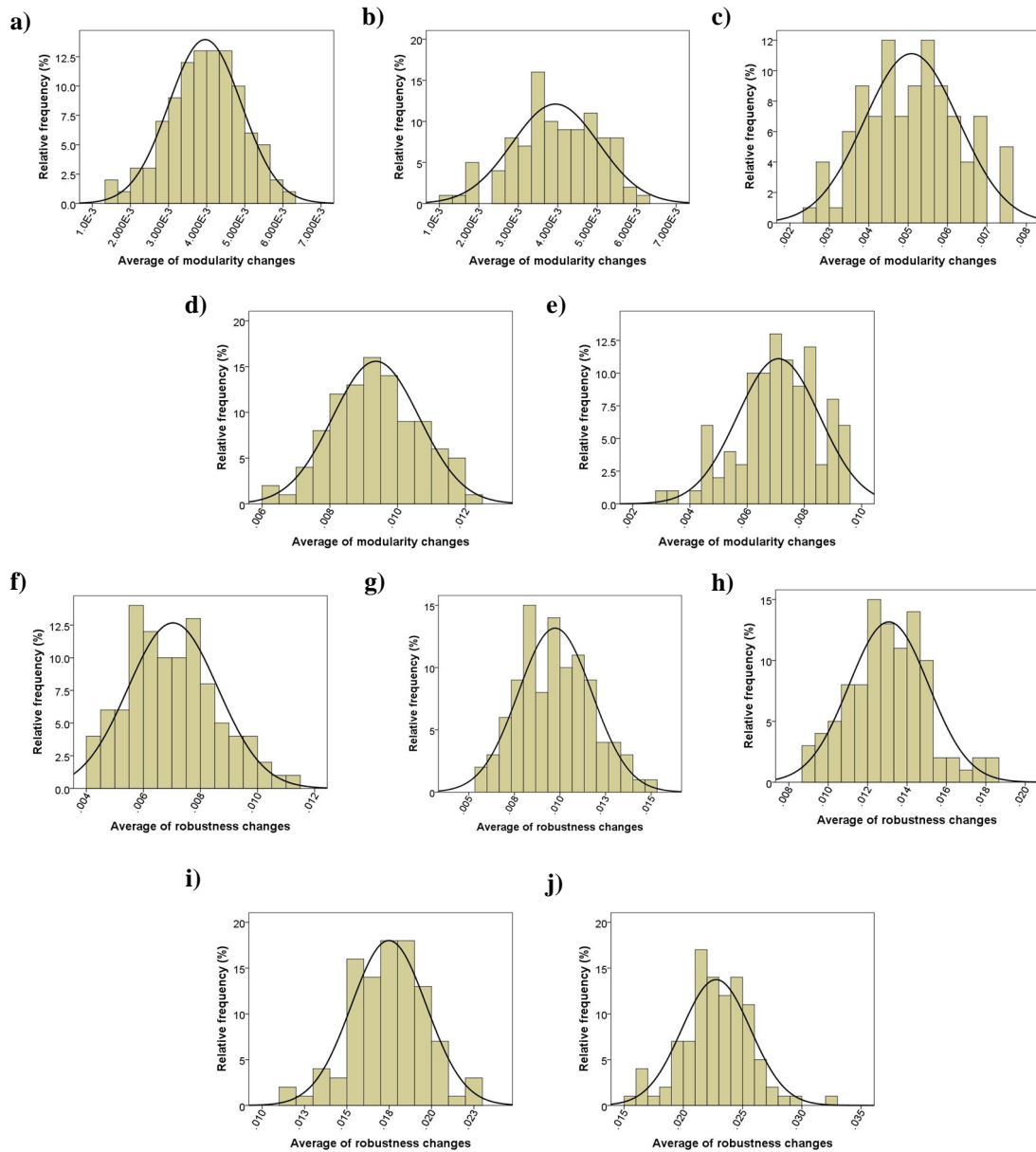


Figure S4.3. Analysis of normal distributions of averages of modularity changes and robustness changes in T-LGL network.

(a)-(e) Results of the average of the modularity change with removal rates of 1% to 5%, respectively. (f)-(j) Results of the average of the robustness change with removal rates of 1% to 5%, respectively. In each subfigure, the average of the modularity or the robustness changes over 50 trials is computed, and this process was repeated 100 times to examine the distribution of the average variable. Kolmogorov-Smirnov test was run and all average values were normally distributed (All P-values > 0.10).

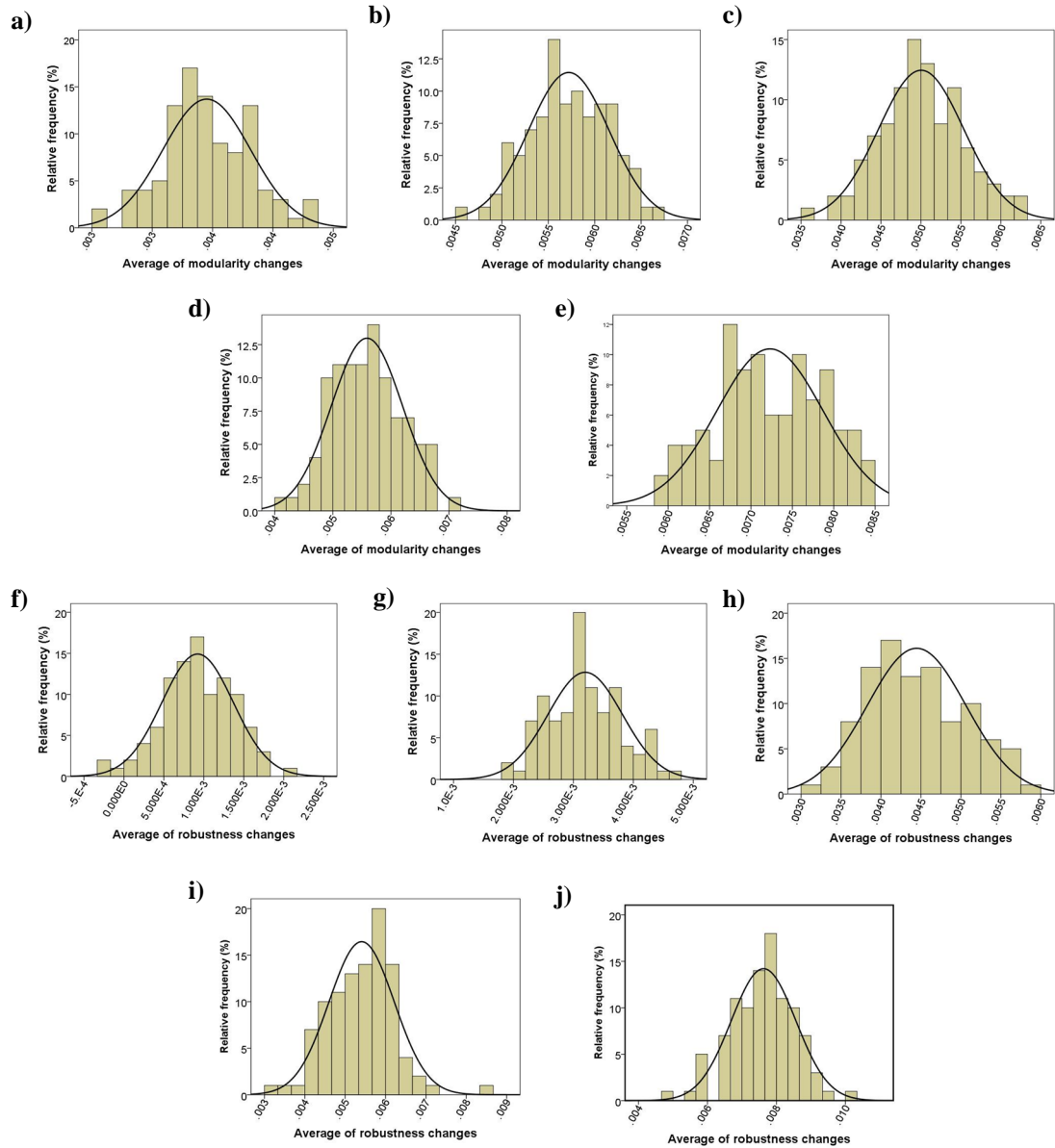


Figure S4.4. Analysis of normal distributions of averages of modularity changes and robustness changes in STF network.

(a)-(e) Results of the average of the modularity change with removal rates of 1% to 5%, respectively. (f)-(j) Results of the average of the robustness change with removal rates of 1% to 5%, respectively. In each subfigure, the average of the modularity or the robustness changes over 50 trials is computed, and this process was repeated 100 times to examine the distribution of the average variable. Kolmogorov-Smirnov test was run and all average values were normally distributed (All P-values > 0.10 except that P-value=0.069 in (e)).

