PROTEIN DOMAIN NETWORKS: ANALYSIS OF ATTACK TOLERANCE UNDER VARIED CIRCUMSTANCES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ŞAZİYE DENİZ OĞUZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
SCIENTIFIC COMPUTING

SEPTEMBER 2010

Approval of the thesis:

PROTEIN DOMAIN NETWORKS: ANALYSIS OF ATTACK TOLERANCE UNDER VARIED CIRCUMSTANCES

submitted by **ŞAZİYE DENİZ OĞUZ** in partial fulfillment of the requirements for the degree of **Master of Science in Department of Scientific Computing, Middle East Technical University** by,

Prof. Dr. Ersan Akyıldız ________________
Director, Graduate School of **Applied Mathematics**

Prof. Dr. Bülent Karasözen ________________
Head of Department, **Scientific Computing**

Assist. Prof. Dr. Hakan Öktem ________________
Supervisor, **Department of Scientific Computing**

Examining Committee Members:

Prof. Dr. Gerhard Wilhelm Weber ________________
Institute of Applied Mathematics, METU

Assist. Prof. Dr. Hakan Öktem ________________
Institute of Applied Mathematics, METU

Prof. Dr. Marat Ubaydulla Akhmet ________________
Department of Mathematics, METU

Assoc. Prof. Dr. Azize Hayfavi ________________
Institute of Applied Mathematics, METU

Assist. Prof. Dr. Kasırga Yıldırak ________________
Department of Economics, University of Trakya

Date: ________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ŞAZİYE DENİZ OĞUZ

Signature :

ABSTRACT

PROTEIN DOMAIN NETWORKS: ANALYSIS OF ATTACK TOLERANCE UNDER VARIED CIRCUMSTANCES

Oğuz, Şaziye Deniz
M.S., Department of Scientific Computing
Supervisor : Assist. Prof. Dr. Hakan Öktem

September 2010, 93 pages

Recently, there has been much interest in the resilience of complex networks to random failures and intentional attacks. The study of the network robustness is particularly important by several occasions. In one hand a higher degree of robustness to errors and attacks may be desired for maintaining the information flow in communication networks under attacks. On the other hand planning a very limited attack aimed at fragmenting a network by removal of minimum number of the most important nodes might have significant usage in drug design.

Many real world networks were found to display scale free topology including WWW, the internet, social networks or regulatory gene and protein networks. In the recent studies it was shown that while these networks have a surprising error tolerance, their scale-free topology makes them fragile under intentional attack, leaving the scientists a challenge on how to improve the networks robustness against attacks.

In this thesis, we studied the protein domain co-occurrence network of yeast which displays scale free topology generated with data from Biomart which links to Pfam database. Several networks obtained from protein domain co-occurrence network having exactly the same connectivity distribution were compared under attacks to investigate the assumption that the different networks with the same connectivity distribution do not need to have the same attack tolerances. In addition to this, we considered that the networks with the same connectivity distribution have higher attack tolerance as we organize the same resources in a better way. Then, we checked for the variations of attack tolerance of the networks with the same connectiviy distributions. Furthermore, we investigated whether there is an evolutionary mechanism for having networks with higher or lower attack tolerances for the same connectivity distribution. As a result of these investigations, the different networks with the same connectivity distribution do not have the same attack tolerances under attack. In addition to this, it was observed that the networks with the same connectivity distribution have higher attack tolerances as organizing the same resources in a better way which implies that there is an evolutionary mechanism for having networks with higher attack tolerance for the same connectivity distribution.

Keywords: Scale-free Networks, Protein Domain Networks, Connectivity Distribution of Networks, Attack Tolerance, Network Resilience

ÖZ

PROTEİN DOMAİN AĞLARI: FARKLI KOŞULLAR ALTINDA SALDIRI TOLERANSININ ANALİZİ

Oğuz, Şaziye Deniz
Yüksek Lisans, Bilimsel Hesaplama Bölümü
Tez Yöneticisi : Yrd. Doç. Dr. Hakan Öktem

Eylül 2010, 93 sayfa

Karmaşık ağların hatalar ve saldırılar karşısında nasıl davrandığı birçok bilim adamının ilgisini çekmektedir. Ağların dayanıklılığı üzerine çalışmalar birçok açıdan önemlidir. Örneğin, iletişim ağlarında bilgi akışının düzgün sağlanabilmesi için hatalara ve saldırılara karşı dayanıklılığının yüksek olması istenmektedir. Diğer yandan, ağları parçalamak için en az sayıda en onemli yapı taşlarını çıkartarak kısıtlı sayıda saldırı planlamak ilaç tasarımında onemli kullanım alanı bulmaktadır.

WWW, internet, sosyal ağlar ve protein ağları gibi birçok gerçek ağların scale free topolojik yapısına sahip olduğu bulunmuştur. Son zamanlarda yapılan çalışmalarda bu ağların şaşırtıcı derecede hatalara karşı dayanıklı oldukları bulunurken diğer yandan scale free topolojik yapısına sahip olmalarından ötürü saldırılara karşı çok hassas oldukları gözlenmiştir. Bu durum bilimadanlarını bu tür özellik gösteren ağların saldırılara karşı dayanıklılığının nasıl arttırılabileceği sorusu ile karşı karşıya bırakmıştır.

Bu tezde mayanın (S. cerevisiae) scale free topolojik yapıya sahip protein domain ağları üzerine calıştık. Çalışmalarımız sırasında Pfam veri bankasına ulaşmamazı sağlayan Biomart' taki verileri kullandık. Bu ağdan aynı bağlantı dağılımına sahip birçok ağ elde ettik. Aynı dağılıma sahip farklı ağların saldırılar altında aynı saldırı toleransını göstermemesi gerektiği varsayımı incelemek için bu birçok ağın saldıralar altında saldırı dayanaklılığını karşılaştırdık. Buna ek olarak, kaynakları daha iyi organize etttikçe aynı bağlantı dağılımına sahip ağların daha yüksek saldırı toleransına sahip olabileceği üzerinde durduk. Sonra, aynı bağlantı dağılımına sahip ağların saldırı toleransı değişimini kontrol ettik. Son olarak, yüksek ya da düşük saldırı toleransına sahip ağların varolması için evrimsel (gelişiminden kaynaklanan) bir mekanizma olup olmadığını baktık. Bütün bu araştırmaların sonucunda, aynı bağlantı dağılımına sahip ağların aynı saldırı toleransını göstermediği belirlendi. Ek olarak, kaynakları daha iyi organize etttikçe aynı bağlantı dağılımına sahip ağların daha yüksek saldırı toleransına sahip olduğu ve bununda yüksek saldırı toleransına sahip ağların varolması için evrimsel bir mekanizma olduğuna işaret ettiği sonucuna varıldı.

Anahtar Kelimeler: Scale Free Ağlar, Protein Domain Ağlar, Ağların Bağlanırlık Dağılımı, Saldırı Toleransı, Ağ Dayananıklılığı

to life

ACKNOWLEDGMENTS

Firstly, I would like to express my deepest gratitude to my supervisor Assist. Prof. Dr. Hakan Öktem for patiently guiding, supporting me and for his suggestions throughout this study.

I also would like to express my special thanks to Stefan Wuchty (Department of Physics, University of Notre Dame, USA) providing me data of Protein Domain Co-occurrence Network of Yeast.

Besides, I would like to thank to my family for all their love, support and encouragements of me in all of my endeavors. I espiacially acknowledge my debt and express my heartfelt thanks to my sister Pınar Oğuz for her encouragement, understanding and guidance throughout this study.

I also wish to thank to the member of METU Mountaineering and Winter Sports Club espiaclly my dear friends Barış, Gülşah, Eylem and Ceren for making the life more enjoyble and for their supports.

I am grateful to the members of METU Computer Center especially Özden, Uğur to offer me so generously a perfect working environment, support, and help whenever I needed it.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Many real world systems can be represented by networks. A network is a set of items which is called vertices or nodes with the connections either directly or indirectly between them, called edges. The existing emprical and theoretical results indicate that complex networks can be divided into two main classes as random (exponential) networks and scale free networks according to the their degree (connectivty) distribution $P(k)$ (the probability that a node in the network is connected to k other nodes). In the first class of networks, $p(k)$ peaks at an average $< k >$ and decays exponentially for large k. This class of networks was first studied by Paul Erdös and Alfred Renyi. Erdös-Renyi (ER) random network is a classical representation of exponential networks [17]. Exponential networks are homogeneous in connectivity which means that most of the nodes in the network have approximately the same number of connections around $< k >$, $k \approx< k >$. But, recent studies show that most real-world systems exhibit scale free structure. These real-world networks include the Worl Wide Web [5], the internet [13], biological networks such as metobolic networks [22], protein domain networks [20], [23] and the author collaboration networks [24]. It was found that structure of these networks can not be described by ER model, so Barabasi et al. introduce a model called BA model which explains the emergence of scale free structure in these networks [15].

Scale free network exhibits a power law degree distribution in the form $P(k) \sim k^{-\gamma}$ where gamma "γ" is free of characteristic scale and its value typically in the range $2 \leq \gamma \leq 3$. As long as gamma "γ" greater or equal to 1, its value may lie outside these bounds [25]. In contrast to the exponential network, scale free networks are heterogenous which means that most node has a few edges while few nodes in the networks has a huge number of edges. This feature in scale free network is deserved to be paid attention which implies that network's property is determined by the most highly connected nodes.

The study of the network robustness receives a growing interest among scientist and is particularly important by several occasions. In one hand a higher degree of robustness to errors and attacks may be desired for maintaining the information flow in communication networks under attacks. On the other hand planning a very limited attack aimed at fragmenting a network by removal of minimum number of the most important nodes might have significant usage in drug design. The robustness of a network can be defined by its behavior under perturbations. There are two general categories of such perturbations; errors or failures: random removal of nodes and attacks: the targeted removal of chosen nodes. The way the nodes are chosen during an attack is called an *attack strategy*. Some attack strategies are introduced in [1], [2]. There are also several methods to measure the robustness of a network under failures and attacks [1], [3], [5], [19]. Using these methods, authors compared two network models- scale free and random networks under failures and attacks. They found that scale-free networks display an unexpected degree of robustness, i.e., the ability of their nodes to communicate being unaffected by even high failure rates. However, these networks are extremely vulnerable to intentional attacks, i.e., to the removal of a few number of highly connected nodes. On the other hand, evolving networks with exponential connectivity distribution are not as robust to random failures, but more resilient to intentional attacks [1], [2], [4], [7], [10].

In this work, we use protein domain co-occurence network of yeast (S. cerevesiae) generated with data from Biomart (Pfam database) data management system. Domains are basic evolutionary units of proteins which are well-defined regions within a protein that can evolve, function and fold independently from the rest of the protein and have their own function. These domains and nature of their interactions determine the function of the protein. In protein domain co-occurence network of yeast (S. cerevesiae), domains are represented by nodes and two domains are considered as connected if they occur together in one protein at least once. Most biological networks including protein domain co-occurrence network were found to exhibit scale free topology [22], [20]. This scale free structure support the expectation from biological network under failures such that as biological networks are required to function in various conditions, it is expected that they should have evolved to own robust structures against structural and enviromental perturbations [26]. In this study we will not compare scale free networks and random networks under attacks which has been studied widely in literature [1], [2], [4]. Instead, we will compare several networks exhibiting scale free structure, which have exactly the same connectivity, under attacks. We analyze the robustness of the network under attacks by studying how the size of the largest connected component varies as a function of the number of removed nodes.

Even the connectivity distribution is an important indicator of a network's qualitative features, different networks with the same connectivity distribution do not need to have the same attack tolerances. Additionally, it can be considered that the networks with same connectivity distribution have higher attack tolerance as we organize the same resources in a better way. Then, it can be checked for the variations of attack tolerance of the networks with the same connectivity distributions. Furthermore, we investigate whether there is an evolutionary mechanism for having networks with higher or lower attack tolerances for the same connectivity distribution. For these purposes, we wrote an algorithm such that we randomly change the links of nodes in protein domain co-occurence network of yeast conserving the connectivity of the network. Later, we analyze the attack tolerance of the randomly modified networks and from these networks we extract the least and the most vulnerable networks under attacks. We continue collecting the least vulnerable network from the networks which are also obtained from the least vulnerable network and the most vulnerable network from the networks which are also obtained from the most vulnerable one.

In this thesis, basic definitions related to graph theory and network's properties and models are given briefly in Chapter 2. In Chapter 3, we introduce the methods to measure attack tolerance of networks which will be used in this work. This chapter also includes comparison of network models under attacks and failures. Chapter 4 contains the application part. Several networks obtained by modifying the connections of nodes in the network are compared under attack which is done by using one of the methods and strategies introduced in Chapter 3. Chapter 5 concludes the thesis and further possible studies to extend the thesis are discussed.

CHAPTER 2

BACKGROUND

2.1 Mathematical Background

Definitions and theorems in this part are mainly taken from [27], [28].

2.1.1 Basic Definitions

Any system of interconnected elements can conveniently be described by means of a diagram consisting of a set of points together with lines joining certain pairs of these points. For example, the points could represent people, with lines joining pairs of friends, or the points might be communication centers, with lines representing communication links. Notice that in such diagram one is mainly interesting in whether or not two given points are joined by line, the manner in which they are joined is immaterial. A mathematical abstraction of situations of this type gives rise to the concept of a graph.

Definition 2.1.1 *A* ***graph*** *or a* ***general graph*** *is an ordered triple $G = (V, E, \phi)$ where*

1. $V \neq \emptyset$,
2. $V \cap E = \emptyset$,
3. $\phi : E \to \boldsymbol{P}^1(V)$ *is a map such that* $\mid \phi(e) \mid \in \{1, 2\}$ *for each* $e \in E$.

The elements of V are the vertices *of G the elements of E are the* edges *of G. The map ϕ is called an* edgemap *and the vertices in $\phi(e)$ are called the* endvertices *of the edge e.*

[1] If S is set, then the set of all subsets of S, denoted $\mathbf{P}(S)$, is called the *power set* of S. For example, If $S = \{1, 2, 3\}$, then $\mathbf{P}(S) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.

Terminology: The total number of vertices in the graph (the cardinality of the set V denoted $\mid V \mid$) is denoted as N and defines the order of the graph. We will refer to N as the size of the network.