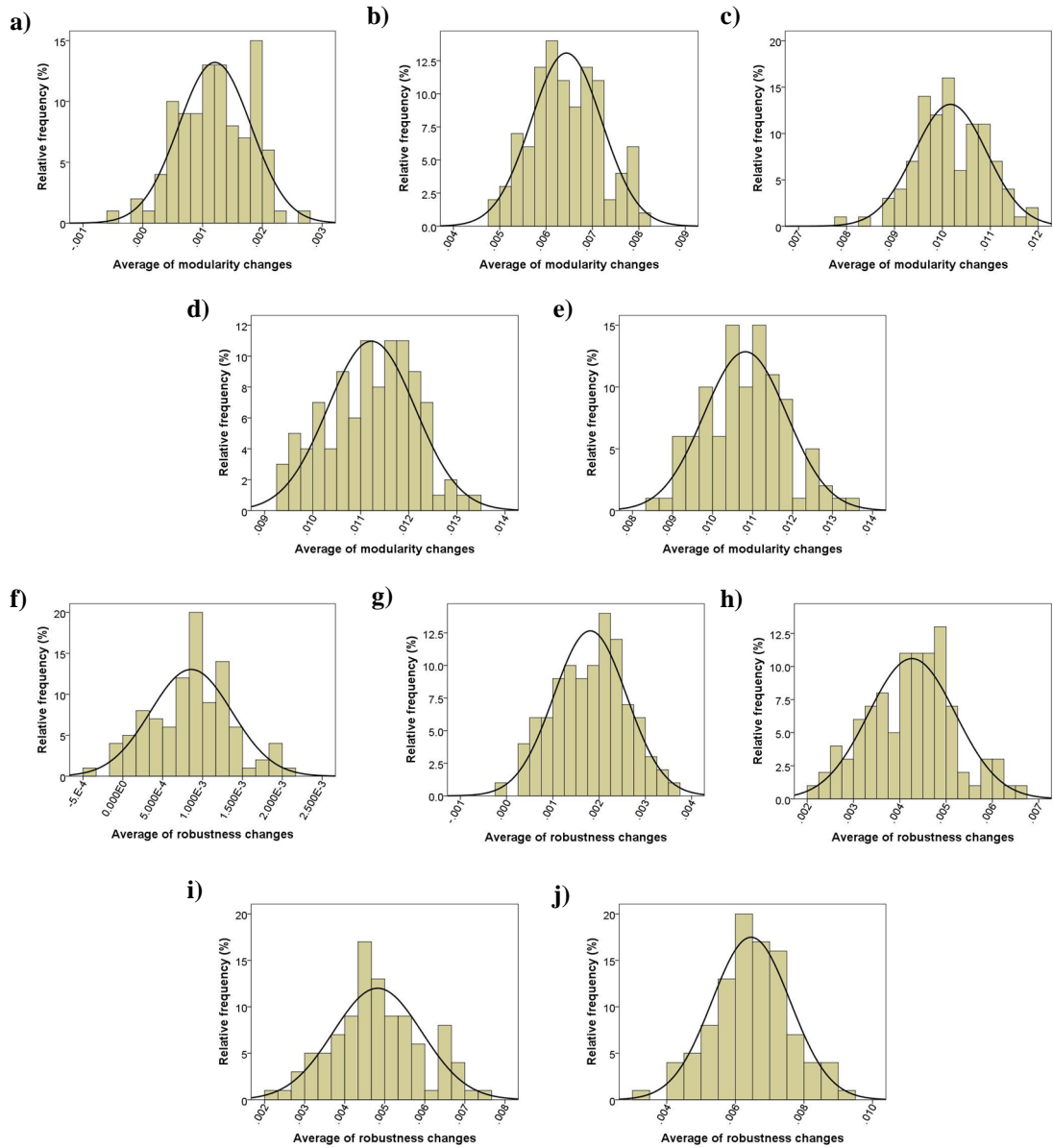


Figure S4.5. Analysis of normal distributions of averages of modularity changes and robustness changes in HIV-1 network.

(a)-(e) Results of the average of the modularity change with removal rates of 1% to 5%, respectively. (f)-(j) Results of the average of the robustness change with removal rates of 1% to 5%, respectively. In each subfigure, the average of the modularity or the robustness changes over 50 trials is computed, and this process was repeated 100 times to examine the distribution of the average variable. Kolmogorov-Smirnov test was run and all average values were normally distributed (All P-values > 0.10 except that P-value=0.053 in (d)).

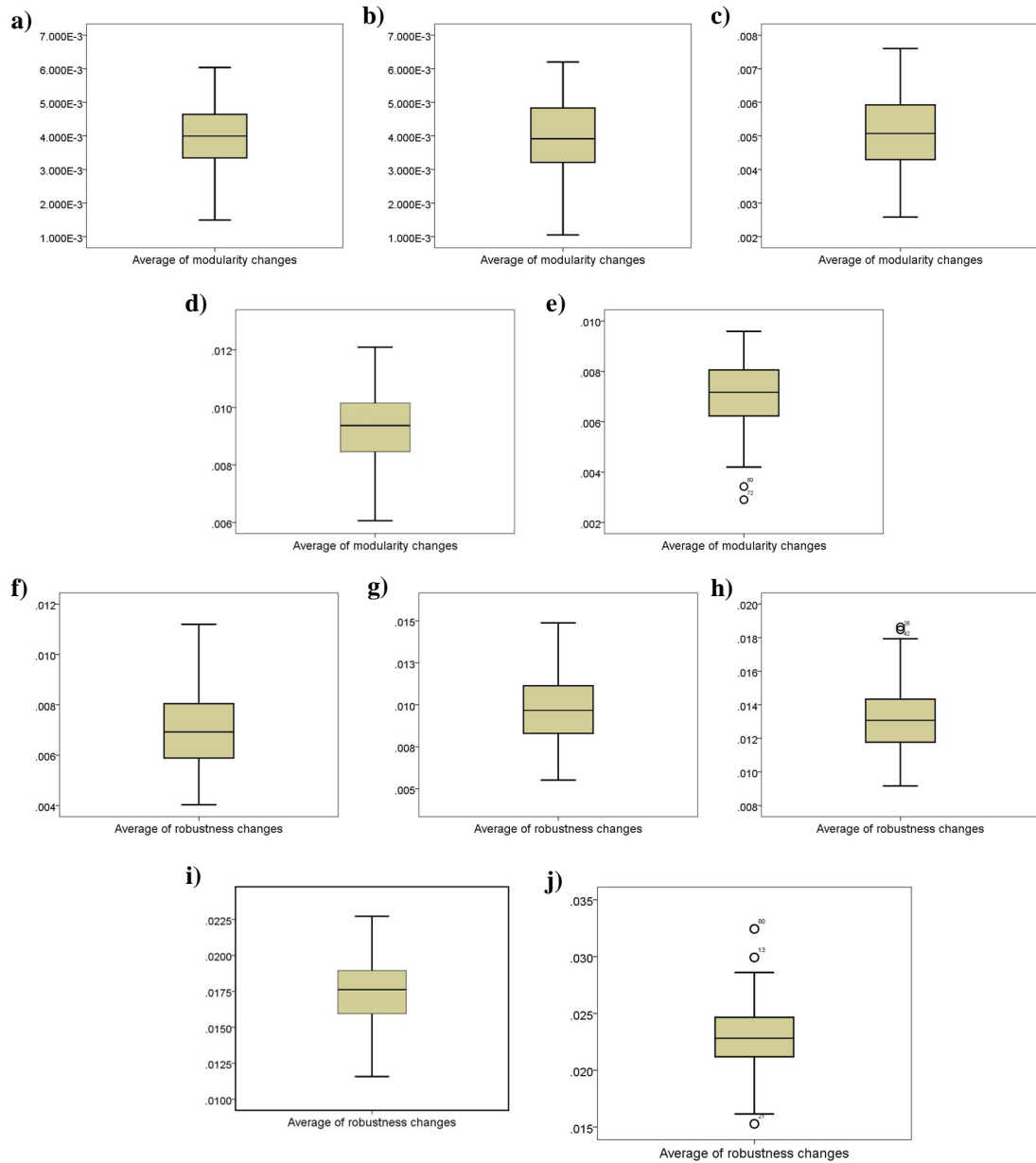


Figure S4.6. Analysis of outliers of averages of modularity changes and robustness changes in T-LGL network.

(a)-(e) Results of the average of the modularity change with removal rates of 1% to 5%, respectively. (f)-(j) Results of the average of the robustness change with removal rates of 1% to 5%, respectively. In each subfigure, the average of the modularity or the robustness changes over 50 trials is computed, and this process was repeated 100 times to examine the distribution of the average variable. I examined outliers by a boxplot inspection and found no significant outliers in all subfigures except for (e), (h) and (j).

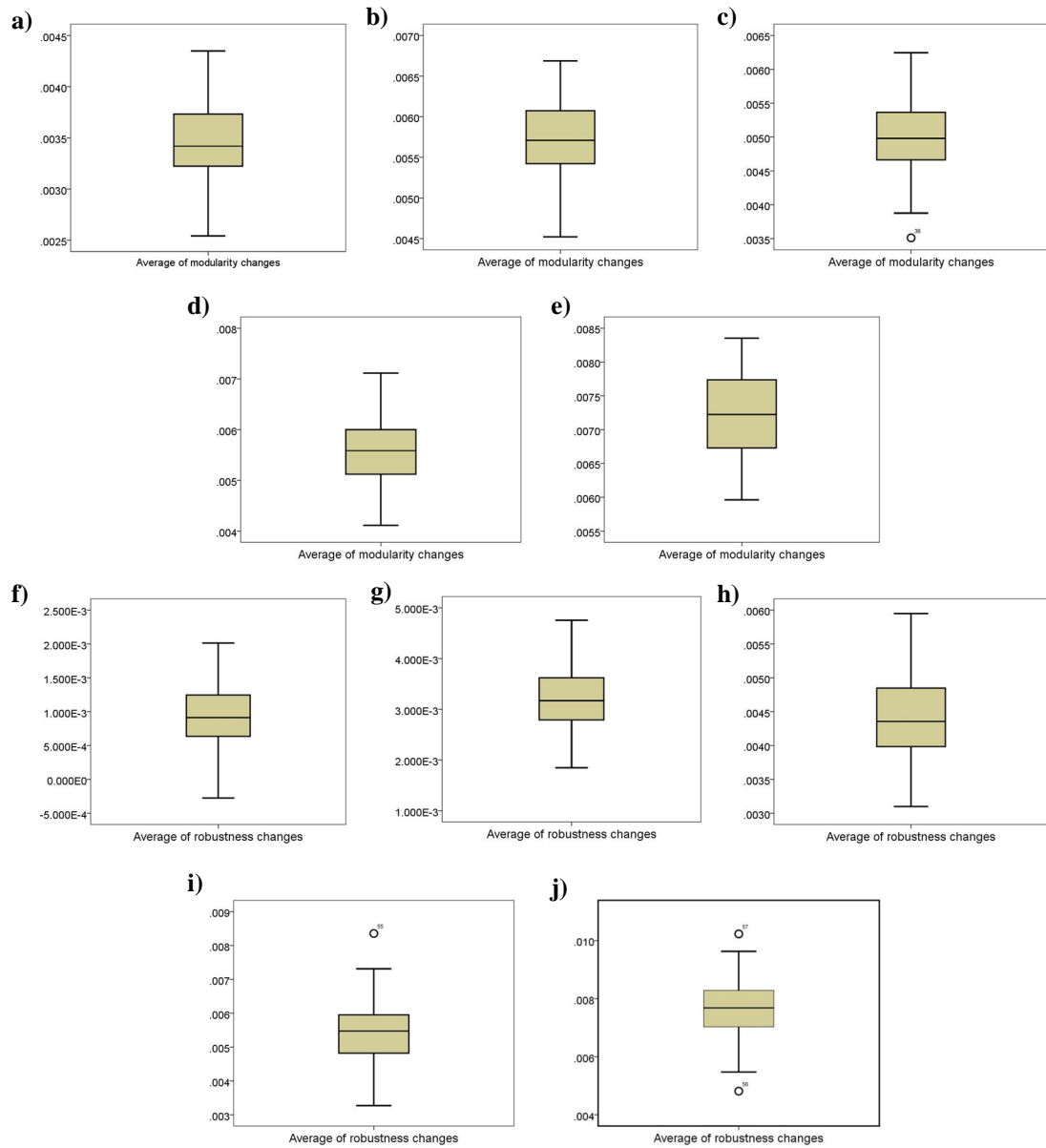


Figure S4.7. Analysis of outliers of averages of modularity changes and robustness changes in STF network.

(a)-(e) Results of the average of the modularity change with removal rates of 1% to 5%, respectively. (f)-(j) Results of the average of the robustness change with removal rates of 1% to 5%, respectively. In each subfigure, the average of the modularity or the robustness changes over 50 trials is computed, and this process was repeated 100 times to examine the distribution of the average variable. I examined outliers by a boxplot inspection and found no significant outliers in all subfigures except for (c), (i) and (j).

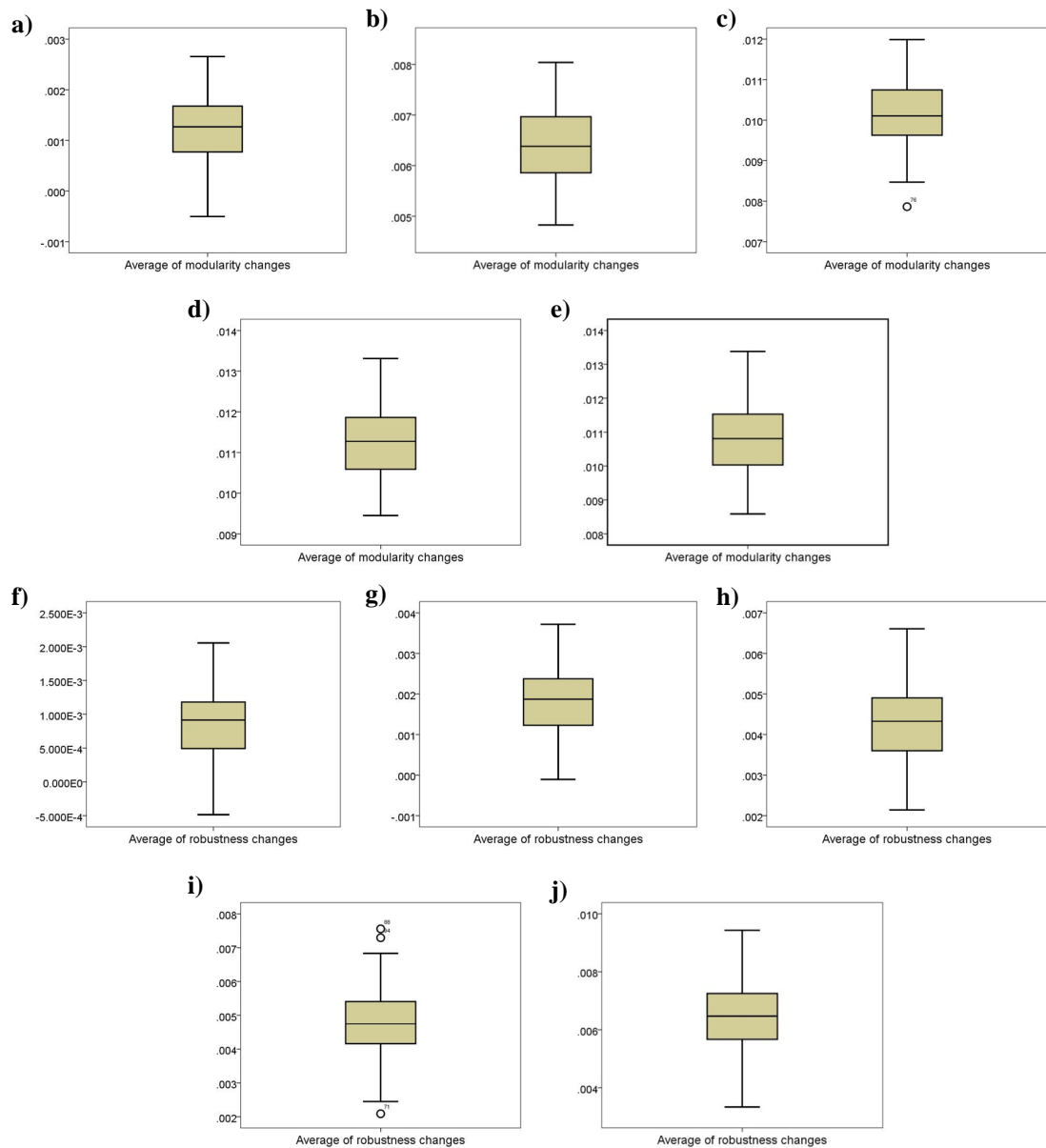


Figure S4.8. Analysis of outliers of averages of modularity changes and robustness changes in HIV-1 network.

(a)-(e) Results of the average of the modularity change with removal rates of 1% to 5%, respectively. (f)-(j) Results of the average of the robustness change with removal rates of 1% to 5%, respectively. In each subfigure, the average of the modularity or the robustness changes over 50 trials is computed, and this process was repeated 100 times to examine the distribution of the average variable. I examined outliers by a boxplot inspection and found no significant outliers in all subfigures except for (c) and (i).

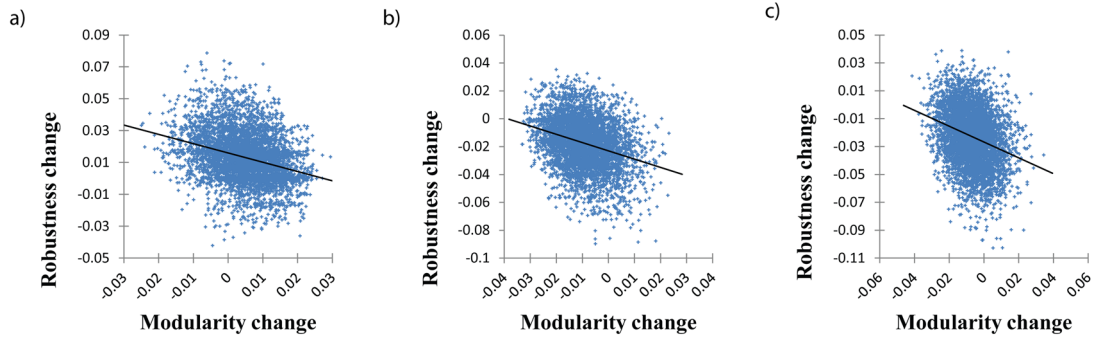


Figure S4.9. Relationship between the changes of the modularity and the robustness in T-LGL signaling network.