Example 2.1.2 *Graphs are in most cases represented by diagrams consisting of dots, which represent the vertices, and curves drawn between the dots representing the endvertices, which represent the edge between vertices.*

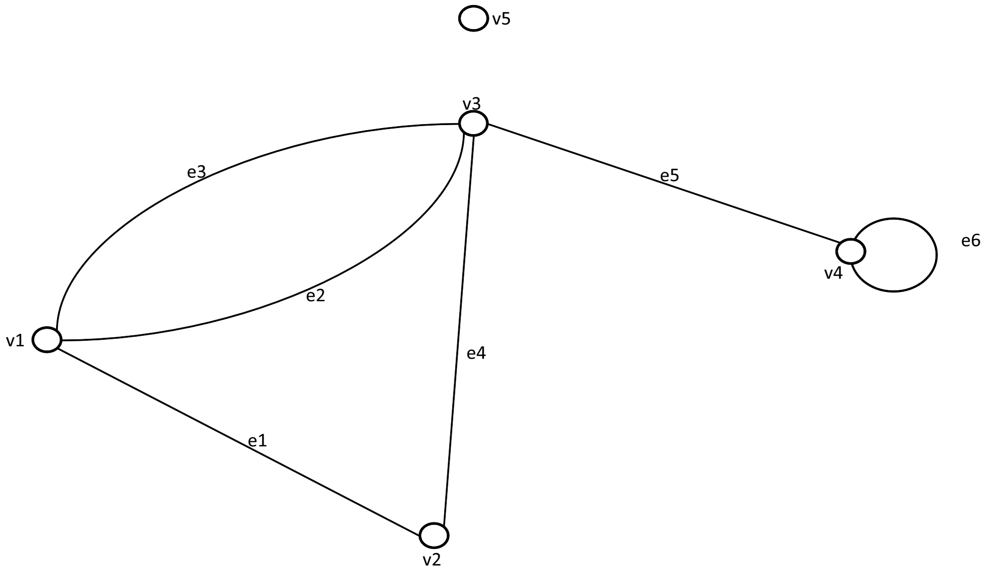


Figure 2.1: A graph with five vertices and six edges.

Consider the diagram shown in Figure 2.1. Here we see that $G = (V, E, \phi)$ *where* $V = \{v_1, v_2, ..., v_5\}$, $E = \{e_1, e_2, ..., e_6\}$ *and the edgemap* ϕ *is given by*

$$\phi(e_1) = \{v_1, v_2\},$$
$$\phi(e_2) = \phi(e_3) = \{v_1, v_3\},$$
$$\phi(e_4) = \{v_2, v_3\},$$
$$\phi(e_5) = \{v_3, v_4\},$$
$$\phi(e_6) = \{v_4\}.$$

The next definition presents the most basic terminology on graphs.

Definition 2.1.3 *Let* $G = (V, E, \phi)$ *be a graph*

1. *Vertices u and v in V are* ***adjacent*** *or* ***neighbors****, if they are the endvertices of some edge* $e \in E$*. That is they are adjacent, if there is an* $e \in E$ *such that* $\phi(e) = \{u, v\}$.

2. *Two distinct edges e and f are* ***adjacent*** *to each other if they have common endvertex. That is* $\phi(e) \cap \phi(f) \neq \emptyset$*.*

3. *A vertex u and an edge are* ***incident*** *if* $u \in \phi(e)$*, that is if u is an endvertex of e.*

4. *A* ***loop*** *is an edge whose endvertices are equal, that is* $\mid \phi(e) \mid= 1$.

5. *We say that* $E^{'} \subseteq E$ *is a set of* ***multiple edges*** *or* ***parallel edges****, if* $\mid E^{'} \mid\geq 2$ *and all* $e^{'} \in E^{'}$ *have the same set of endvertices. That is* $\phi(e^{'}) = \phi(f^{'})\ \forall e^{'}, f^{'} \in E^{'}$.

6. *A vertex u is called* ***isolated****, if it is not an endvertex of any edge. That is* $u \notin \phi(e)$ $\forall e \in E$.

When tackling a problem that can be phrased in graph-theoretic terms, often it can be reduced to a problem involving a graph having no multiple edges or any loop. Such graphs constitute an important class called *simple graph*. Since there is at most one edge between a pair of vertices in a simple graph, the edges are in one-to-one correspondence with their distinc end vertices. Therefore, a simple graph can be defined without the edgemap ϕ from Definition 2.1.1. Because of the importance of simple graphs, it is convenient to state their formal definition separately.

Definition 2.1.4 *A* ***simple graph*** *is an ordered pair G = (V, E), where V is a nonempty set of vertices and E is a set of 2-element subset of V, that is a simple graph is a graph that has no self-loops or multi-edges*

$$E \subseteq \{X : X \subseteq V, \mid X \mid= 2\} = \{\{u, v\} : u, v \in V, u \neq v\}. \tag{2.1}$$

Definition 2.1.5 *The* ***complete graph*** *on n vertices is the simple graph* $G = K_n$*, where*

$$V(K_n) = \{v_1, v_2, ..., v_n\},$$
$$E(K_n) = \{\{v_i, v_j\} : 1 \leq i < j \leq n\}.$$

That is, every pair of distinct vertices is connected by an edge (Figure 2.2).

Note: The complete graph K_n has $\frac{n(n-1)}{2}$ edges.

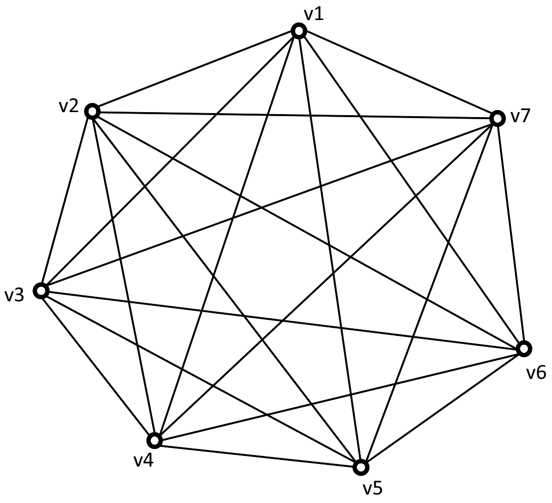


Figure 2.2: Complete graph K_7.

Many properties of graphs must be stated in terms of numerical values associates with the graph or some of its components. The first such attribute, it is defined the degree of a vertex.

Definition 2.1.6 *Let $G = (V, E, \phi)$ be a graph and $v \in V$ a vertex of G. The **degree** of v, denoted d(v), is defined by*

$$d(v) =| \{e \in E : v \in \phi(e), | \phi(e) |= 2\} | +2 | \{e \in E : v \in \phi(e), | \phi(e) |= 1\} | . \qquad (2.2)$$

Remark: For a graph G and a vertex $v \in V(G)$ we have:

- If G is a simple graph, then $d(v)$ is the number of neighbors of v in G.
- If G is a general graph, then $d(v)$ is the number of edges having v as an endvertex, where the loops are counted twice.

Theorem 2.1.7 *For a graph G we have*

$$\sum_{u \in V} d(u) = 2 \mid E \mid . \tag{2.3}$$

Definition 2.1.8 *Let $G = (V, E)$ be a graph and k be the degree of a vertex. To calculate the **average degree**, all degrees are summed and divided by the total number of vertices in the network:*

$$< k >= \frac{\sum_{i=}^{N} k(v_i)}{N}, \tag{2.4}$$

where N is the total number of nodes in the network.

When studying a specific graph, many times our attention is focused solely on a special part of the graph, perhaps on a smaller graph lying inside the larger graph . This motivates the following definition of a subgraph.

Definition 2.1.9 *For graphs $G^{'} = (V^{'}, E^{'}, \phi^{'})$ and $G = (V, E, \phi)$, we say that $G^{'}$ is a **subgraph** of G, if*

1. *$V^{'} \subseteq V$,*

2. *$E^{'} \subseteq E$,*

3. *$\phi^{'}(e) = \phi(e)\ \forall e \in E^{'}$.*

It is denoted the subgraph relation on graphs using the standard subset notation $\subseteq$. That is, if $G^{'}$ is a subgraph of G, then it is written $G^{'} \subseteq G$. In particular, for simple graphs $G^{'} = (V^{'}, E^{'})$ and $G = (V, E)$, $G^{'}$ is a subgraph of G if $V^{'} \subseteq V$ and $E^{'} \subseteq E$.

Example 2.1.10 *Consider the graph G given in Figure 2.3 (a). Figure 2.1.1 (b) shows a subgraph $G^{'} o f G$.*

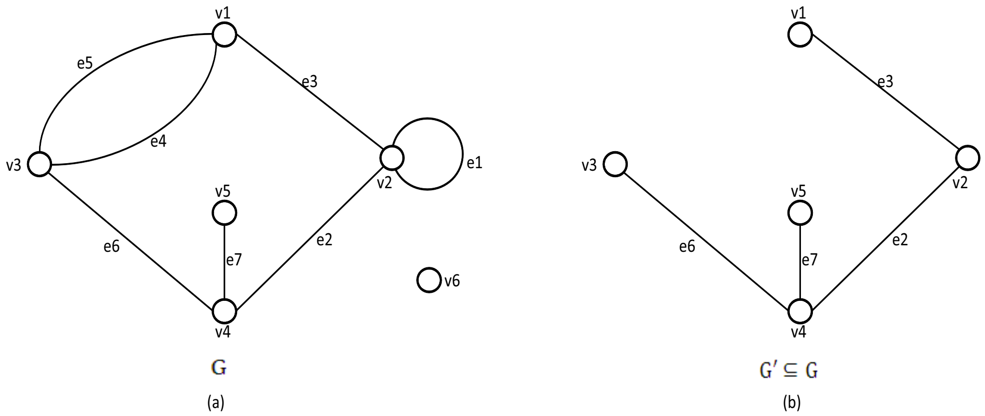


Figure 2.3: Subgraph.

Definition 2.1.11 *A **directed graph** or **digraph** is an ordered triple $\vec{G} = (V, E, \eta)$ where*

1. *$V \neq \emptyset$,*

2. *$V \cap E = \emptyset$,*

3. *$\eta : E \mapsto V \times V$ is a map.*

The set V is the set of vertices, and the set E the directed edges, or arcs. If $\eta(e) = (u, v)$, then u is called the tail of e and v the head of e.

*If $\eta(e) = (v, v)$, then e is called a **directed loop**.*

Two directed edges e and $e^{'}$ are said to be parallel edges if $\eta(e) = \eta(e^{'})$. That is, the edges are mapped onto the same ordered pair of vertices

Example 2.1.12 *Consider the digraph $\vec{G} = (V, E, \eta)$ with five vertices and six directed edgesas shown in Figure 2.4. Here, $\vec{G} = (V, E, \eta)$, where $V = \{v_1, v_2, v_3, v_4, v_5\}$, $E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$, and the edgemap η is given by*

$$\eta(e_1) = (v_1, v_2),$$
$$\eta(e_2) = (v_3, v_1),$$
$$\eta(e_3) = (v_1, v_3),$$
$$\eta(e_4) = (v_2, v_3),$$
$$\eta(e_5) = (v_3, v_4),$$
$$\eta(e_6) = (v_4, v_4).$$

In particular, the directed edge e_6 is the directed loop.

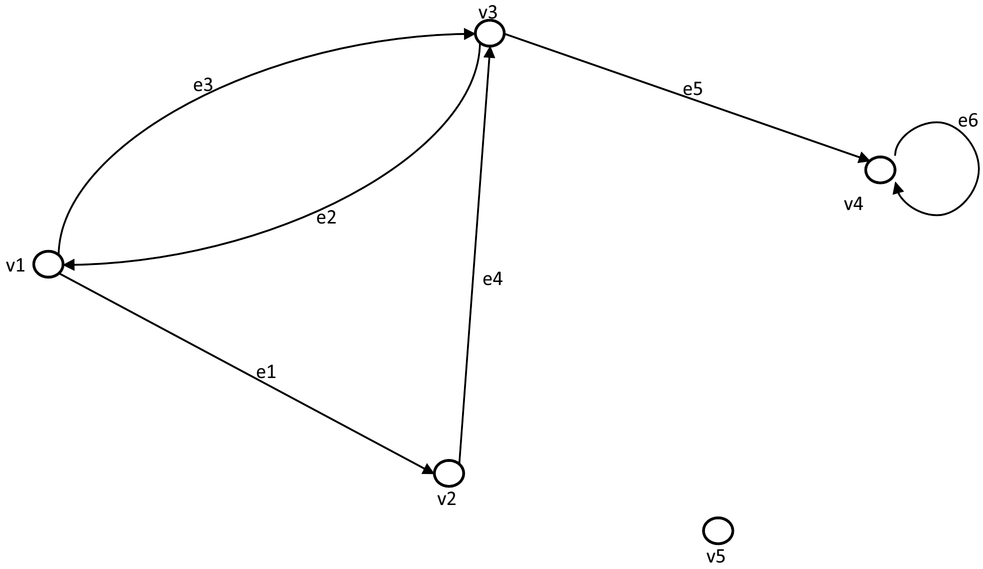


Figure 2.4: A digraph with five vertices and six directed edges.

Definition 2.1.13 *A digraph having no directed loops and no parallel directed edges is called a **simple digraph**.*

Any simple digraph $\vec{G}$ can be presented as an ordered pair $\vec{G} = (V, E)$, where V is a set of vertices and $E \subseteq V \times V$. Note that there is a slight difference between a simple digraph and having the edges represented by a subset of $V \times V$:

- A digraph $\vec{G}$ has representation $\vec{G} = (V, E)$, where $E \subseteq V \times V$ if, only if, $\vec{G}$ has no parallel directed edges.
- A digraph is simple if, only if, $\vec{G}$ has a representation $\vec{G} = (V, E)$, where $E \subseteq V \times V$ and $\vec{G}$ has no directed loops.

2.1.2 The Basic Representations for Graphs

From a mathematical point of view, it is convenient to define a graph by means of the adjacency matrix $\mathbf{A} = (a_{ij})_{i,j=1,2,...,N}$. This is a $N \times N$ matrix defined in the following definition:

Definition 2.1.14 *We define **the adjancency matrix*** $\mathbf{A}$ *of a graph G(V, E) as the* $\mid V \mid \times \mid V \mid$*:*

$$a_{ij} = \begin{cases} 1, & \text{if } (i, j) \in E \\ 0, & \text{if } (i, j) \notin E. \end{cases} \tag{2.5}$$

We define the adjancency matrix representation of a digraph G in the same way as for an undirected graph. The adjacency matrix for a graph is symmetric, while the adjanceny matrix for a digraph is asymmetric.