(a)-(c) Results in the cases that the edge-removal rate is 3%, 4%, and 5%, respectively. Each plot is a result of 5,000 trials. Correlation coefficients of (a)-(c) were -0.30652, -0.30684, and -0.28626, respectively (All P-values <0.0001).

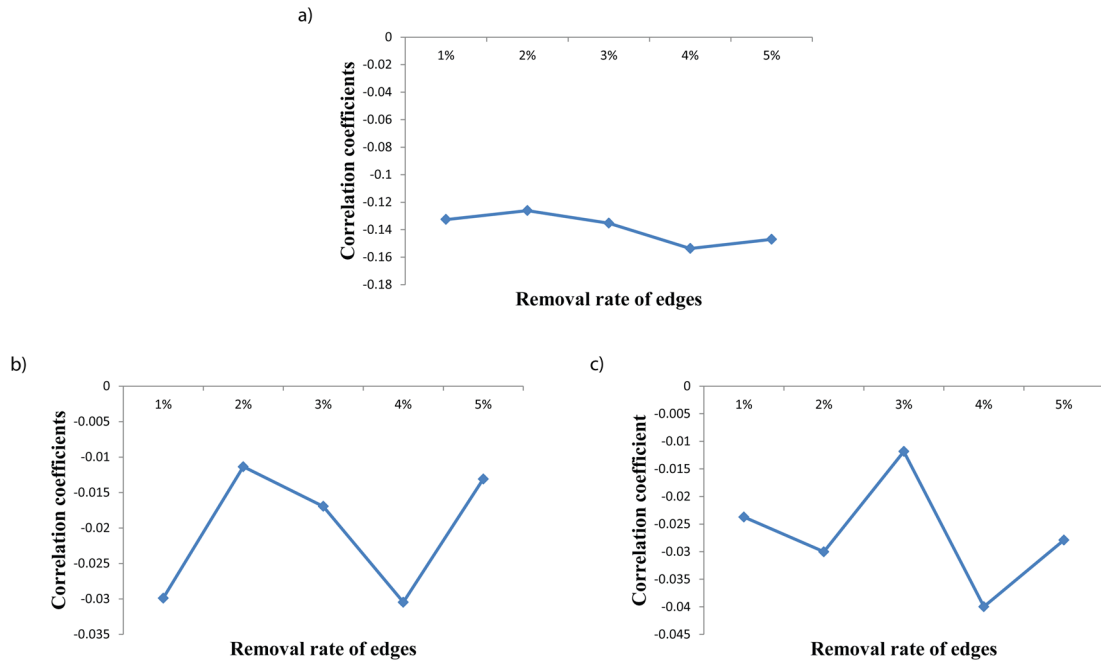


Figure S4.10. Relationship between the changes of the modularity and the robustness in random networks.

(a)-(c) Results of the random networks shuffled from T-LGL, STF, and HIV-1 signaling networks, respectively. In each subfigure, a set of 100 random networks were generated and 500 trials of edge-removals were tested for each network (Hence, each correlation coefficient was obtained over a total of 50,000 samples). I analyzed the relationship by varying the edge-removal rate from 1% to 5%. All cases showed significantly negative relationship (All P-values < 0.0001).

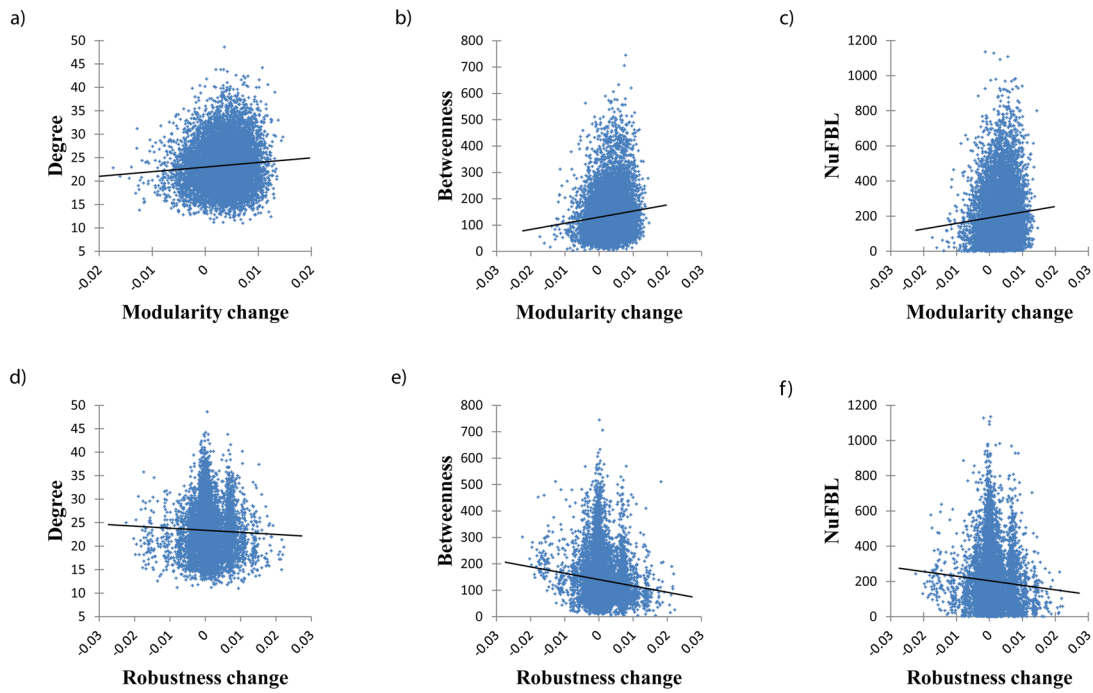


Figure S4.11. Relationship of each of the changes of the modularity and the robustness with the structural properties in STF signaling network.

The removal rate was set 1%, and a total of 5,000 trials of removals were examined. **(a)-(c)** Relations of the change of modularity with edge-based degree, EBEW, and NuFBL, respectively. The change of modularity was significantly positively correlated with all structural properties (Correlation coefficients were 0.07621, 0.10084, and 0.07762, respectively, with all P-values<0.0001). **(d)-(f)** Relations of the change of robustness with edge-based degree, EBEW, and NuFBL, respectively. The change of robustness was significantly negatively correlated with all structural properties (Correlation coefficients were -0.03749, -0.11430, and -0.06860, respectively, with all P-values<0.0001).

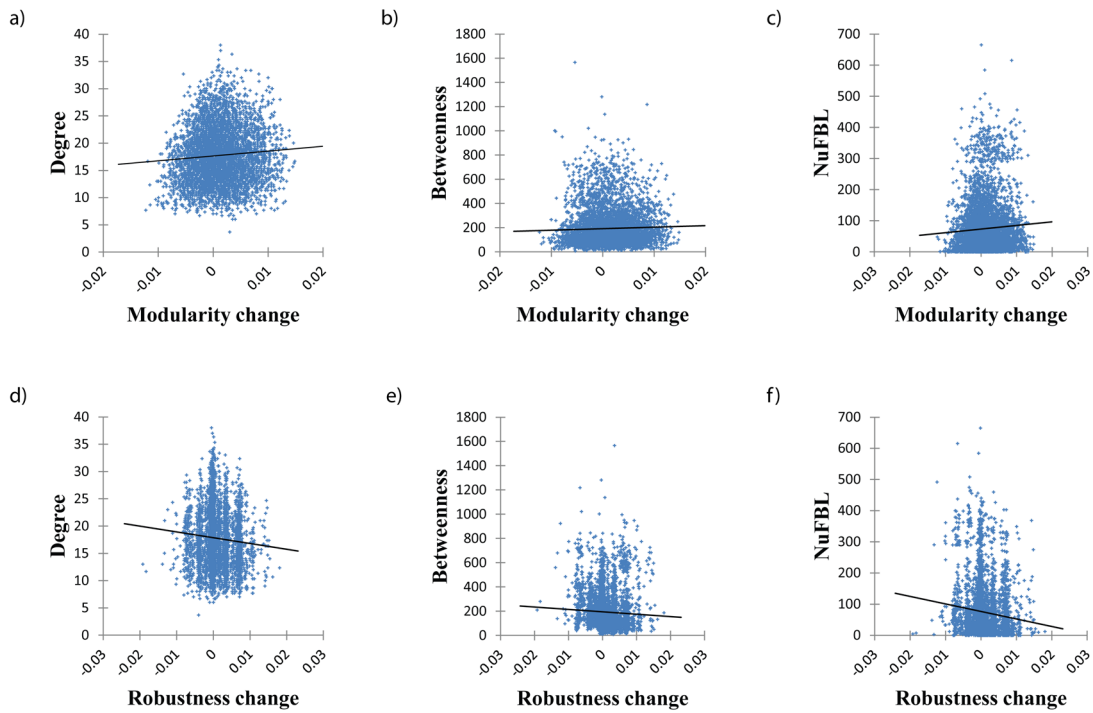


Figure S4.12. Relationship of each of the changes of the modularity and the robustness with the structural properties in HIV-1 signaling network.

The removal rate was set 1%, and a total of 5,000 trials of removals were examined. **(a)-(c)** Relations of the change of modularity with edge-based degree, EBEW, and NuFBL, respectively. The change of modularity was significantly positively correlated with all structural properties (Correlation coefficients were 0.07549, 0.03526, and 0.05970, respectively, with all P-values<0.0001). **(d)-(f)** Relations of the change of robustness with edge-based degree, EBEW, and NuFBL, respectively. The change of robustness was significantly negatively correlated with all structural properties (Correlation coefficients were -0.07664, -0.04671, and -0.10621, respectively, with all P-values<0.0001).