Definition 2.1.15 *A **weighted graph** is a graph that has a numeric label w(e) associated with each edge e, called the weight of edge e. Edges weights can be integers, rational numbers, or real numbers, which represent a concept such as distance, connection costs, or affinity.*

2.1.2.1 Paths and Cyles

Definition 2.1.16 1. *A **walk** in a graph $G = (V, E, \phi)$ is an alternating sequence*

$$(v_0, e_1, v_1, e_2, ..., e_k, v_k)$$

of vertices and edges that begins and ends with a vertex. For each $i \in \{1, ..., k\}$ the endvertices of e_i are v_{i-1} and v_i. That is

$$\phi(e_i) = \{v_{i-1}, v_i\},$$

The vertex v_o is the initial vertex of the walk. The vertex v_k is the final vertex of the walk. The initial or final vertex of a walk is also called an endvertex of the walk. The natural number k is the lenght *of the walk.*

2. *A **trail** in G is a walk with all of its edges $e_1, e_2, ..., e_k$ distinct*

3. *A **path** in G is a walk with all of its vertices $v_0, v_1, ..., v_k$ distinct.*

4. *For vertices u and v in G, a u, v-walk (u, v-trail or u, v-path) is walk (respectively trail or path) with initial vertex u and final vertex v.*

5. *A walk or trail of length at least one is **closed** if the initial vertex and the final vertex are the same. A closed trail is also called a **circuit**.*

6. *A **cycle** is a closed walk with distinc vertices except for the initial and final vertex, which are the same.*

Assuming G is a simple graph, then a walk, trail, path, or cycle can be specified by a sequence of vertices $v_0, v_1, ..., v_k$ instead of $(v_0, e_1, v_1, e_2, ..., e_k, v_k)$. This is because a pair of adjacent vertices v_{i-1} and v_i completely determine the only edge $e_i = \{v_{i-1}, v_i\}$ between them.

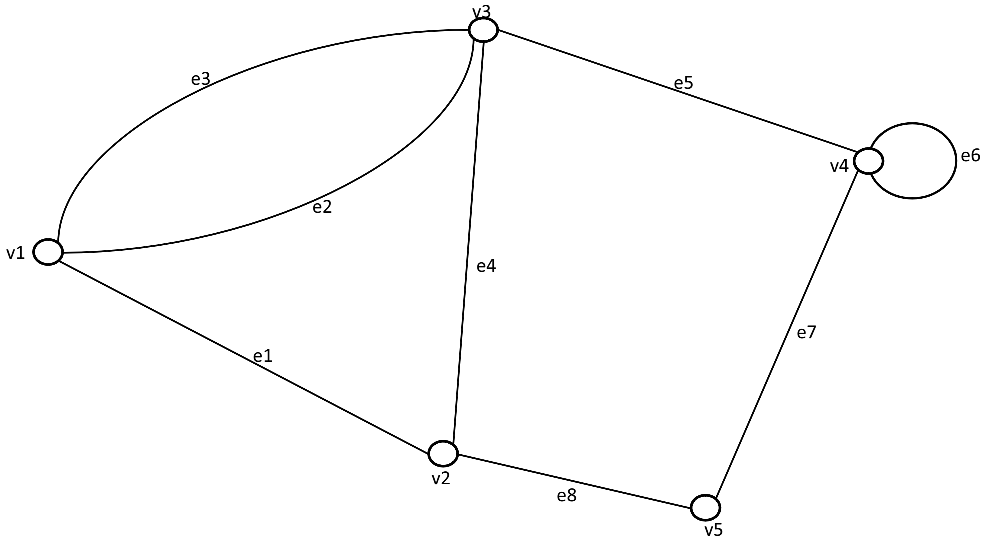


Figure 2.5: Graph used to illustrate walks, trails, paths, and cycles.

Example 2.1.17 *We illustrate Definition 2.1.16 using the graph G = (V, E) with five vertices and seven edges shown in Figure 2.5.*

- $w = (v_1, e_2, v_3, e_3, v_1, e_1, v_2, e_4, v_3, e_4, v_2, e_8, v_5, e_7, v_4)$
  Here w is a walk of lenght seven. Note that w is not a trail, since the edge e_4 appears twice. A walk that is not a trail is clearly not a path. Hence, w is not a path.

- $t = (v_1, e_2, v_3, e_3, v_1, e_1, v_2, e_8, v_5 e_7, v_4)$
  The walk t is a trail of length five. Note that t is not a path, since the vertex v_1 appears twice.

- $p = (v_1, e_2, v_3, e_4, v_2, e_8, v_5, e_7, v_4)$
  The walk p is a path of length four.

- $c_1 = (v_3, e_2, v_1, e_2, v_3, e_4, v_2, e_8, v_5, e_7, v_4, e_5, v_3)$

The walk c_1 is a closed walk of length six.

- $c_2 = (v_3, e_2, v_1, e_3, v_3, e_4, v_2, e_8, v_5, e_7, v_4, e_5, v_3)$
  The walk c_2 is a circuit of length six.

- $c_3 = (v_3, e_4, v_2, e_8, v_5, e_7, v_4, e_5, v_3)$
  The walk c_3 is a cycle of length four.

Figure 2.6 explains the relationship of the items from Definition 2.1.16 in any given graph G. An arrow from one oval down to another means that the first (upper) oval contains the items in the secand (lower) oval, since the items in the first are more general than the items in the secand oval. So, for example, walks are more general than trails. Going down the diagram in figure means that we are restricting the structure more and more. The diagram in Figure 2.6 is actually an example of a directed graph and, in fact, a poset.

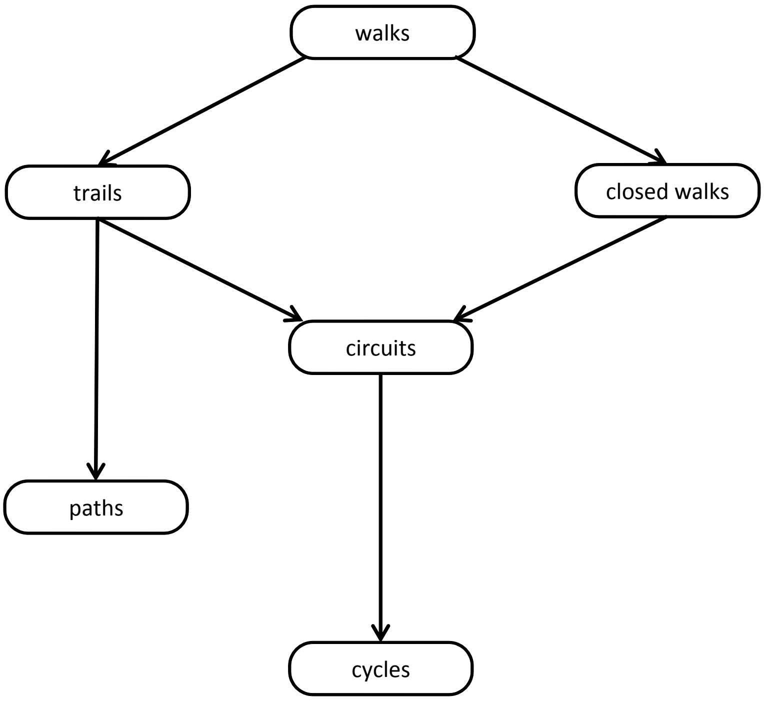


Figure 2.6: Relationships between various subgraphs and walks.

2.1.2.2 Connectivity

Definition 2.1.18 *A graph G is* ***connected****, if for every pair of distinct vertices $u, v \in V$, the graph G has a u,v-path. Otherwise we say that the graph is disconnected.*

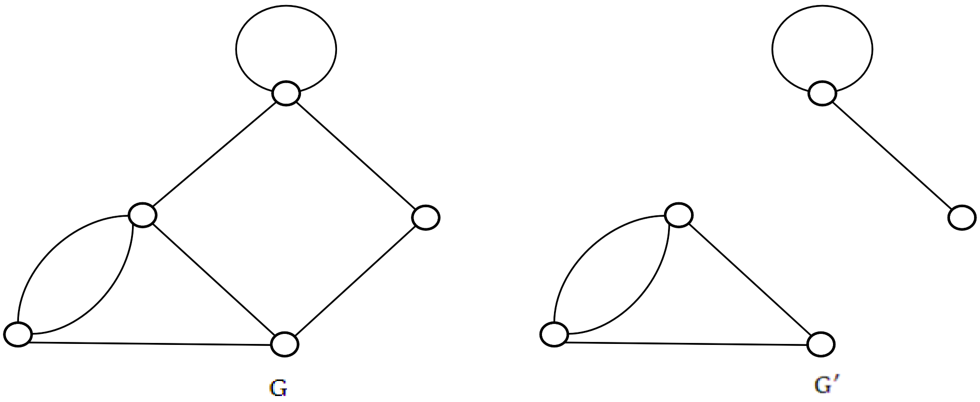


Figure 2.7: Graph G is connected, but $G^{'}$ is not.

As it can be seen in Figure 2.7, the graph $G^{'}$ is a made up of two connected parts. Each segment is subgraph of $G^{'}$ that is itself connected. Such a connected part of a graph is called a *component (or connected component)* of the graph. The following definition states this situation more formally.

Definition 2.1.19 *Let G be a graph. Let $H_1, ..., H_k$ be connected subgraphs of G whose vertex sets and edge set are pairwise disjoint and such that they cover all the vertices and edges of G. That is,*
$V(G) = V(H_1) \cup ... \cup V(H_k)$
$E(G) = E(H_1) \cup ... \cup E(H_k)$
where $V(H_i) \cap V(H_j) = E(H_i) \cap E(H_j) = \emptyset$ for each distinct i and j.
Then each of the subgraphs H_i is called ***component*** *or* ***connected component*** *of G.*

Theorem 2.1.20 *Every graph G has a unique collection of connected component $H_1, ..., H_k$. In particular, the number k of connected components of G is uniquely determined by G.*

Definition 2.1.21 *Let G be a graph. The minimum number of vertices of G, whose removal* *disconnects G, or creates a graph with a single vertex, is called the connectivity of G and is denoted by $\kappa(G)$. If $k \leq \kappa(G)$, then we say that G is $k-$ connected.*

Definition 2.1.22 *The local connectivity $\kappa(x, y)$ of two non-adjacent vertices is the minimum number of vertices seperating x from y. If x and y are adjacent vertices, their local connectivity is defined as $\kappa_H(x, y) + 1$ where $H = G - xy$.*

Theorem 2.1.23 ***(Menger)****: Let $x, y \in G$, $x \neq y$. There exists a set of $\kappa(x, y)$ independent paths between x and y and this set is maximal.*

2.1.2.3 Digraph Connectivity

Basically, there are two natural ways to view connectivity in digraph. One is simply to adapt directly the definition from graphs, and say that a digraph $\vec{G}$ is connected if, and only if, its underlying graph is connected. In that case we say that $\vec{G}$ is *weakly connected*. By a *component* of $\vec{G}$ it is simply mean that the subdigraph induced by the vertices of the corresponding component of the underlying graph G.

The following definition, however, is a more common way of describing connectivity in digraphs.

Definition 2.1.24 *A* ***directed walk*** *$\vec{w}$ in a digraph $\vec{G}$ is an alternating sequence*

$$\vec{w} = (v_0, e_1, v_1, e_2, ..., e_k, v_k)$$

of vertices and directed edges, where for each $i \in \{1, 2, ..., k\}$ the tail and head of e_i are v_{i-1} and v_i, respectively. That is,

$$\eta(e_i) = (v_{i-1}, v_i).$$

The notion of an initial vertex, a final vertex, the lenght, a directed trail, a directed path, and a directed u, v-walk, trail, path are the same as in Definition 2.1.16, but here it is added the words “directed” before each of the words “walk”, “trail”, and “path”. Likewise, the notion of closed directed trail and path are called directed circuit and directed cycle, respectively.

Remark: Since each directed edge has a unique tail and head, there is no ambiguity in writing a directed walk as

$$\vec{w} = (e_1, e_2, ..., e_k),$$

where it is understood that the initial vertex of $\vec{w}$ is the tail of e_1 and the final vertex is the head of e_k.
Likewise, if our digraph $\vec{G}$ has no parallel directed edges, and hence the set of directed edges is given by a subset $E \subseteq V \times V$, then we can write a directed walk as a sequence of vertices

$$\vec{w} = (v_0, v_1, ..., v_k).$$

It can be now stated the more common definition for connectivity in digraphs, called *strong* connectivity, to emphasize its difference from weak connectivity.

Definition 2.1.25 *A digraph $\vec{G}$ is **strongly connected** if for every pair $u, v \in V(\vec{G})$ of distinct vertices there is a directed walk from u to v in $\vec{G}$.*
*The **strong components** of $\vec{G}$ are maximal strongly connected subdigraphs of $\vec{G}$*

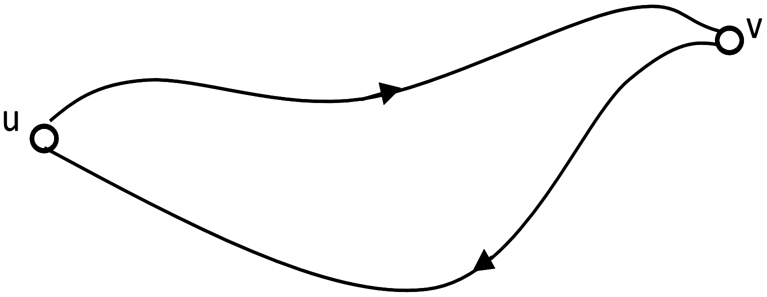


Figure 2.8: The condition for strong connectivity in a diagraph.

Note: The definition 2.1.25 implies that for very pair $u, v \in V(\vec{G})$ of vertices there is a directed path from u to v and a directed path from v to u as well (by reversing the role of u and v.) Also note that these two directed paths, one from u to v and the other from v to u, are not necesssarily edge disjoint!

As it can be seen in Figure 2.8, the condition of digraph $\vec{G}$ being strongly connected implies that for every pair $u, v \in \vec{G}$ of vertices there is a directed walk from u to v in $\vec{G}$, as well as a directed walk from v back to u in $\vec{G}$.