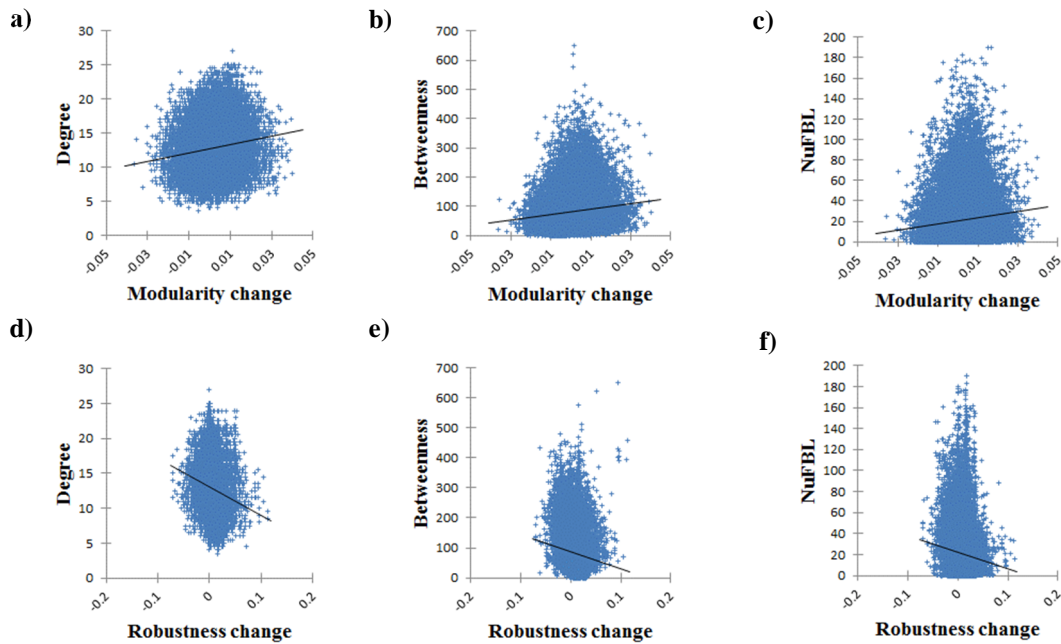


Figure S4.13. Relationship of each of the changes of the modularity and the robustness with the structural properties in random networks shuffled from T-LGL network.

In each subfigure, a set of 100 random networks were generated and 500 trials of edge-removal were tested for each network (Hence, each correlation coefficient was obtained over a total of 50,000 samples). The removal rate was set 1%. **(a)-(c)** Relations of the change of modularity with edge-based degree, EBEW, and NuFBL, respectively. The change of modularity was significantly positively correlated with all structural properties (Correlation coefficients were 0.14388, 0.11878, and 0.10775, respectively, with all P-values<0.0001). **(d)-(f)** Relations of the change of robustness with edge-based degree, EBEW, and NuFBL, respectively. The change of robustness was significantly negatively correlated with all structural properties (Correlation coefficients were -0.14851, -0.11281, and -0.08690, respectively, with all P-values<0.0001).

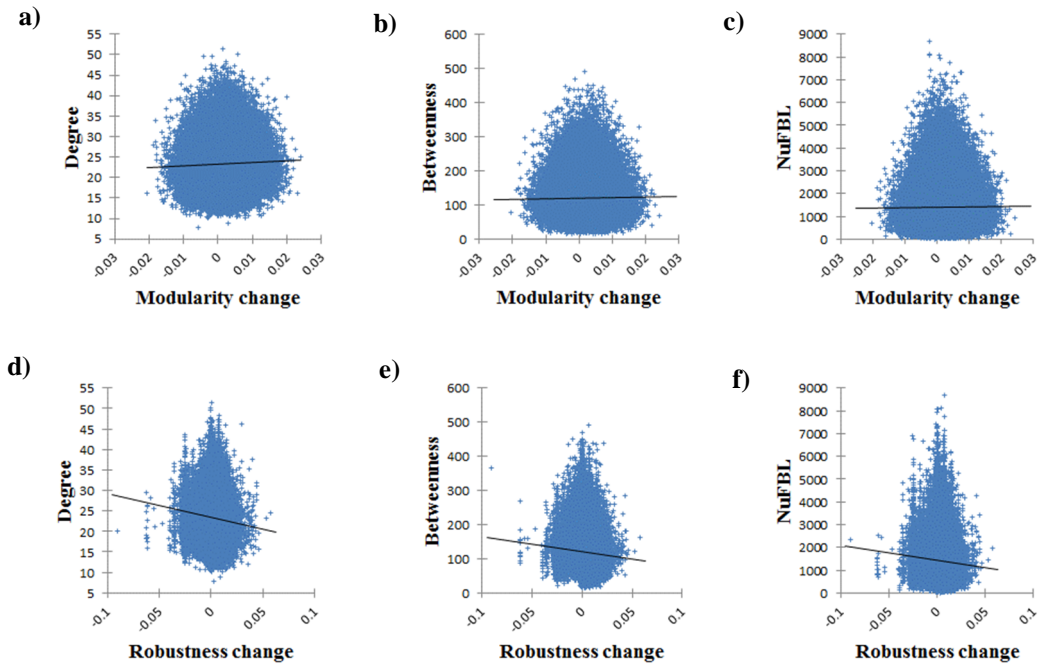


Figure S4.14. Relationship of each of the changes of the modularity and the robustness with the structural properties in random networks shuffled from STF network.

In each subfigure, a set of 100 random networks were generated and 500 trials of edge-removal were tested for each network (Hence, each correlation coefficient was obtained over a total of 50,000 samples). The removal rate was set 1%. **(a)-(c)** Relations of the change of modularity with edge-based degree, EBEW, and NuFBL, respectively. The change of modularity was significantly positively correlated with all structural properties (Correlation coefficients were 0.03228, 0.01516, and 0.01062, respectively, with all P-values<0.0001). **(d)-(f)** Relations of the change of robustness with edge-based degree, EBEW, and NuFBL, respectively. The change of robustness was significantly negatively correlated with all structural properties (Correlation coefficients were -0.07121, -0.05301, and -0.04716, respectively, with all P-values<0.0001).

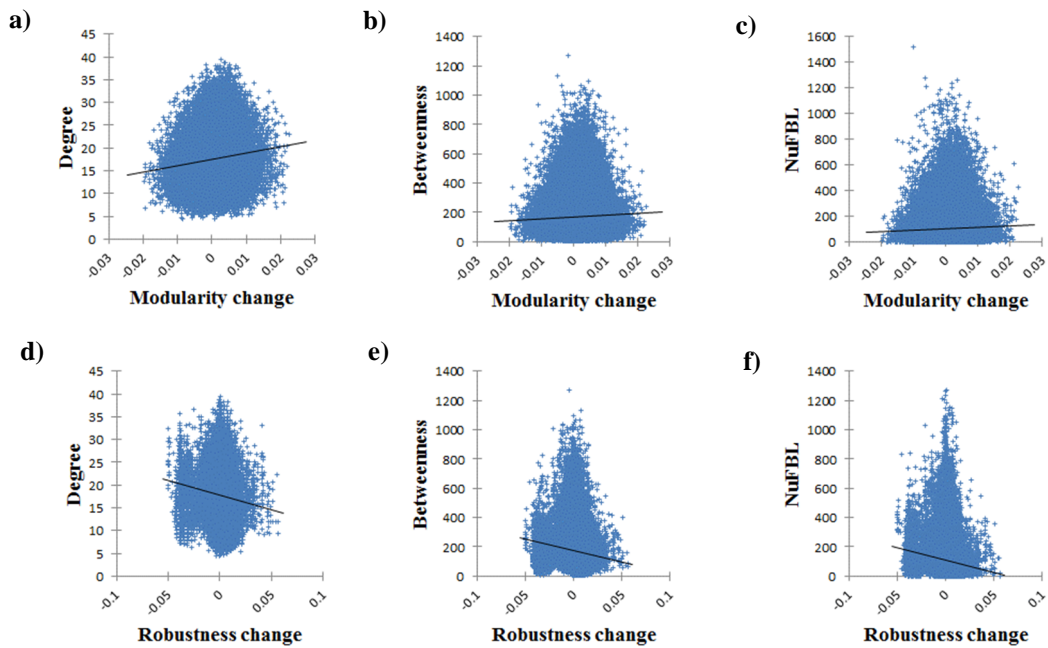


Figure S4.15. Relationship of each of the changes of the modularity and the robustness with the structural properties in random networks shuffled from HIV-1 network.

In each subfigure, a set of 100 random networks were generated and 500 trials of edge-removal were tested for each network (Hence, each correlation coefficient was obtained over a total of 50,000 samples). The removal rate was set 1%. **(a)-(c)** Relations of the change of modularity with edge-based degree, EBEW, and NuFBL, respectively. The change of modularity was significantly positively correlated with all structural properties (Correlation coefficients were 0.11393, 0.04112, and 0.04064, respectively, with all P-values<0.0001). **(d)-(f)** Relations of the change of robustness with edge-based degree, EBEW, and NuFBL, respectively. The change of robustness was significantly negatively correlated with all structural properties (Correlation coefficients were -0.08353, -0.08649, and -0.09906, respectively, with all P-values<0.0001).