2.1.3 Giant Connected Component

The characteristic of networks discussed so for do not allow us to imagine their global topology. To get real image of a network, we have to know its percolation property. As Figure 2.9 demonstrates, a network may consist of a number of disjoint parts-connected component. The standart notion of percolating cluster and percolation threshold for networks are introduced in the following way.

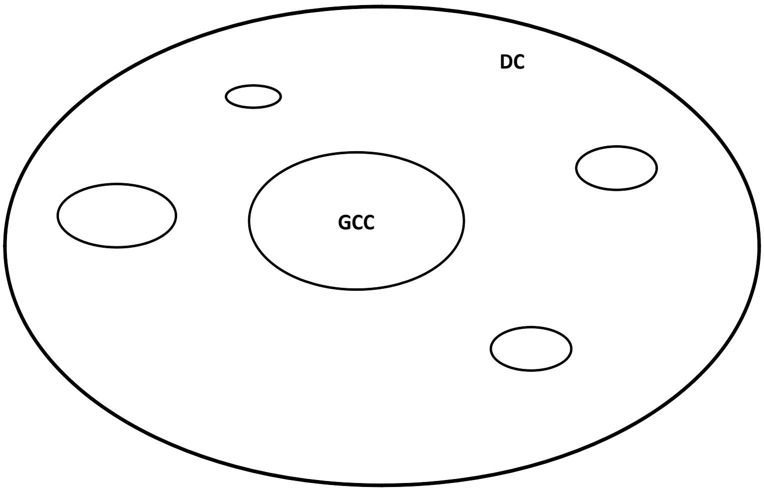


Figure 2.9: The general structure of an undirected network with the giant connected component (GCC). The GCC plays the role of a percolating cluster. The rest of the network includes separate finite-size clusters: finite connected components. Usually, this part is referred to as disconnected components (DC).

To begin with, suppose that the network is undirected. A distinct connected component of a network is a set of mutually reachable vertices. The size of a connected component is the total number of vertices in it. When the relative sizes of all connected components of a network tend to zero as the number of vertices in the network approaches infinity, the network is below the percolation threshold. If the relative size of the largest connected approaches a finite (non-zero) value in the limit of a large network, the network is above the percolation threshold. In such an event, the huge connected component plays the role of a percolating cluster. In graph theory, this is called the **giant connected component (GCC)**. The general structure of an undirected graph, when the giant connected componet is present, shown in Figure 2.9. The rest of the network consists of separate finite connected components. Traditionally, these parts are together referred to as **disconnected component (DC)**.

Definition 2.1.26 ***The Giant Connected Component (GCC)*** *of an undirected graph* G =

(V, E)*, where* V *is the set of all vertices and* E *is the set of all edges, is the maximal set of vertices* $U \subset V$ *such that every pair of vertices u and v in U are reachable from each other.*

The notion of the giant connected components are truly important. They characterize a network as a unit “organisim” and indicate its “health”. For example, an undirected graph is only a set of separate clusters if the GCC is absent.

2.2 Properties of Networks

2.2.1 Clustering Coefficient

Nodes in many real systems exhibit a tendency to cluster, which can be qualified using the clustering coefficient [12], a measure of the degree to which the neighbors of a particular nodes are connected to each other. For example in a friendship network, C reflects the degree to which friend of a particular person are friends with each other as well. Formally the clustering coefficent of node i is defined as

$$C_i = \frac{2n_i}{k_i(k_i - 1)}, \tag{2.6}$$

where n_i *denotes the number of links connecting the* k_i *neighbors of node i to each other.*

Accordingly, it can be defined ***the average clustering coefficient*** *as*

$$< C >= \frac{1}{N} \sum_{i=1}^{N} C_i. \tag{2.7}$$

Note: The average degree $< k >$ and average clustering coefficient $< C >$ depends on the number of nodes and links in the network. By contrast $P(k)$ (see Subsection 2.2.2) are independent of the network’s size and therefore capture a network’s generic features which allows them to be used to classify various networks.

2.2.2 Degree (Connectivity) Distribution

Generally the degree of vertices in random networks are statistically distributed. We define p_k to be the fraction of vertices in the network that have degree k. Equivalently p_k is the probability that a vertex chosen uniformly at random has degree k. A plot of p_k for any given network can be formed by making a histogram of the degree of vertices. This histogram is the degree distribution for the network.

Knowing the degree distribution of each vertices in a network, it easy to find the total degree distribution

$$P(k, N) = \frac{1}{N} \sum_{s=1}^{N} p(k, s, N), \tag{2.8}$$

where $p(k, s, N)$ is the probability that the vertex s in the network of size N has k connections.

The following examples demonstrate typical degree distribution for networks.

The poission distribution:

$$P(k) = \frac{e^{-<k>}(< k >)^k}{k!}. \tag{2.9}$$

Here, $< k >$ is the average degree. A “classical random graph ” asymptotically has just this degree distribution, if its number of vertices approaches infinity under the constraint that the mean degree is fixed.

Classical random graph: This graph is defined by the following simple rules:

- The total number of vertices is fixed.
- Randomly chosen pairs of vertices are connected via undirected edges.

Exponential Distribution

$$P(k) \sim e^{-\frac{k}{<k>}}. \tag{2.10}$$

For instance, this is the degree distribution of “the growing random graph ”.

Growing random graph:

- At each time step, a new vertex is added to the graph.
- Simultaneously, a pair of randomly chosen vertices is connected by an edges.

The Power-law distribution:

$$P(k) \sim k^{-\gamma}, k \neq 0. \tag{2.11}$$

Here, $\gamma \geq 1$ is the exponent (or parameter) of the distribution whose value is typically in the range $2 \leq \gamma \leq 3$, although occasionally it may lie outside these bounds.

The power-law distribution contrast with the Poisson and Exponential distributions. It has no natural scale and, hence, may be called scale-free. Networks with such distribution are called scale-free.

2.2.3 Network Models

Recent technological advances such as availability of computers and others increased the gathering of topological data on large network. This availability of topological data and recent theoretical advances led to great advances in the understanding of the general feature of network structure and development [1], [5], [11], [12]. The existing empirical and theoretical results indicate that complex networks can be divided into two major classes based on their connectivity distribution $P(k)$ (representing the probability that a node in the network is connected to k other nodes): exponential networks and scale-free networks. The main role of the network models is to explain the emergence and behavior of some of the most important network characteristics. As they play a crucial role in shaping the understanding of complex networks, it is needed to pay attention to some of the more important models.

2.2.3.1 Exponential Networks (Random Networks)

This class of networks is characterized by a $P(k)$ that peaks at an average $< k >$ and decays exponentially for large k. Because of this exponential behavior, random networks can be called Exponential network. The most investigated examples of such exponential networks are the random graph model of Erdös and Renyi [11] and the small-world model of Watts and Strogatz [12], both leading to a fairly homogeneous network, in which each node has approximately the same number of links, $k \simeq< k >$. This property is illustrated in Figure 2.13.

The Erdös - Renyi (ER) model of random graph: The simplest complex network model was proposed by Paul Erdös and Alfred Renyi in 1959. The Erdös-Renyi (ER) model of a random network (see Figure 2.11) starts with N nodes and connects each pair of nodes with probability p, creating a graph with approximately $\frac{pN(N-1)}{2}$ randomly distributed links. The distribution of connection of networks generated by this model follows a Poission distribution for large N (see Figure 2.10) which indicates that most nodes have roughly the same number of links, approximately equal to the network's average degree, $< k >$. The tail of the degree distribution $P(k)$ decreases exponentially, which indicates that nodes that significantly deviate from the average are extremely rare.

Despite its elegance, recent findings indicate that ER model can not explain the topological properties of real networks [7], [33]. In contrast to the Poisson distribution, for many social and technological networks the number of nodes with a given degree follows a power-law.

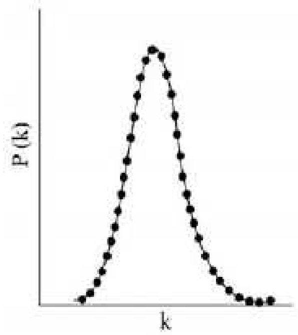


Figure 2.10: Degree distribution for random network. Random network present a peak distribution [33].

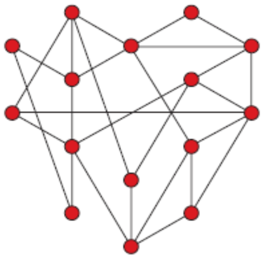


Figure 2.11: Model of a random network [33].

2.2.3.2 Scale-free Networks

Results on the World-Wide Web (WWW) [5], the Internet [13] and other large networks, [14], [15] indicate that many systems belong to a class of inhomogeneous networks (this property is illustrated in Figure 2.13), called scale-free networks, for which $P(k)$ decays as a power-law, that is $P(k) \sim k^{-\gamma}$, free of a characteristic scale. Whereas the probability that a node has a very large number of connections $k \gg< k >$ is practically prohibited in exponential networks, highly connected nodes are statistically significant in scale-free networks.

The Barabasi-Albert (BA) model of scale-free network: The inhomogeneous connectivity distribution of many real netwoks is reproduced by the scale-free model [14], [15] which is based on two basic rules: *growth* and *preferential attachment*. The model start with m_0 nodes. At every time step t a new node with M links is added to the network, which connects to already existing node I with probability $\prod_I = \frac{k_I}{\sum_J k_J}$, where k_I is the degree of node I and J is the index denoting the sum over network nodes. The network that is generated by this growth process has a power-law degree distribution that is characterized by the degree exponent $\gamma = 3$. Such distribution are seen as a straight line on a log-log plot (see Figure 2.12).

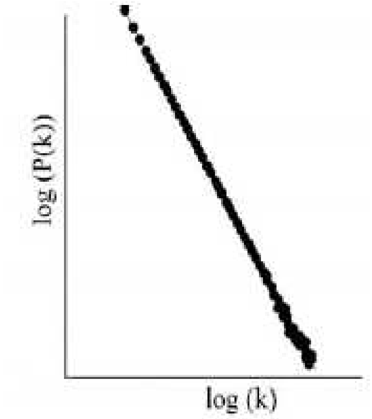


Figure 2.12: Degree distribution for scale-free networks. Scale free networks present a straight line in the log-log plot [33].

Growth and preferential attachment are jointly responsible for the emergence of the scale free property in complex network and these two fundamental process have a key role in the development of real networks. For example, the World Wide Web has grown from 1 to more than 3-billion web pages over a 10-year period (growth process) and on the World Wide Web we are more familiar with highly connected web pages, and therefore are more likely to link them (preferential attachment) [33].

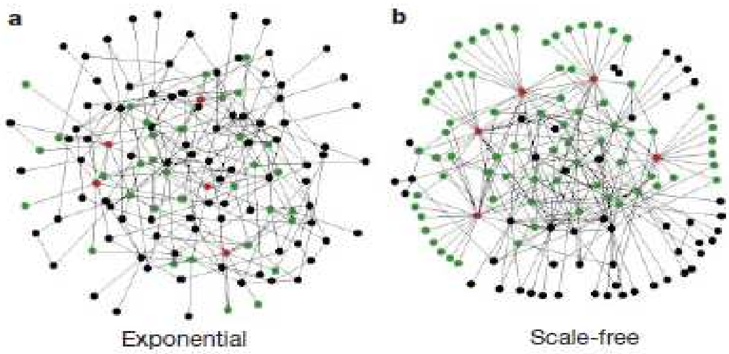


Figure 2.13: Visual illustration of the difference between an exponential and a scale-free network. This figure is taken from [1].

2.2.4 Scale-free Networks

An important information to characterize a graph G is the degree of a vertex i. That is, the number k_i of edges incident with vertex i. Networks with power law degree distribution have been focus of great deal of attention in the literature [16], [17]. They are referred as scale-free networks [14], although it is only their degree distributions that are scale-free (The term scale-free refers to an functional form $f(x)$ that remains unchanged to within a multiplicative factor under a rescaling of the independent variable x. In fact this means power law forms since these are the only solutions to $f(ax) = bf(x)$ and hence power-law and scale-free are for the purposes, synonyms). It indicates the absence of a typical node in the network (one that could be used to characterize the rest of the nodes). This is in strong contrast to random networks, for which the degree of all nodes is in the vicinity of the average degree, which could be considered typical. However, scale-free networks could easily be called scale-rich as well, as their main feature is the coexistence of nodes of widely different degrees (scales), from nodes with one or two links to major hubs.

Barabasi et al. focused their attention on $P(k)$, the degree distribution of a network, and showed that many real large network as the World Wide Web, the internet, metabolic and protein networks are scale free that is their degree distribution follows a power law for large k [7], [5], [18]. That is a scale-free network is one in which the probability of a node having k connection is proportional to $k^{-\gamma}$, where γ is the degree exponent ($2 \leq \gamma \leq 3$). In this sense scale-free network are heterogenous. Heterogenous means that while most vertices are lowly connected, a huge number of edges is concentrated in a small number of nodes. Thus network's properties are determined by hubs (the most highly connected nodes) (see Figure 2.14).

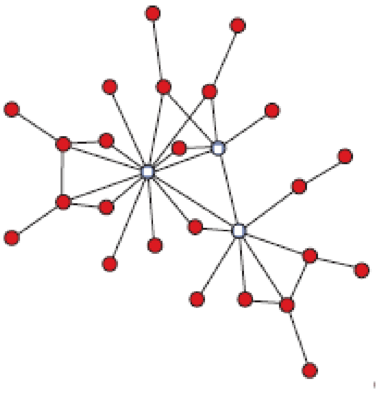


Figure 2.14: The network's properties are determined by hubs (white nodes) [33] .

2.3 Protein Domain Networks

Since proteins are the most essential structural components for living things to maintain their cell functions properly, proteins are widely studied in biology. Although proteins are unique, they share certain common properties. For example, well-defined regions within a protein can evolve, function and fold independently from the rest of the protein and have their own function. They are called protein domains, and served as protein building blocks. These domains and nature of their interactions determine the function of the protein. Thus, it can be said that domains are basic evolutionary units of proteins.

Many biological systems can be represented by networks [20], [22], [26]. One of these networks is protein domain networks which include two networks type; protein domain interaction networks and protein domain co-occurence networks. The domain architecture of proteins was studied by considering protein domains as nodes and their co-occurrence in proteins as links, documenting again the emergence of a scale-free architecture [20]. Although methods and sources of domain information in [20] were different, the scale-free features of the networks were found to be robust [18].

In this work, we focus on Protein domain co-occurrence network of yeast (S.cerevisiae) which is formally defined by a set of nodes consisting of all domains which occur in the protein sequence of the yeast proteome. Two domains are regarded as being indirectly linked if they occur in one of these protein sequences [21]. Protein domain co-occurrence network of yeast

(S.cerevisiae) which is obtained from Biomart (Pfam domain database) data management system is weighted and undirected.

CHAPTER 3

NETWORK RESILIENCE

The property of resilience of networks to the removal of their vertices which is related to degree (connectivity) distribution is of great interest in the literature. Most of the networks is considered for their function on their connectivity, that is, the existence of paths between pairs of vertices. If we remove vertices from a network, the length of these paths will increase, and eventually vertex pairs will become disconnected and some of the vertices will become unreachable. As a result the network communication between them will be destroyed. Response of resilience of networks to such vertex removal can vary according to network model which can be classified by their degree distributions.

The presence of a giant connected component in a network indicates the unity of a network. In other words, if a network does not have a giant connected component, we can say that that network consists of disconnected clusters (components). If we wonder answer of the question ”When networks are tried to be destroyed what is the response of networks to such an attack?”, the variation of the giant component must also be studied. if such an attack does not crucially diminish the giant component, the damage is not serious , but the damage is fatal if it eliminates the giant component.

There are also a variety of different ways in which we remove vertices and different networks show varying levels of resilience to such a vertex removal. For example, one could remove vertices at random from a network,

Failure - Nodes are removed randomly which might be represent random perturbations, environmental factors etc.

or one could target some specific set of vertices, such as those with the highest degrees.

Attack - Selected nodes are removed from the network which might represent more organized and selective effects to the network like virus attacks.

The way the nodes are chosen during an attack is called an *attack strategy*. There are different attack strategies which is introduced in [1], [2]. But, the most widely studied attack strategy has been introduced in [1]. At each step the nodes with maximal degree is removed by decreasing order of their degree. This attack can be called as the *classical attack strategy* [2].

There are several measures (methods) to analyze the resilience of the network to failures or attacks. Some of these are introduced in the following section.

3.1 Methods to Measure Error and Attack Tolerance of Complex Networks

3.1.1 Average Shortest Path

Average shortest path is a typical network statistic to measure the network distance of a network. To define the average of the shortest path lengths between two vertices, L, it is needed first to construct the shortest path length d_{ij} between two vertices measured as the miminum number of edges in the network for all nodes from i to j. By definition, $d_{ij} \geq 1$ with $d_{ij} = 1$ if there exists a direct edge between i and j. The characteristic path length L of graph G is defined as the average of the shortest path lengths between two generic vertices:

$$L(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} d_{ij}, \tag{3.1}$$

where N is the number of nodes in the network.

As it can be seen from Equation (3.1), this definition is valid only if G is totally connected, which means that there must exist at least a path connecting any couple of vertices with a finite number of steps. Otherwise, when from i^* it can not be reached to j^* then $d_{i^* j^*} = +\infty$ and consequentely L as given in (3.1), being divergent. When studying how the resilience of a network are affected by the removal of nodes, one often encounters non-connected networks.

In such cases the alternative formalism can be used such that the efficiency of network [3].

If the network is connected, the average shortest path is greater when the network undergoes disruptions. As some of the nodes are deleted, the shortest path may be longer due to the deletion of these nodes, because some of these nodes may form part of the shortest path in the network before disruptions. However, if the disruptions of the network are so severe that the network is fragmented, the shortest paths between the fragmented nodes and other nodes are infinity. Then the average shortest path becomes infinity. In this situation, the average shortest path can only show that the network is fragmented [19].

3.1.2 Efficiency of the Network

To define the efficiency of G, suppose that every node sends information along the network, through its edges and assume that the efficiency ϵ_{ij} in the communication between node i and j is inversely proportional to the shortest distance: $\epsilon_{ij} = \frac{1}{d_{ij}} \ \forall\ i, j$. With this definition, when there is no path in the graph between i and j, $d_{ij} = +\infty$ and consistently ϵ_{ij}= 0. The global efficiency of the graph G can be defined as:

$$E_{glob} = \frac{\sum_{i \neq j \in G} \epsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}, \tag{3.2}$$

the global efficiency is normalized, that is: $0 \leq E_{glob}(G) \leq 1$. The maximum value of the efficiency $E_{glob}(G) = 1$ are obtained in the ideal case of a completely connected graph, i.e. in the case in which the graph G has all the $\frac{N((N-1)}{2}$ possible edges and $d_{ij} = 1 \ \forall\ i, j$ [3].

As it is mentioned, if large number of nodes are removed, the network becomes unconnected (consequentely L as given in 3.1,being divergent, is an illdefined quantity). To overcome this problem, the efficiency of the network can be used to measure the response of the networks to external factors because network efficiency can be extended to non-connected networks. That is, using of the efficiency measure to characterize scale-free networks allows to avoid problems due to the divergence of the average distance [3].

The global efficiency of the graph G decreases when the network undergoes disruptions [3].

3.1.3 Diameter of the Network

Diameter is defined as the longest shortest path between any pair of nodes of a network. It characterizes the ability of two nodes to communicate with each other.

$$Dia = max\{d_{ij} \mid i, j = 1, 2, ..., N\}, \tag{3.3}$$

where N is the number of nodes in the network.

Networks with a very large number of nodes can have quite a small diameter: the smaller d is, the shorter is the expected path between them; for example, the diameter of the WWW, with over 800 million nodes, is around 19 [5] whereas social networks with over six billion individuals are believed to have a diameter of around six [6].

3.1.4 Size of the Giant Connected Component

The resilience of the network to failures or attacks can be analyzed by studying how the size of the largest connected component varies as a function of the number of removed nodes. That is, during the network fragmentation process it can be analyzed the disruption of the network topology by measuring properties of the giant cluster that remains connected, including size S as fraction of the size of initially-connected network.

This method is introduced by Barabasi et al. [1] in the following way: “When nodes are removed from a network, clusters of nodes whose links to the system disappear may be cut off (fragmented) from the main cluster. This fragmentation process is investigated by measuring the size of the largest cluster, S, shown as a fraction of the total system size, when a fraction f of the nodes are removed either randomly or in an attack mode. It is found that for the exponential network, as f increases, S displays a threshold-like behaviour such that for $f \geq f_c$, $S \simeq 0$, where f_c is the threshold value. However, the response of a scale-free network to attacks and failures is rather different. For random failures no threshold for fragmentation is observed; On the other hand, the response to attack of the scale-free network is similar (but swifter) to the response to attack and failure of the exponential network.”

3.2 Literature Review on the Comparison of the Behavior of Scale Free and Random Network under Failures and Attacks

As we mentioned, complex networks can be divided into two major classes according to the connectivity distribution $p(k)$: exponential networks and scale-free networks. Exponential networks are homogeneous in connectivity, which means most nodes in network have approximately the same number of connections around $< k >$, $k \approx< k >$. In contrast, many real networks belong to a class of inhomogeneous, scale-free networks. Different from exponential networks, in scale-free network, a few highly connected nodes ($k \gg \quad < k >$) exist and thus play a significantly important role in the network's connectivity.

Previous study [1], [8], [9], [10] has shown that scale-free networks display an unexpected degree of robustness, i.e., the ability of their nodes to communicate being unaffected by even unrealistically high failure rates. However, these networks are extremely vulnerable to intentional attacks, i.e., to the removal of a few of highly connected nodes. This property is rooted in their heterogenity. On the other hand, evolving networks with exponential connectivity distribution (exponential network) are not as robust to random failures, but more resilient to intentional attacks. This property is due to their homogeneity, exhibit a similar tolerance with respect to errors and attacks.

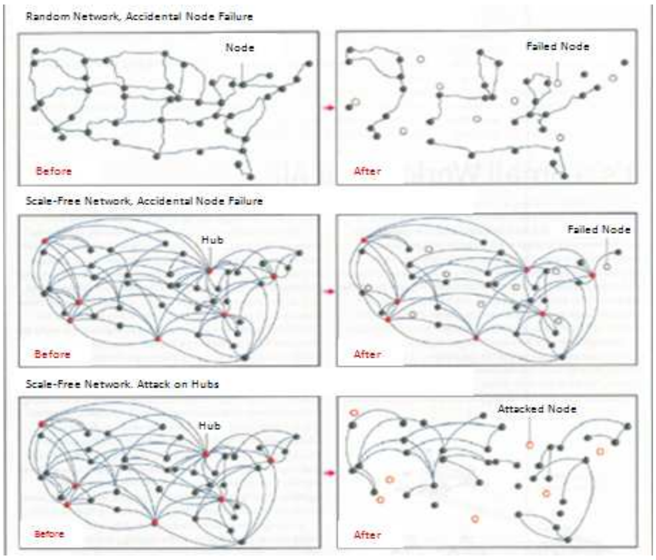

Random Network, Accidental Node Failure
Node
Failed Node
Before
After
Scale-Free Network, Accidental Node Failure
Hub
Failed Node
Before
After
Scale-Free Network. Attack on Hubs
Hub
Attacked Node
Before
After

Figure 3.1: The simulations of the effects of failures and attacks on scale-free and exponential random networks. This figure is taken from [7].

By Barabasi et al. [7], the effects of failures and attacks on scale-free and exponential random networks are simulated. In this simulation shown in Figure 3.1, the diagrams are constructed using U.S. highway system which resembles to random networks and U.S. airline system for scale-free networks. When the nodes are removed randomly, scale-free networks seem to be much more robust than exponential random networks in terms of network connectivity, but they seem more vulnerable to intentional attacks. In failure, since nodes are removed randomly, the probability that the removed nodes with low-degree is higher than the removed nodes with high-degree (hubs), as a result effects of failure on network connectivity is expected to be small. On the other hand, in a intential attack the nodes with high degree are removed, thus effect of attack on network connectivity is expected to be big, that is the network connectivity can be heavily damaged. These results can be observed from the simulation shown in Figure 3.1. The first two diagrams show the effects of random failure of nodes on a random network (U.S. highway system) connectivity. The second two diagrams show the effect of random node removal on a scale free network (U.S. airline system) connectivity. The last two diagrams show the effect of a hub attack on the scale-free network. Comparing the second and the last two diagrams, it is seen that the network connectivity is heavily damaged under attacks. All these results observed in this analysis can be explained by presence of hubs in scale-free networks [7].

Barabasi et al. [1] again investigate the effect of failures and attacks on scale free and exponential random networks. But in this case changes in the diameter of the two types of networks are examined as a function of the fraction of nodes being removed. They compare the exponential (E) and scale-free (SF) network models, each containing $N = 10,000$ nodes and 20, 000 links under failures and attacks in Figure 3.2 **(a)**. The triangle and square symbols in Figure 3.2 **(a)** correspond to the diameter of the exponential and scale-free networks respectively when a fraction f of the nodes are removed randomly (error tolerance). The diamond and circle symbols in Figure 3.2 **(a)** show the response of the exponential and the scale-free networks to attacks respectively, when the most highly connected nodes are removed. In Figure 3.2 **(b)**, changes in the diameter of the Internet (scale-free network) under random failures (squares) and attacks (circles) are shown. In Figure 3.2 **(c)**, changes in the diameter of the World-Wide Web (scale-free network) under random failures (squares) and attacks (circles) are shown. The results in Figure 3.2 agree with the previous observation that scale-free networks are more robust than exponential networks under failures, but more vulnerable to attacks. It can be also observed that failures or attacks on exponential random networks causes almost the same amount of damage to the network.

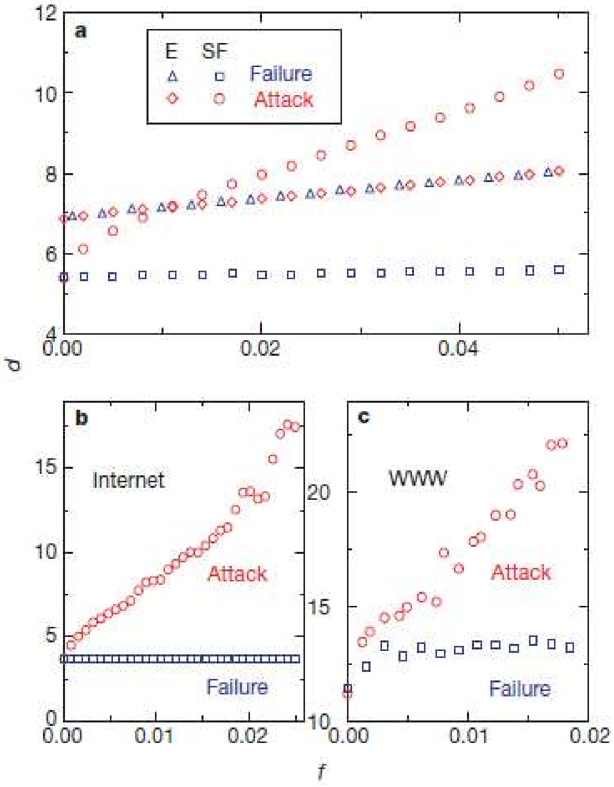


Figure 3.2: Changes in the diameter of the network as a function of the fraction of the removed nodes. This figure is taken from [1].

Barabasi et al. [1] continue to investigate the effect of failures and attacks on the two types of networks by studying another method. They measured the size of the giant connected componet and the average size of the disconnected component (or isolated components) as a function of the fraction of nodes are removed either in a targeted way or randomly. In Figure 3.3, changes in the relative size of the largest cluster S (open symbols) and the average size of the isolated clusters $< s >$ (filled symbols) are shown as function of the fraction of removed nodes f for the same systems. The size S is defined as the fraction of nodes contained in the largest cluster. In Figure 3.3 **(a)**, fragmentation of the exponential network under random failures (squares) and attacks (circles) are shown. In Figure 3.3 **(b)**, fragmentation of the scale-free network under random failures (squares) and attacks (circles) are shown. The inset shows the error tolerance curves for the whole range of f, indicating that the main cluster falls apart only after it has been completely deflated. In Figure 3.3, (**c**) and (**d**) show the effect of failure and attack on the Internet and www, respectively. Again, It is observed that scale-free networks are more robust than exponential networks and vulnarable to attacks, in contrast to exponential network behave the same under failures and attacks.