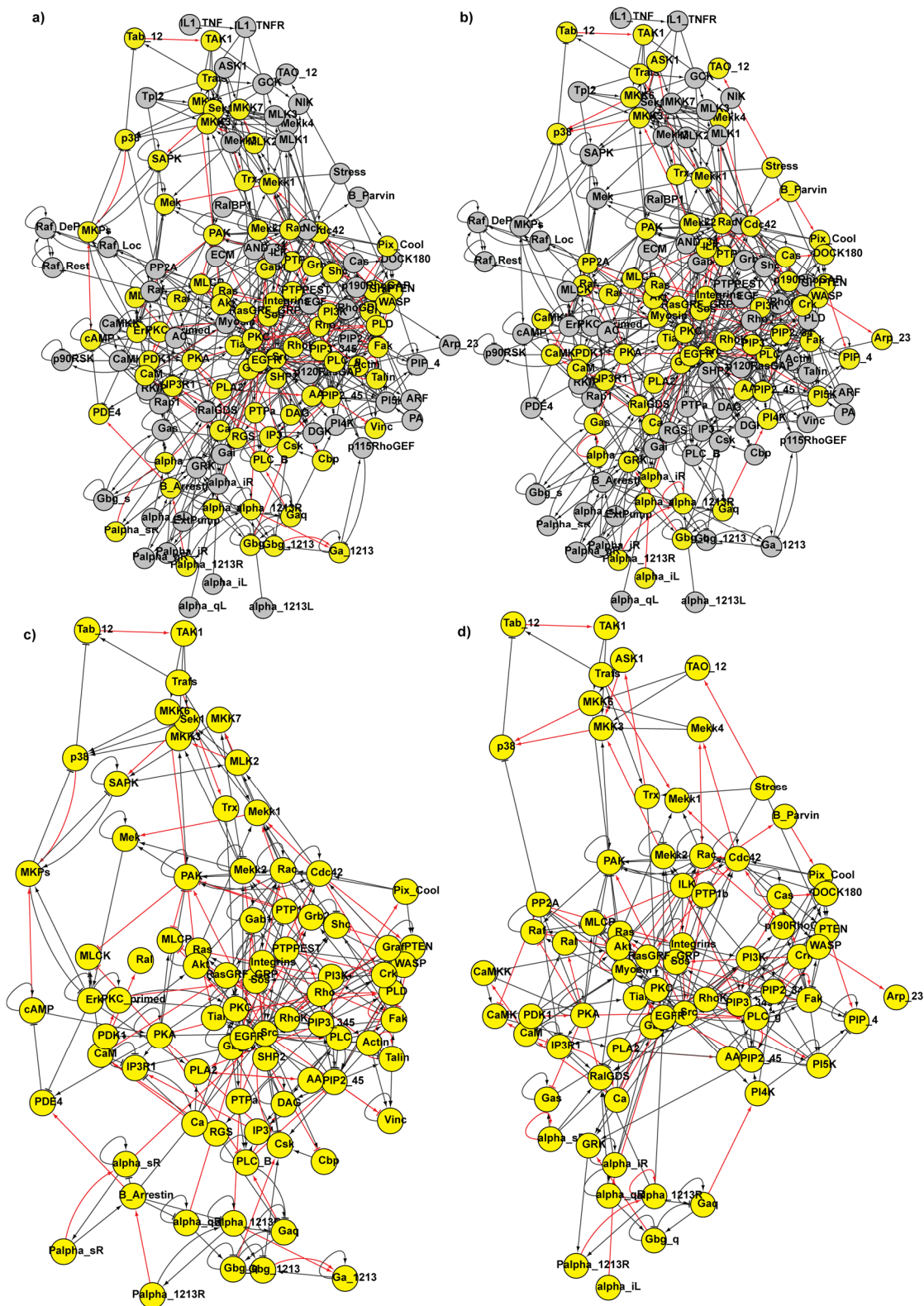


Figure S4.16. Topological distributions of High-MI/High-RD edges and their incident nodes in STF signaling network.

(a)-(b) Distributions of High-MI and High-RD edges, respectively, and their incident nodes. (c)-(d) Subgraphs with respect to High-MI-incident and High-RD-incident nodes, respectively. Red link and yellow node represent High-MI edge and High-MI-incident node, respectively, in both (a) and (c), whereas they represent High-RD edge and High-RD-incident node, respectively, in both (b) and (d).

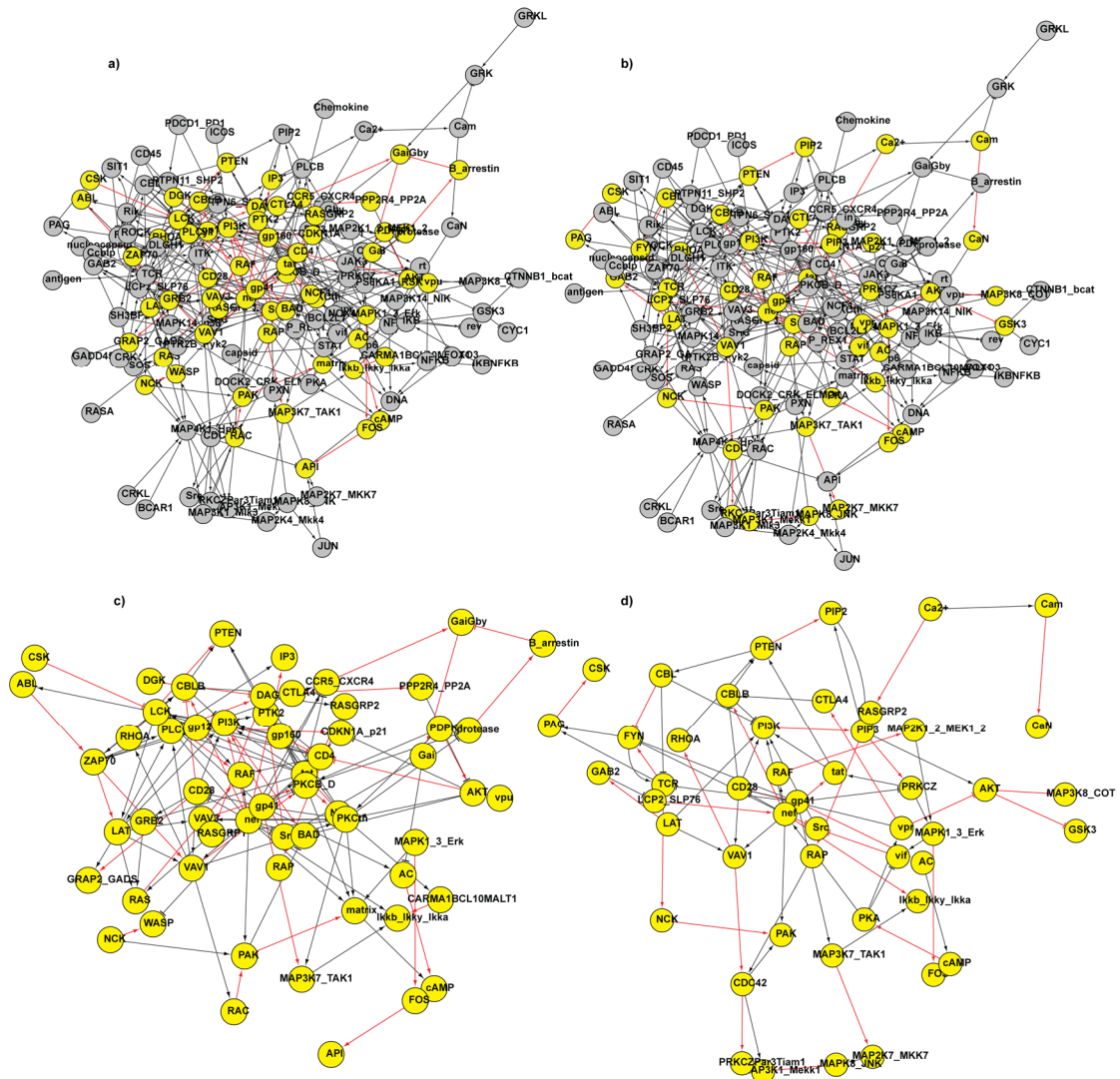


Figure S4.17. Topological distributions of High-MI/High-RD edges and their incident nodes in HIV-1 signaling network.

(a)-(b) Distributions of High-MI and High-RD edges, respectively, and their incident nodes. (c)-(d) Subgraphs with respect to High-MI-incident and High-RD-incident nodes, respectively. Red link and yellow node represent High-MI edge and High-MI-incident node, respectively, in both (a) and (c), whereas they represent High-RD edge and High-RD-incident node, respectively, in both (b) and (d).