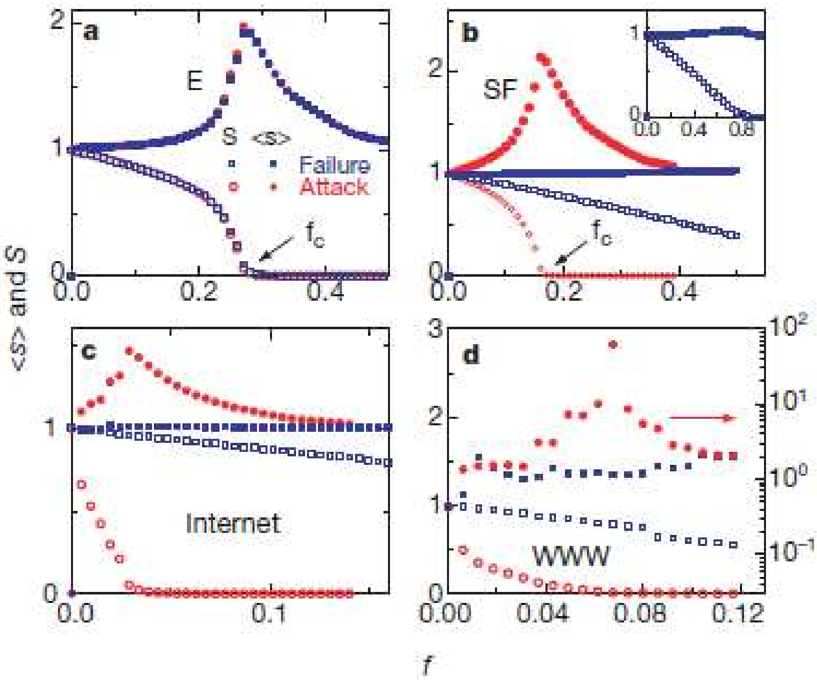


Figure 3.3: Changes in the relative size of the largest cluster S (open symbols) and the average size of the isolated clusters $< s >$ (filled symbols) as function of the fraction of removed nodes f for the same systems. This figure is taken from [1].

CHAPTER 4

APPLICATIONS

In the previous section, the methods to measure the attack tolerance of complex networks are presented. In order to use these methods and analyze the network, some changes must be made on the network. Protein domain co-occurence network of yeast (S.cerevisiae) which is formatted into readable format by Pajek program is used to analyse the attack tolerance of several networks. In the first part of this chapter, we first briefly introduce the structure of Pfam protein domain co-occurrence network of yeast. Then, we mention about those changes made on the network by using Pajek program[1]. In the second part of this chapter we give some statistics of protein domain co-occurrence network of yeast. In this thesis, we compare the attack tolerance of the several networks which is obtained from protein domain co-occurence network of yeast. In the last part of this chapter, how we get those several network from protein domain co-occurence network of yeast and the analysis of the attack tolerance of those networks are introduced. In order to get these several networks which have exactly the same connectivity distribution with Protein domain co-occurrence network of yeast, a program written in MATLAB is used.

4.1 Descriptions and Preparations of the Network Used in the Applications

Many biological systems can be represented by networks [20], [22], [26]. One of these networks is protein domain networks which include two network types; protein domain interaction networks and protein domain co-occurence networks. As we mentioned in Section 2.3, a protein domain is a part of protein sequence and structure that can evolve, function, and exist independently of the rest of the protein chain. Each domain forms a compact three-

[1] Pajek is a program for large -network analysis and visualization (Batagelj and Mrvar 1998).

dimensional structure and often can be independently stable and folded. Many proteins consist of several structural domains. One domain may appear in a variety of evolutionarily related proteins.

Table 4.1: Some examples of Pfam domains

Domain label (accession)	Description of domain
PF00069	Protein kinase domain
PF00730	HhH-GPD superfamily base excision DNA repair protein
PF02844	Phosphoribosylgycinamide synthetase, N domain

Protein domain co-occurrence network of yeast[2] which is used in this application is constructed from Biomart data management system which links to Pfam domain database. In Pfam domain database[3], each domain is labeled. Table 4.1 demonstrates some examples of such domains.

Table 4.2: Pfam protein domain co-occurrence network of yeast obtained from Biomart which links to Pfam database.

domain1	domain2	occurrence in proteins
PF06747	PF08583	1
PF01288	PF00809	1
PF00293	PF05026	1
PF00293	PF09297	1
PF01118	PF02774	2
PF01119	PF08676	2
PF09261	PF07748	1
PF01053	PF01212	1
PF05436	PF04648	2
PF08022	PF08030	8
PF00033	PF00032	1
PF01734	PF00027	1
PF01237	PF00023	2
PF00730	PF00633	1
PF00730	PF07934	1
PF01233	PF02799	1
PF01232	PF08125	1
.	.	.
.	.	.
.	.	.

[2] The construction of Pfam domain co-occurrence network of yeast was made by Stefan Wuchty.

[3] The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs).

In Pfam domain co-occurrence network of yeast, domains are connected if they co-appear in the same proteins in yeast. Also, each link comes with a weight, reflecting the number of proteins the domains co-appeared in. Table 4.2 shows Pfam domain co-occurrence network of yeast before we make some changes on this data. In this table the domain in the first column (domain1) are connected with the domains in the second column (domain2) and the third colum reflects the number of proteins the domains co-appeared in. In Appendix A, the whole Pfam domain co-occurrence network of yeast can be found. In order to analyse this network, we made some changes on the network by using Pajek program. The structure of a Pajek network file is a simple text file that can be typed out in any word processor that exports plain text (see [29] for detailed description of Pajek program). In this simple text file, we attribute serial numbers to the vertices ranging from 1 to the number of vertices and Pajek automatically labeled the vertices. Table 4.3 shows the structure of the pajek domain co-occurrence network file[4]. In this tabel, first, the data file specifies the number of vertices ("vertices 1007"). Then, each vertex is identified on a separate line by a serial number, a textual label (enclosed in quotation marks (" ")) and three real numbers between 0 and 1, which indicate the position of the vertex in three-dimensional space if the network is drawn. This information is unnecessary for us so we ignore these numbers. However, it is important to note that the text label is crucial for the identification of vertices, because serial numbers of vertices may change during the analysis. We ignore the weight of the network to simplfy the analysis. As a result our network is undirected and unweighted network. In addition to this, our network is a simple network that has no self-loops or multi-edges. After the network are formatted in pajek network file format, the network are ready for analysis.

[4] This is a huge network file. It is avaliable, if you contact me via e-mail: doguz@metu.edu.tr

Table 4.3: Partial listing of protein domain co-occurence network of yeast data file for Pajek.

```
*Vertices 1007
    1       “v1”                 0.1000   0.5000    0.5000
    2       “v2”                 0.1000   0.4975    0.5000
    3       “v3”                 0.1000   0.4950    0.5000
    4       “v4”                 0.1001   0.4925    0.5000
    5       “v5”                 0.1001   0.4900    0.5000
    6       “v6”                 0.1002   0.4875    0.5000
    7       “v7”                 0.1003   0.4850    0.5000
    8       “v8”                 0.1004   0.4825    0.5000
    9       “v9”                 0.1005   0.4800    0.5000
   10       “v10”                0.1006   0.4775    0.5000
    .        .                     .        .         .
    .        .                     .        .         .
    .        .                     .        .         .

  *Edges
  790      945        1
  339      260        1
  103      736        1
  103      1002       1
  308      499        2
  309      966        2
  998      840        1
  296      327        1
  760      698        2
  854      856        8
    .        .        .
    .        .        .
    .        .        .
```

Our network is unconnected network; it has 231 disconnected components and a giant connected component. Since we use the method; “size of the giant connected component” to analyze the attack tolerance of the networks, we extract the giant connected component of protein domain co-occurence network of yeast by using Pajek program. After all we are ready to produce several randomly modified networks, which have the same connectivity distribution with protein domain co-occurrence network, from the giant connected component of yeast network by using MATLAB (Part of the MATLAB code used in this process is written by me and can be found in Appendix B).

4.2 Statistics of Protein Domain Co-occurrence Network of Yeast

To capture the network's generic features and to get an idea about the structure of the network, we calculate some statistics of protein domain co-occurrence network of yeast.

Table 4.4: Statistics of protein domain co-occurrence networks of yeast (S.cerevisiae).

organism	N_{nodes}	N_{edges}	N_{cc}	N^{gc}_{nodes}	N^{gc}_{edges}	$< k >$	C	$< k >^{gc}$	γ
S. cerevisiae	1007	1280	231	334	556	2.54	0.39	3.33	1.5

Table 4.4 summarizes the basic statistics of the domain co-occurrence network of S.cerevisia, which are calculated with the help of Pajek and Matlab, that contain a giant connected component 'gc' incorporates the majority of the domain N^{gc}_{nodes}, co-existing with many small, connected component N_{cc} and N^{gc}_{edges} is the number of edges in the largest component. Our network is not a huge network, 1007 vertices and 1280 edges, comparing to the other networks in [1], [5], [13], [24] and has a giant connected component with 334 vertices. Average degree $< k >$ and clustering coefficient C of organisims are compared in [25] and it was found that as level of organisims development increase both the average degree $< k >$ and the clustering coefficient C gradually increases. We do not use the information of average degree $< k >$ and clustering coefficient C anywhere in this anaysis. We just gave these values as an information about the network.

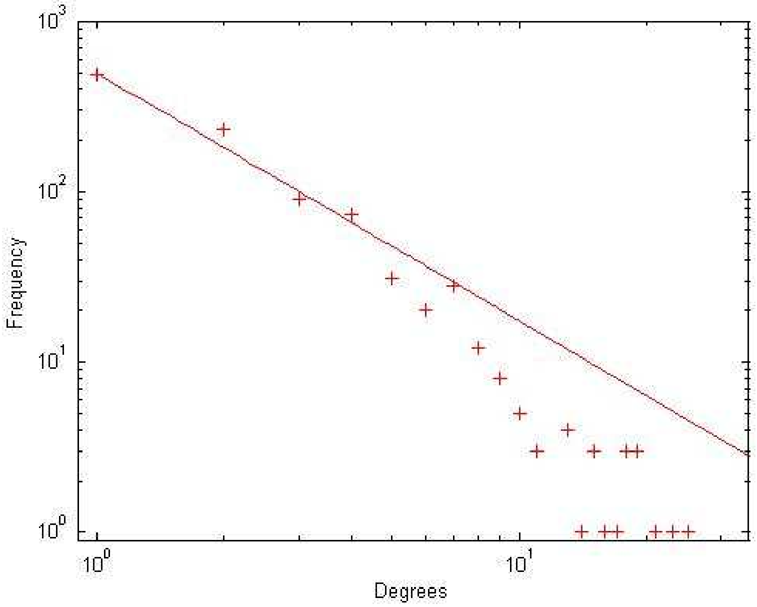


Figure 4.1: Protein domain co-occurrence network of S.cerevisiae displays scales free behaviour. A network feature is characterized by the power law in the degree distribution $P(k) \sim k^{-\gamma}$ (see Table 4.4 for detailed values).

As it was mentioned, the nodes are domains and two domains are connected by an undirected edge if they occur together in one protein at least once. These connections define the edge set of the network. Therefore the degree of a domain is the number of other domains to which it is connected. In our network, frequency distribution of degree reveals the presence of scale free topology. Thus frequency distribution follows a power law $P(k) \sim k^{-\gamma}$. The degree distribution $P(k)$ of the protein domain co-occurrence network of S. cerevisiae follows a power law degree distribution with exponent $\gamma = 1.5$ (Figure 4.1). The result of this analysis is that the scale free network topologically is dominated by few highly connected hubs (most highly connected nodes). In addition to this, empirically, domain co-occurrence network displays power law degree distribution, power-law distributed, resulting in few vertices having many edges and many vertices having few edges. In Table 4.5, the number of nodes which have one link (edge) is 485, and two links is 232, i.e., the number of nodes decreases as the links they have increase. As it can be seen from the Table 4.5, at the end of the table the number of nodes having 25 links is one. Thus, Table 4.5 shows emprically that the network displays power law degree distribution.

Table 4.5: Frequency distribution of degree in protein domain co-occurance network of yeast (S.cerevisiae).

Degree	Freq	Freq%	CumFreq	CumFreq%	Representative(Vertex Label)
1	485	48.1629	485	48.1629	v5
2	232	23.0387	717	71.2016	v8
3	91	9.0367	808	80.2383	v11
4	73	7.2493	881	87.4876	v3
5	31	23.0387	912	90.5660	v24
6	20	1.9861	932	92.5521	v20
7	28	2.7805	960	95.3327	v4
8	12	1.1917	972	96.5243	v7
9	8	0.7944	980	97.3188	v2
10	5	0.4965	985	97.8153	v13
11	3	0.2979	988	98.1132	v63
13	4	0.3972	992	98.5104	v69
14	1	0.0993	993	98.6097	v35
15	3	0.2979	996	98.9076	v45
16	1	0.0993	997	99.0070	v558
17	1	0.0993	998	99.1063	v921
18	3	0.2979	1001	99.4042	v6
19	3	0.2979	1004	99.7021	v1
21	1	0.0993	1005	99.8014	v21
23	1	0.0993	1006	99.9007	v93
25	1	0.0993	1007	100.0000	v133
Sum	1007	100.0000			

4.3 Analysis of Attack Tolerances of the Networks

In this thesis, we compare several networks exhibiting scale free structure, which have exactly the same connectivity with the original network, under attacks. We analyze the robustness of the network to attacks by studying how the size of the largest connected component varies as a function of the number of removed nodes.

Even the connectivity distribution is an important indicator of a network's qualitative features, different networks with the same connectivity distribution do not need to have the same attack tolerances. In addition to this, it can be considered that the networks with same connectivty distribution have higher attack tolerance as we organize the same resources in a better way. Then, it can be checked for the variations of attack tolerance of the networks with the same connectivity distributions. Furthermore, we investigate whether there is an evolutionary mechanism for having networks with higher or lower attack tolerances for the same connectivity distribution. For these purposes, we wrote an algorithm to produce several randomly modified networks having the same connectivity distribution with the original network , then as we attack those networks we collect one of those network with a strategy which will be mentioned in this chapter. Before we check these assumptions and investigations, we want to mention about our attack strategy and the method to measure the attack tolerance of the network.

Attack Strategy: The way the nodes are chosen during an attack is called an *attack strategy*. Some attack strategies introduced in [1], [2]. We use clasical attack strategies which introduced in [1]. In this attack strategy, we first begin to remove the most highly connected nodes and continue to remove nodes by decreasing order of their degree.

The Method: Size of the Giant Connected Component: This method introduced by Barabasi et al. in [1] is; "When nodes are removed from a network, clusters of nodes whose links to the system disappear may be cut off (fragmented) from the main cluster. This fragmentation process is investigated by measuring the size of the largest cluster, S, shown as a fraction of the total system size, when a fraction f of the nodes are removed either randomly or in an attack mode. It is found that for the exponential network under attacks and failures and the scale free networks under attack (see Subsection 3.1.4), as f increases, S displays a threshold-like behaviour such that for $f \geq f_c, S \simeq 0$, where f_c is the threshold value." We slightly modified this method in the following way: We removed a fraction f of the nodes in an attack mode (the most highly connected nodes removed first and continue remove nodes by decreasing order of their degree) like the method introduced in [1]. But, all nodes which were removed in our analysis are belong to the giant connected component. Since other most connected nodes not belong to the giant connected component do not effect the size of giant connected component, we ignored these nodes. In doing so, we accelerated the analysis. Also, Concerning the thresholds values, we considered that the threshold was reached whenever the size of the giant connected component of the network becomes smaller than 2% of the whole system size and 5% of the begining size of the giant connected component of the network.

The Strategy of Analysis: First produce four modified networks from the original network having the same connectivity distribution with the original network by applying the algorithm mentioned below (also see Table 4.6 and in Appendix B, it can be found whole matlab code for this algorithm). After finding attack tolerances of those networks, we collect the least vulnerable network (displaying higher attack tolerance with respect to the others) and the most vulnerable network (displaying lower attack tolerance with respect to the others) among those networks under attack. At this point we divide the analysis into two part:

In the first part, we produce four modified networks from the most vulnerable network by applying the algorithm. After finding attack tolerances of those networks, we collect the most vulnerable network among those networks. We continue appliying the same procedure *n* times (in this analysis, we applied it 10 times). *In the second part*, we produce four modified networks from the least vulnerable networks by applying the algorithm. After attack those networks, we collect the least vulnerable network among those networks. We continue appliying the same procedure *n* times (in this analysis, we applied it 10 times).

It is important to note that all networks in this analysis have the same connectivity distribution with the orginal networks. The strategy of the analysis mentioned above is illustrated in Figure 4.2

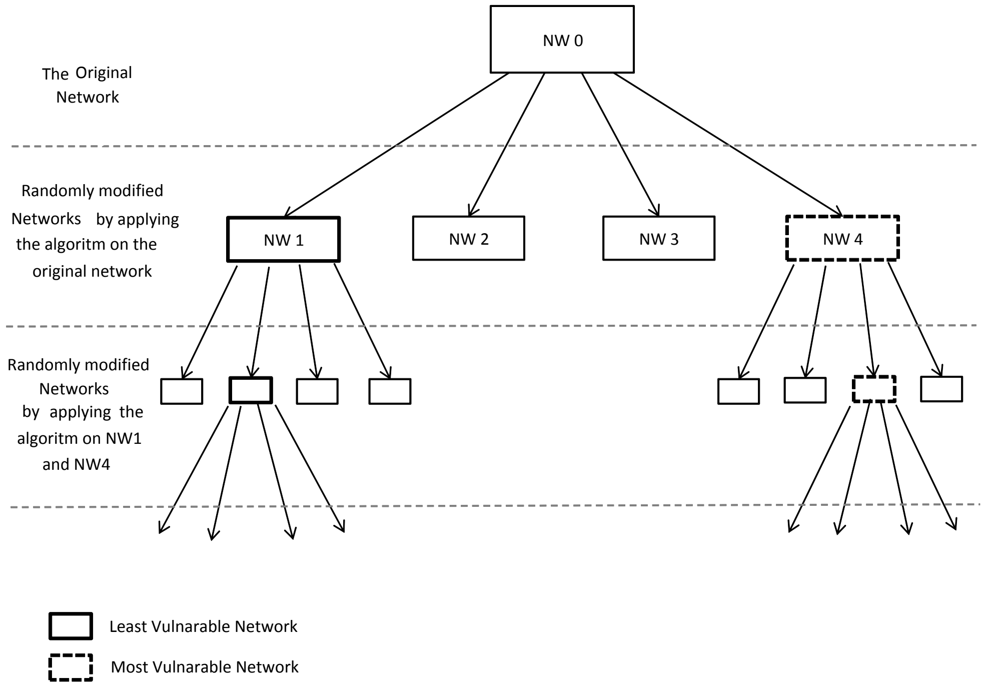

NW 0
The Original Network
Randomly modified Networks by applying the algoritm on the original network
NW 1
NW 2
NW 3
NW 4
Randomly modified Networks by applying the algoritm on NW1 and NW4
Least Vulnarable Network
Most Vulnarable Network

Figure 4.2: Illustration of the strategy of analysis.

To check our assumptions mentioned above, we wrote an algoritm (Table 4.6). In this algorithm, we randomly change the links of the nodes in protein domain co-occurrence network of yeast conserving the connectivity of the network. We can explain this process in the following way: we randomly find a node say n_1 has 4 links ($deg(n_1) = 4$) and a node say n_2 has 2 links ($deg(n_2) = 2$) which they are connected. After we found these two nodes we randomly search for two nodes say n_3 and n_4, which must be unconnected, must have degree $deg(n_1) - 1 = 3$ and $deg(n_2) - 1 = 1$, respectively. After found those nodes, we break the link of n_1 an n_2 and we connect n_3 and n_4. Now n_1 and n_2 have degree $deg(n_1) = 3$ and $deg(n_2) = 1$ respectively while degree of n_3 and n_4 become $deg(n_3) = 4$ and $deg(n_4) = 2$. As a result the connectivity of the network does not changed. This process is illustrated in Figure 4.3. We applied this process 3% of the size of the giant connected component.

Example 4.3.1 *We illustrate* ***the algorithm*** *using the graph G = (V, E) with seven vertices and seven edges shown in Figure 4.3. Here we see that* $G = (V, E)$ *with* $V = \{v_1, ..., v_7\}$, $E = \{(v_1, v_2), (v_1, v_3), (v_1, v_4), (v_1, v_5), (v_2, v_5), (v_3, v_6, (v_3, v_7)\}$. *After applying the algorithm, it is obtanied the modified graph* $G^{'} = (V^{'}, E^{'})$ *with seven vertices and seven edges, where* $V^{'} = \{v_1, ..., v_7\}$, $E^{'} = \{(v_1, v_3), (v_1, v_4), (v_1, v_5), (v_2, v_5), (v_3, v_4), (v_3, v_6, (v_3, v_7)\}$.

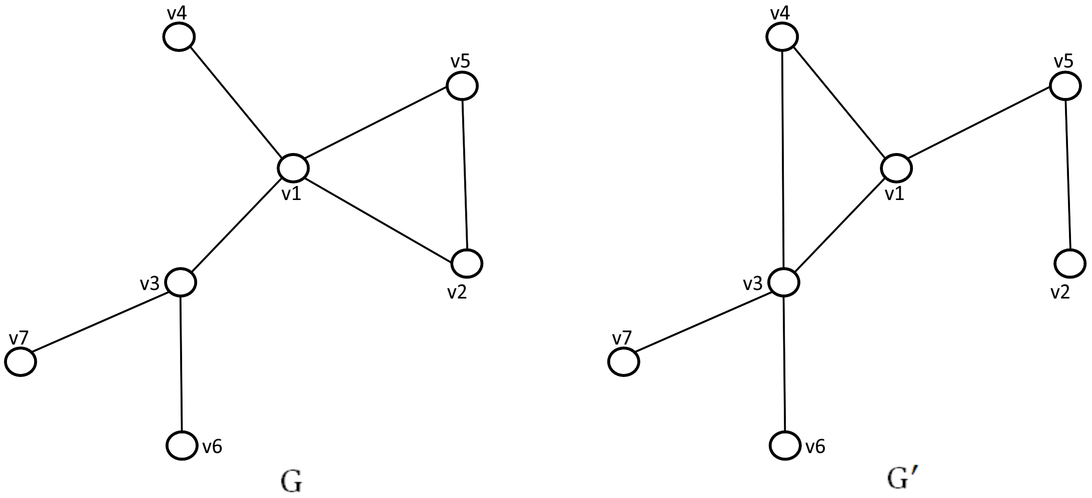


Figure 4.3: Graph used to illustrate the algorithm.

Table 4.6: The algorithm of the analysis.

Algorithm: An algorithm for Randomly modifying the connections of vertices in a Graph $G = (V, E)$, where $V = \{v_1, v_2, ..., v_n\}$, $E = \{(v_i, v_j)\}$, $i, j = 1, 2, ..., n$.

Input: A simple, connected, undirected graph $G = (V, E)$

Output: A simple, connected, undirected randomly modified graph $G^{'} = (V^{'}, E^{'})$

```
1    Find adjancency matrix $A(G)_{n\times n}$ of $G$.
2    Compute degrees of $G$ using $A$.
     for $p = 1 : 10$
3    Randomly find two vertices $v_a$, $v_b$ such that $(v_a, v_b) \in E$, i.e., $A(v_a, v_b) = 1$.
4    Find $deg(v_a)$ and $deg(v_b)$.
5    Search randomly for two vertices $v_c$, $v_d$ and $deg(v_c) = deg(v_a) - 1$,
     $deg(v_d) = deg(v_b) - 1$.
        if $A(v_c, v_d) = 1$
          repeat the step 5
        else
          set $A(v_c, v_d) = 1$, $A(v_a, v_b) = 0$
     end
```

By applying the algorithm (just mentioned above) to the giant connected component of protein domain co-occurrence network, we get four randomly modified networks having the same connectivity distribution with the original network. Later, we analyze the attack tolerance of these four randomly modified networks and among these networks we extract the most vulnerable network and the least vulnerable network under attack. The most vulnerable network means that it is fragmented faster than the other networks and the least vulnerable network means that it is fragmented slower than the other networks as we remove nodes. We determine the most vulnerable and the least vulnerable networks by comparing the thresholds values of the networks. After extracting the most vulnerable and the least vulnerable networks, we again get four randomly modified networks from the most vulnerable and four randomly modified networks from the least vulnerable network by applying the algorithm. At this point we divide the analysis into two ways. In the first way, we attack the four networks obtained from the most vulnerable network and we collect the most vulnerable network from these networks. We repeat this process and continue collecting the most vulnerable network from the networks which are obtanied from the most vulnerable network. In the second way, we attack the four networks obtained from the least vulnerable network and we collect the least vulnerable network from these networks. We repeat this process and continue collecting the least vulnerable network from the networks which are obtanied from the least vulnerable network (see “the strategy of analysis” mentioned above). In Figures 4.4-8, changes in the relative size of the giant connected component S as function of the fraction of removed nodes f in the original network and first four networks in Tables 4.7 and 4.8 are simulated.

Results of this analysis are shown in Table 4.7 and 4.8. In these tables threshold values at which the network is fragmented are given. In Tables 4.7 and 4.8, the check sign “✓” which is nearby the networks means that the network is the most vulnerable network among other four networks in Table 4.7 and the least vulnerable network among other four networks in Table 4.8, and the star sign “$\star$” which is nearby the networks in Table 4.7 means that the network is fragmented later than the original network. In Table 4.7, the first analysis (Analysis1) in which four networks collected from the original network, network1_2 having the lowest threshold value is the most vulnerable network. We choose this network to make the Analysis 2 and continue like this. If we compare all threshold values in Table 4.7, all networks have different thresholds values or same threshold values. Also, only the smallest part of the (5% of the whole) networks has threshold values greater than the original network. In Table 4.8, the first analysis (Analysis1) in which four networks collected from the original network, Network1_1 having the highest threshold value is the least vulnerable network. We choose this network to make the Analysis 2 and continue like this. If we compare all threshold values in Table 4.8, all networks have different threshold values or same threshold values and all threshold values of the network after Analysis 2 are larger than the threshold value of the original network. But, in contrast to the results in Table 4.7, as we continue our analysis the threshold values of the networks increase. But, in Table 4.7, threshold values of the networks do not decrease but just become smaller than the threshold value of the original network as we continue the analysis.

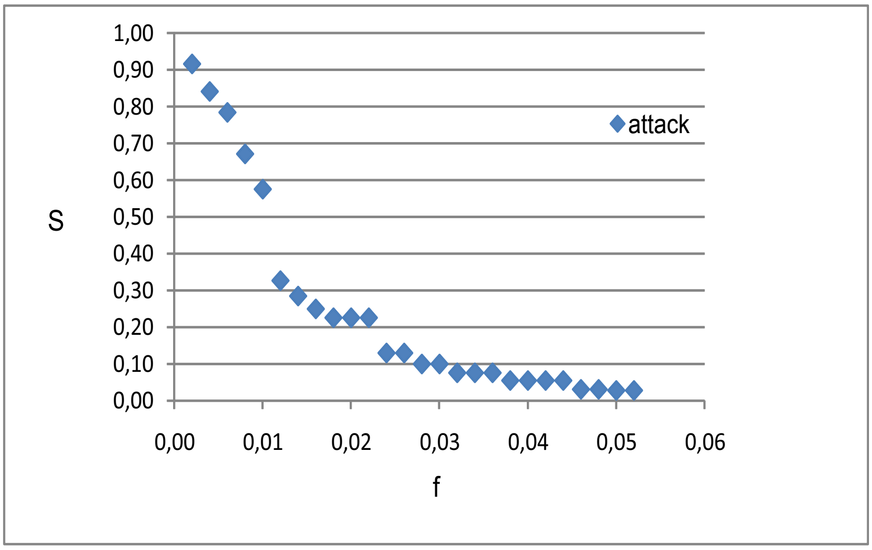
Results of this analysis indicate that the networks having the exactly the same connectivity distribution have different attack tolerances. From the most vulnerable network analysis, it can be observed that there is a selection which means that as we choose the worst network (the worst network means that threshold value of this network is less than the original network) , the probability that the network obtained from the worst network to become the worst one is high. This observation is derived from the result in the most vulnerable network analysis: only smallest part of the (5% of the whole) networks have threshold values greater than the original network. But from the least vulnerable network analysis, we can observe a pattern such that as we select the least vulnerable network from the least vulnerable network, threshold value of the networks (i.e., attack tolerance of the network) increases. This result and comparing two analyses (the least vulnerable and the most vulnerable network analyses) indicate that as we organize the same resources in a better way we can get a network with a higher atack tolerance. Additionally, all these results show that there is an evolutionary mechanism for having networks with higher attack tolerance for the same connectivity distribution.