APPENDIX D

Type of GO analysis	GO term	High-MI-incident (%)	The rest of genes (%)	P-value
Modularity change	Protein kinase activator activity	75.00	25.00	360.0E-6
	Negative regulation of cysteine-type endopeptidase activity	66.84	40.11	1.1E-6
	Response to mechanical stimulus	74.13	37.07	93.0E-9
	Activation of immune response	75.09	27.66	120.0E-18
	Protein phosphatase binding	84.82	21.21	7.1E-9
	Mitogen-activated protein kinase kinase binding	75.00	25.00	1.9E-6
	Growth factor receptor binding	76.73	30.69	50.0E-6
	Insulin receptor binding	89.82	17.96	450.0E-9
		High-RD-incident (%)	The rest of genes (%)	P-value
Robustness change	Rho guanyl-nucleotide exchange factor activity	75.00	25.00	760.0E-6
	Protein serine/threonine/tyrosine kinase activity	62.50	37.50	44.0E-12
	Cellular response to carbohydrate stimulus	66.67	33.33	47.0E-6
	Response to epidermal growth factor	60.00	40.00	1.4E-6
	Antigen receptor-mediated signaling pathway	72.73	27.27	42.0E-9
	Lipopolysaccharide-mediated signaling pathway	75.00	25.00	190.0E-6
	Immune response-activating cell surface receptor signaling pathway	71.70	31.37	55.0E-18
	Neurotrophin TRK receptor signaling pathway	64.60	43.07	8.5E-6

Table S4.1. GO analysis results between High-MI-incident/High-RD-incident group and the rest of genes in STF network.

All P-values were calculated by using Bonferroni test.

Type of GO analysis	GO term	High-MI-incident (%)	The rest of genes (%)	P-value
Modularity change	Negative regulation of kinase activity	61.54	38.46	720.0E-12
	Virus receptor activity	75.00	25.00	470.0E-6
	Regulation of defense response to virus	60.00	40.00	100.0E-6
	Positive regulation of immune response	64.34	39.06	560.0E-36
	Response to growth hormone	66.67	33.33	45.0E-9
	Phosphatase binding	63.64	36.36	510.0E-12
	Phosphotyrosine binding	60.00	40.00	5.5E-9
	Protein phosphatase binding	66.67	33.33	7.9E-9
		High-RD-incident (%)	The rest of genes (%)	P-value
Robustness change	Receptor signaling protein serine/threonine kinase activity	72.73	27.27	2.5E-12
	MAP kinase kinase kinase activity	75.00	25.00	7.0E-6
	Response to axon injury	75.00	25.00	550.0E-6
	Mast cell activation involved in immune response	61.39	46.04	200.0E-9
	Stimulatory C-type lectin receptor signaling pathway	60.00	40.00	180.0E-12
	Lipopolysaccharide-mediated signaling pathway	60.00	40.00	11.0E-6
	Negative regulation of ERBB signaling pathway	75.00	25.00	91.0E-6
	Negative regulation of epidermal growth factor receptor signaling pathway	75.00	25.00	83.0E-6

Table S4.2. GO analysis results between High-MI-incident/High-RD-incident group and the rest of genes in HIV-1 network.

All P-values were calculated by using Bonferroni test.

REFERENCES

- Alves, R., Antunes, F. and Salvador, A. (2006) Tools for kinetic modeling of biochemical networks, *Nat Biotech*, **24**, 667-672.
- Ananthasubramaniam, B. and Herzel, H. (2014) Positive Feedback Promotes Oscillations in Negative Feedback Loops, *PLOS ONE*, **9**, e104761.
- Barabási, A.-L. and Albert, R. (1999) Emergence of Scaling in Random Networks, *Science*, **286**, 509-512.
- Bhalla, U.S., Ram, P.T. and Iyengar, R. (2002) MAP Kinase Phosphatase As a Locus of Flexibility in a Mitogen-Activated Protein Kinase Signaling Network, *Science*, **297**, 1018.
- Bindea, G., *et al.* (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, *Bioinformatics*, **25**, 1091-1093.
- Binns, D., *et al.* (2009) QuickGO: a web-based tool for Gene Ontology searching, *Bioinformatics*, **25**, 3045-3046.
- Bonacich, P. (1987) Power and Centrality: A Family of Measures, *American Journal of Sociology*, **92**, 1170-1182.
- Bournazou, E. and Bromberg, J. (2013) Targeting the tumor microenvironment: JAK-STAT3 signaling, *JAK-STAT*, **2**, e23828.
- Brazas, M.D., *et al.* (2011) The 2011 bioinformatics links directory update: more resources, tools and databases and features to empower the bioinformatics community, *Nucleic Acids Research*, **39**, W3-W7.
- Campbell, C. and Albert, R. (2014) Stabilization of perturbed Boolean network attractors through compensatory interactions, *BMC Systems Biology*, **8**, 1-16.
- Carninci, P., *et al.* (2005) The Transcriptional Landscape of the Mammalian Genome, *Science*, **309**, 1559.
- Ciliberti, S., Martin, O.C. and Wagner, A. (2007) Robustness Can Evolve Gradually in Complex Regulatory Gene Networks with Varying Topology, *PLoS Comput Biol*, **3**, e15.
- Cline, M.S., *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape, *Nat. Protocols*, **2**, 2366-2382.
- Consortium, T.G.O. (2008) The Gene Ontology project in 2008, *Nucleic acids research*, **36**, D440-D444.
- Consortium, T.U. (2015) UniProt: a hub for protein information, *Nucleic acids research*, **43**, D204-D212.
- Cumbo, F., *et al.* (2014) GIANT: A Cytoscape Plugin for Modular Networks, *PLoS ONE*, **9**, e105001.
- de Reus, M.A., *et al.* (2014) An edge-centric perspective on the human connectome: link communities in the brain, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **369**.
- Erdős, P. and Rényi, A. (1959) On random graphs, I, *Publicationes Mathematicae (Debrecen)*, **6**, 290-297.
- Estrada, E. and Rodríguez-Velázquez, J.A. (2005) Subgraph centrality in complex networks, *Physical Review E*, **71**, 056103.
- Freeman, L. (1977) A Set of Measures of Centrality Based on Betweenness, *Sociometry*, **40**, 35-41.
- Freeman, L.C. (1977) A Set of Measures of Centrality Based on Betweenness, *Sociometry*, **40**, 35-41.

- Girvan, M. and Newman, M.E.J. (2002) Community structure in social and biological networks, *Proceedings of the National Academy of Sciences*, **99**, 7821-7826.
- Graudenzi, A., *et al.* (2011) Robustness Analysis of a Boolean Model of Gene Regulatory Network with Memory, *J Comput Biol*, **18**, 559-577.
- Han, J.-D.J., *et al.* (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network, *Nature*, **430**, 88-93.
- Harris, S.E., *et al.* (2002) A model of transcriptional regulatory networks based on biases in the observed regulation rules, *Complexity*, **7**, 23-40.
- Helikar, T., *et al.* (2012) The Cell Collective: Toward an open and collaborative approach to systems biology, *BMC Systems Biology*, **6**, 96.
- Hintze, A. and Adami, C. (2008) Evolution of Complex Modular Biological Networks, *PLOS Computational Biology*, **4**, e23.
- Hirabayashi, T., Murayama, T. and Shimizu, T. (2004) Regulatory Mechanism and Physiological Role of Cytosolic Phospholipase A₂, *Biological and Pharmaceutical Bulletin*, **27**, 1168-1173.
- Holme, P. (2011) Metabolic Robustness and Network Modularity: A Model Study, *PLoS ONE*, **6**, e16605.
- Ingolia, N.T. (2004) Topology and Robustness in the Drosophila Segment Polarity Network, *PLOS Biology*, **2**, e123.
- Jeong, H., *et al.* (2001) Lethality and centrality in protein networks, *Nature*, **411**, 41-42.
- Jeong, H., *et al.* (2000) The large-scale organization of metabolic networks, *Nature*, **407**, 651-654.
- Jordan, J.D., Landau, E.M. and Iyengar, R. (2000) Signaling Networks: The Origins of Cellular Multitasking, *Cell*, **103**, 193-200.
- Kaneko, K. (2007) Evolution of Robustness to Noise and Mutation in Gene Expression Dynamics, *PLOS ONE*, **2**, e434.
- Kauffman, S. (2004) A proposal for using the ensemble approach to understand genetic regulatory networks, *Journal of Theoretical Biology*, **230**, 581-590.
- Kauffman, S., *et al.* (2003) Random Boolean network models and the yeast transcriptional network, *Proceedings of the National Academy of Sciences*, **100**, 14796-14799.
- Kauffman, S., *et al.* (2004) Genetic networks with canalizing Boolean rules are always stable, *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 17102-17107.
- Kim, H. and Anderson, R. (2012) Temporal node centrality in complex networks, *Physical Review E*, **85**, 026107.
- Kim, J.-R., Yoon, Y. and Cho, K.-H. (2008) Coupled Feedback Loops Form Dynamic Motifs of Cellular Networks, *Biophysical Journal*, **94**, 359-365.
- Kitano, H. (2004) Biological robustness, *Nat Rev Genet*, **5**, 826-837.
- Kreimer, A., *et al.* (2008) The evolution of modularity in bacterial metabolic networks, *Proceedings of the National Academy of Sciences*, **105**, 6976-6981.
- Kwon, Y.-K. and Cho, K.-H. (2008) Coherent coupling of feedback loops: a design principle of cell signaling networks, *Bioinformatics*, **24**, 1926-1932.
- Kwon, Y.-K. and Cho, K.-H. (2008) Quantitative analysis of robustness and fragility in biological networks based on feedback dynamics, *Bioinformatics*, **24**, 987-994.
- Kwon, Y.-K., Choi, S. and Cho, K.-H. (2007) Investigations into the relationship between feedback loops and functional importance of a signal transduction network based on Boolean network modeling, *BMC Bioinformatics*, **8**, 384.

- Le, D.-H. and Kwon, Y.-K. (2011) The effects of feedback loops on disease comorbidity in human signaling networks, *Bioinformatics*, **27**, 1113-1120.
- Le, D.-H. and Kwon, Y.-K. (2011) NetDS: A Cytoscape plugin to analyze the robustness of dynamics and feedforward/feedback loop structures of biological networks, *Bioinformatics*.
- Le, D.-H. and Kwon, Y.-K. (2011) NetDS: a Cytoscape plugin to analyze the robustness of dynamics and feedforward/feedback loop structures of biological networks, *Bioinformatics*, **27**, 2767-2768.
- Le, D.-H. and Kwon, Y.-K. (2013) A coherent feedforward loop design principle to sustain robustness of biological networks, *Bioinformatics*, **29**, 630-637.
- Leicht, E.A. and Newman, M.E.J. (2008) Community Structure in Directed Networks, *Physical Review Letters*, **100**, 118703.
- Li, M., *et al.* (2015) A Topology Potential-Based Method for Identifying Essential Proteins from PPI Networks, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **12**, 372-383.
- Lin, Y.-S., *et al.* (2007) Proportion of Solvent-Exposed Amino Acids in a Protein and Rate of Protein Evolution, *Molecular Biology and Evolution*, **24**, 1005-1011.
- Little, J.W., Shepley, D.P. and Wert, D.W. (1999) Robustness of a gene regulatory circuit, *The EMBO Journal*, **18**, 4299-4307.
- Lopes, C.T., *et al.* (2010) Cytoscape Web: an interactive web-based network browser, *Bioinformatics*, **26**, 2347-2348.
- Mabuchi, S., *et al.* (2015) The PI3K/AKT/mTOR pathway as a therapeutic target in ovarian cancer, *Gynecologic Oncology*, **137**, 173-179.
- Maslov, S. and Sneppen, K. (2002) Specificity and Stability in Topology of Protein Networks, *Science*, **296**, 910-913.
- Maslov, S., Sneppen, K. and Alon, U. (2002) Correlation profiles and motifs in complex networks. In, *Handbook of Graphs and Networks*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 168-198.
- Morris, J., *et al.* (2011) clusterMaker: a multi-algorithm clustering plugin for Cytoscape, *BMC Bioinformatics*, **12**, 436.
- Naldi, A., *et al.* (2010) Diversity and Plasticity of Th Cell Types Predicted from Regulatory Network Modelling, *PLoS Comput Biol*, **6**, e1000912.
- Ng, P.C. and Henikoff, S. (2003) SIFT: predicting amino acid changes that affect protein function, *Nucleic Acids Research*, **31**, 3812-3814.
- Nikolaou, K., Sarris, M. and Talianidis, I. (2013) Molecular Pathways: The Complex Roles of Inflammation Pathways in the Development and Treatment of Liver Cancer, *Clinical Cancer Research*, **19**, 2810.
- Noack, A. (2009) Modularity clustering is force-directed layout, *Physical Review E*, **79**, 026102.
- Oyeyemi, O.J., *et al.* (2015) A logical model of HIV-1 interactions with the T-cell activation signalling pathway, *Bioinformatics*, **31**, 1075-1083.
- Paroni, A., *et al.* (2016) CABERNET: a Cytoscape app for augmented Boolean models of gene regulatory NETWORKS, *BMC Bioinformatics*, **17**, 64.
- Pomerening, J.R., Sontag, E.D. and Ferrell, J.E. (2003) Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2, *Nat Cell Biol*, **5**, 346-351.
- Puniya, B.L., *et al.* (2016) Systems Perturbation Analysis of a Large-Scale Signal Transduction Model Reveals Potentially Influential Candidates for Cancer Therapeutics, *Frontiers in Bioengineering and Biotechnology*, **4**, 10.

- Rivera, C., Vakil, R. and Bader, J. (2010) NeMo: Network Module identification in Cytoscape, *BMC Bioinformatics*, **11**, S61.
- Rual, J.-F., *et al.* (2005) Towards a proteome-scale map of the human protein-protein interaction network, *Nature*, **437**, 1173-1178.
- Saadatpour, A., *et al.* (2011) Dynamical and Structural Analysis of a T Cell Survival Network Identifies Novel Candidate Therapeutic Targets for Large Granular Lymphocyte Leukemia, *PLOS Computational Biology*, **7**, e1002267.
- Shannon, P., *et al.* (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks, *Genome Research*, **13**, 2498-2504.
- Shen-Orr, S.S., *et al.* (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*, *Nat Genet*, **31**, 64-68.
- Shimbel, A. (1953) Structural parameters of communication networks, *Bulletin of Mathematical Biology*, **15**, 501-507.
- Shimbel, A. (1953) Structural parameters of communication networks, *Bulletin of Mathematical Biophysics*, **15**, 501-507.
- Slomovitz, B.M. and Coleman, R.L. (2012) The PI3K/AKT/mTOR Pathway as a Therapeutic Target in Endometrial Cancer, *Clinical Cancer Research*, **18**, 5856-5864.
- Smoot, M.E., *et al.* (2011) Cytoscape 2.8: new features for data integration and network visualization, *Bioinformatics*, **27**, 431-432.
- SPSS (2012) IBM SPSS Statistics, [www.ibm.com/products/spss-statistics].
- Steinway, S.N., *et al.* (2015) Inference of Network Dynamics and Metabolic Interactions in the Gut Microbiome, *PLoS Comput Biol*, **11**, e1004338.
- Stelzl, U., *et al.* (2005) A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome, *Cell*, **122**, 957-968.
- Suderman, M. and Hallett, M. (2007) Tools for visually exploring biological networks, *Bioinformatics*, **23**, 2651-2659.
- Szalay-Bekó, M., *et al.* (2012) ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality, *Bioinformatics*, **28**, 2202-2204.
- Tadaka, S. and Kinoshita, K. (2016) NCMine: Core-peripheral based functional module detection using near-clique mining, *Bioinformatics*.
- Takemoto, K. and Kihara, K. (2013) Modular organization of cancer signaling networks is associated with patient survivability, *Biosystems*, **113**, 149-154.
- Tran, T.-D. and Kwon, Y.-K. (2013) The relationship between modularity and robustness in signalling networks, *Journal of The Royal Society Interface*, **10**.
- Tran, T.-D. and Kwon, Y.-K. (2013) The relationship between modularity and robustness in signalling networks, *Journal of the Royal Society Interface*, **10**, 20130771.
- Trinh, H.-C. and Kwon, Y.-K. (2015) Effective Boolean dynamics analysis to identify functionally important genes in large-scale signaling networks, *Biosystems*, **137**, 64-72.
- Trinh, H.-C. and Kwon, Y.-K. (2016) Edge-based sensitivity analysis of signaling networks by using Boolean dynamics, *Bioinformatics*, **32**, i763-i771.
- Trinh, H.-C., Le, D.-H. and Kwon, Y.-K. (2014) PANET: A GPU-Based Tool for Fast Parallel Analysis of Robustness Dynamics and Feed-Forward/Feedback Loop Structures in Large-Scale Biological Networks, *PLoS ONE*, **9**, e103010.
- Truong, C.-D., Tran, T.-D. and Kwon, Y.-K. (2016) MORO: a Cytoscape app for relationship analysis between modularity and robustness in large-scale biological networks, *BMC Systems Biology*, **10**, 122.

- Variano, E.A., McCoy, J.H. and Lipson, H. (2004) Networks, Dynamics, and Modularity, *Physical Review Letters*, **92**, 188701.
- Viana, M.P., *et al.* (2009) Modularity and robustness of bone networks, *Molecular BioSystems*, **5**, 255-261.
- von Dassow, G. and Munro, E. (1999) Modularity in animal development and evolution: Elements of a conceptual framework for EvoDevo, *Journal of Experimental Zoology*, **285**, 307-325.
- Wang, S.W. and Sun, Y.M. (2014) The IL-6/JAK/STAT3 pathway: Potential therapeutic strategies in treating colorectal cancer (Review), *Int J Oncol*, **44**, 1032-1040.
- Winterhalter, C., *et al.* (2014) Pepper: cytoscape app for protein complex expansion using protein-protein interaction networks, *Bioinformatics*.
- Wuchty, S. and Stadler, P.F. (2003) Centers of complex networks, *Journal of Theoretical Biology*, **223**, 45-53.
- Yi, T.-M., *et al.* (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control, *Proceedings of the National Academy of Sciences*, **97**, 4649-4653.
- Yoon, J., Blumer, A. and Lee, K. (2006) An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality, *Bioinformatics*, **22**, 3106-3108.
- Zhang, Y., *et al.* (2014) Comprehensive analysis of microRNA-regulated protein interaction network reveals the tumor suppressive role of microRNA-149 in human hepatocellular carcinoma via targeting AKT-mTOR pathway, *Molecular Cancer*, **13**, 253.