Table 4.7: The results of the most vulnerable network analysis.

	Networks	Threshold values (f_c)
	The Original Network	0,046
Analysis1	Network1_1	0,046
✓	Network1_2	0,036
	Network1_3	0,038
	Network1_4	0,036
Analysis2 ✓	Network2_1	0,036
	Network2_2	0,040
	Network2_3	0,046
	Network2_4	0,042
Analysis3	Network3_1	0,044
	Network3_2	0,046
✓	Network3_3	0,036
	Network3_4	0,038
Analysis4	Network4_1	0,046
✓	Network4_2	0,046
	Network4_3	0,046
	Network4_4	0,046
Analysis5	Network5_1	0,040
	Network5_2	0,046
	Network5_3	0,040
✓	Network5_4	0,038
Analysis6 ✓	Network6_1	0,036
	Network6_2	0,044
	Network6_3	0,038
	Network6_4	0,046
Analysis7 ✓	Network7_1	0,038
	Network7_2	0,046
	Network7_3	0,040
	network7_4	0,046
Analysis8 ✓	Network8_1	0,038
	Network8_2	0,038
	Network8_3	0,046
	Network8_4	0,046
Analysis9 ⋆	Network9_1	0,058
	Network9_2	0,046
✓	Network9_3	0,046
⋆	Network9_4	0,060
Analysis10 ⋆	Network10_1	0,058
⋆	Network10_2	0,060
	Network10_3	0,046
⋆	Network10_4	0,052

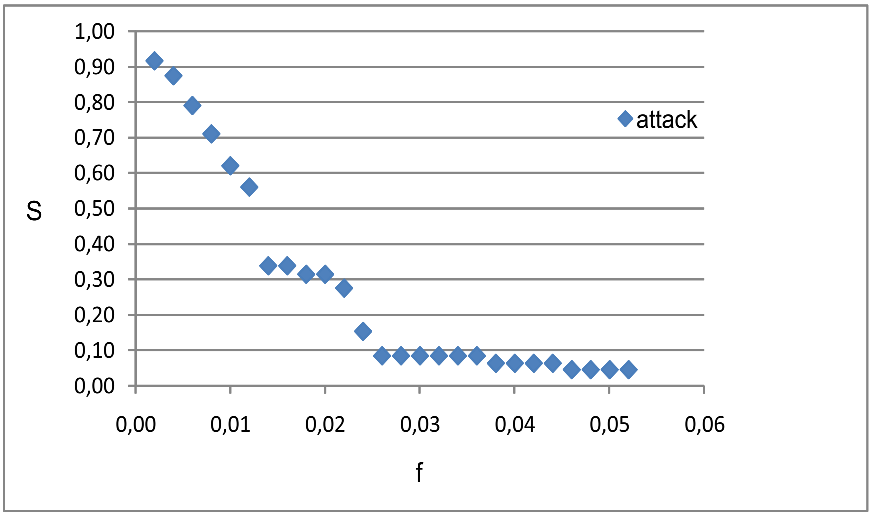
Table 4.8: The results of the least vulnerable network analysis.

	Networks	Threshold values (f_c)
	The Original Network	0,046
Analysis1 ✓	Network1_1	0,046
	Network1_2	0,036
	Network1_3	0,038
	Network1_4	0,036
Analysis2	Network2_1	0,036
✓	Network2_2	0,046
	Network2_3	0,046
	Network2_4	0,030
Analysis3 ✓	Network3_1	0,058
	Network3_2	0,046
	Network3_3	0,048
	Network3_4	0,046
Analysis4	Network4_1	0,058
	Network4_2	0,058
✓	Network4_3	0,066
	Network4_4	0,058
Analysis5	Network5_1	0,064
	Network5_2	0,046
✓	Network5_3	0,064
	Network5_4	0,048
Analysis6	Network6_1	0,058
	Network6_2	0,058
	Network6_3	0,062
✓	Network6_4	0,064
Analysis7 ✓	Network7_1	0,066
	Network7_2	0,058
	Network7_3	0,066
	network7_4	0,066
Analysis8	Network8_1	0,046
	Network8_2	0,066
✓	Network8_3	0,066
	Network8_4	0,066
Analysis9	Network9_1	0,070
	Network9_2	0,060
✓	Network9_3	0,070
	Network9_4	0,068
Analysis10	Network10_1	0,066
	Network10_2	0,068
	Network10_3	0,080
	Network10_4	0,068



The Original Network

Figure 4.4: Changes in the relative size of the giant connected component S as function of the fraction of removed nodes f in the original network.



Network1_1

Figure 4.5: Changes in the relative size of the giant connected component S as function of the fraction of removed nodes f in Network1_1.

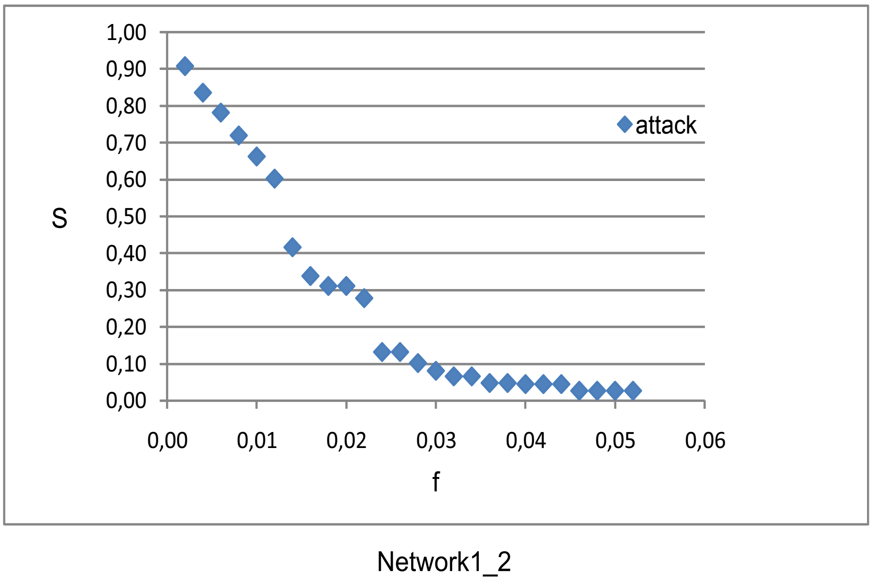


Figure 4.6: Changes in the relative size of the giant connected component S as function of the fraction of removed nodes f in Network1_2.

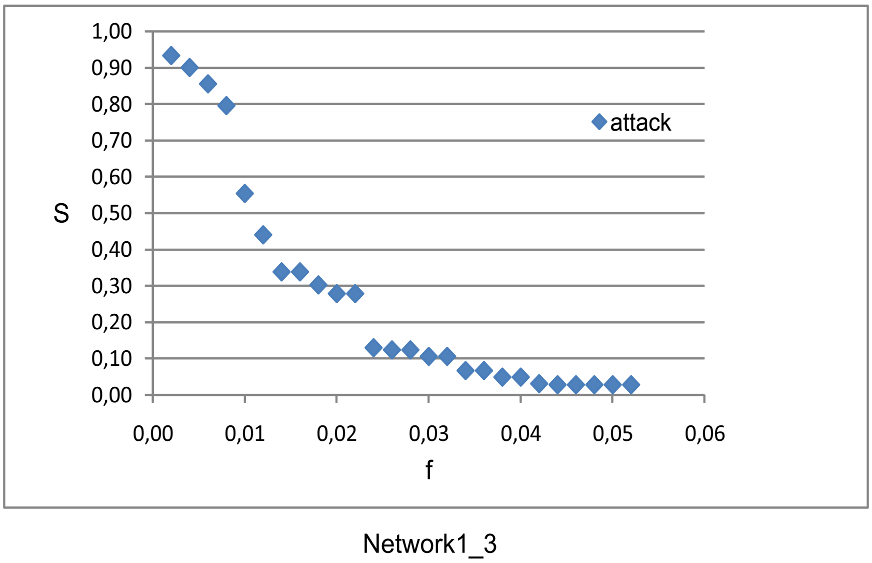
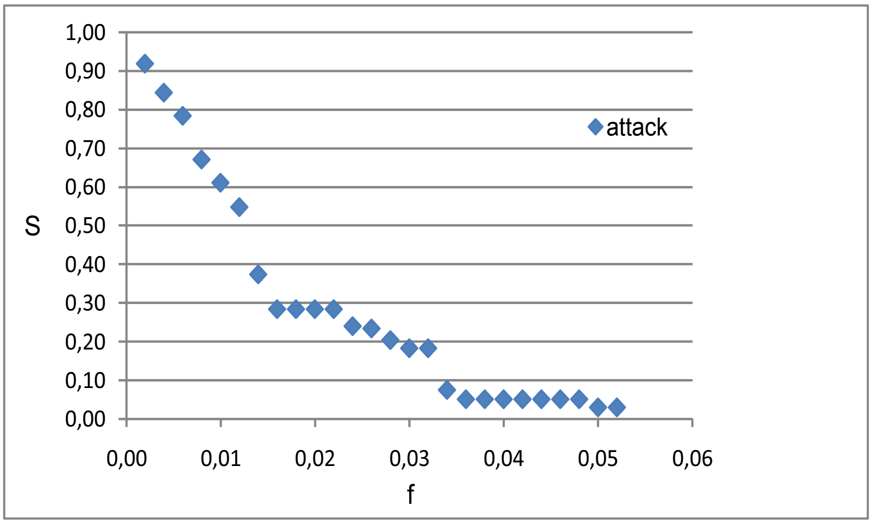


Figure 4.7: Changes in the relative size of the giant connected component S as function of the fraction of removed nodes f in Network1_3.



Network1_4

Figure 4.8: Changes in the relative size of the giant connected component S as function of the fraction of removed nodes f in Network1_4.

CHAPTER 5

CONCLUSION

The stability or survivability of some complex networks under different circumstances has received a growing interest among scientists. The study of the network robustness is particularly important by several occasions. In one hand a higher degree of robustness to errors and attacks may be desired for maintaining the information flow in communication networks under attacks. On the other hand, planning a very limited attack aimed at fragmenting a network by removal of minimum number of the most important nodes might have significant usage in drug design. In this thesis, we studied protein domain co-occurrence network of yeast generated with data from Biomart which links to Pfam database. Several networks obtained from protein domain co-occurrence network having exactly the same connectivity distribution were compared under attacks. In this work, we investigated the assumption that the different networks with the same connectivity distribution do not need to have the same attack tolerances. In addition to this, we considered that the networks with same connectivity distribution have higher attack tolerance as we organize the same resources in a better way. Then, we checked for the variations of attack tolerance of the networks with the same connectiviy distributions. Furthermore, we investigated whether there is an evolutionary mechanism for having networks with higher or lower attack tolerances for the same connectivity distribution.

Firstly, we checked whether protein domain co-occurrence network of yeast displays scale free topology or not. We found that our network are scale free network (displaying power law degree distribution). Investigation of the scale free topology is particularly important, since many real world networks are scale free networks. Scale free networks display unexpected degree of robustness, i.e., the ability of their nodes to communicate being unaffected by even high failure rates. However these networks are extremely vulnerable to intentional attacks.

Then, since we want to compare the attack tolerance of the networks, we determined the attack strategy: remove nodes by decreasing order of their degrees and the method: size of the giant connected component. To analyze the robustness of the network to attacks by studying how the size of the largest connected component varies as a function of the number of removed nodes, first we had to determine the size of the giant connected component and extract the giant connected component from the network. Extraction of the giant connected component of the network is particularly important for us, since we made all analysis on this component.

To check our assumption and investigations we wrote an algorithm. We randomly changed the links of nodes in the giant connected component of protein domain co-occurrence network of yeast while conserving the connectivity of the network. In doing so, we obtained several randomly modified networks which have the same connectivity distribution with the original network. Then we attacked to those networks and collected the most vulnerable network (the most vulnerable network means that it is fragmented faster than the other networks) and the least vulnerable network (the least vulnerable network means that it is fragmented slower than the other networks) as we remove nodes. We determined the most vulnerable network and the least vulnerable network by comparing the threshold values of the networks. We applied the algorithm on the most vulnerable network and the least vulnerable network to obtain several randomly modified networks. We again attacked to those networks. Then, we continued to collect the most vulnerable network from the networks which also obtained from the most vulnerable network and the least vulnerable network from the networks which also obtained from the least vulnerable network.

As a result, the networks having the exactly the same connectivity distribution have different attack tolerance under attacks. In addition to this, from the most vulnerable network analysis, we observed that there is a selection which means that as we choose the worst network (the worst network means that thresholds value of this network is less than the original network), the probability that the network obtaining from the worst network to become the worst one is high. This observation is derived from the result in the analysis: only the smallest part of the (5% of the whole) networks has threshold values greater than the original network. But, from the least vulnerable network analysis, we observed a pattern such that as we select the least vulnerable network from the networks which also obtained from the least vulnerable network, threshold values of the networks (i.e., attack tolerance of the network) incerease. This result and comparing two analyses (the least vulnerable and the most vulnerable network analysis) indicate that as we organize the same resources in a better way we can get a network with an higher attack tolerance. Also, all these results show that there is an evolutionary mechanism for having networks with higher attack tolerance for the same connectivity distribution.

The most important observations from this work is that there is a pattern such that as the network with higher attack tolerance is selected from the network which shows higher attack tolerance, attack tolerances of the networks increase. This observation indicates that an evolutinary mechanisim for having networks with higher attack tolerance for the same connectivity distribution can be constructed. From these observations, a guestion comes into mind; given the connectivity of the network, how the network is organized in the best way to show high attack tolerance under attacks. For this purpose, a method to organize the network in a better way can be developed. In addition to this, another question can be asked; without making an attack tolerance analysis and just only looking at the structure (structure may mean that how the links of the nodes in the network is organized) of the networks with the same connectivity distribution, can it be observed how the network behaves under attacks? Is it possible to develop an algorithm for this purpose. All these questions will be investigated as a future work.

REFERENCES

[1] Albert R., Jeong H., Barabasi A.-L., *Attack and error tolerance of complex networks*, Nature, 406: 378-382 (2000).

[2] Guillaume J.-L., Latapy M., Magnien C., *Comparison of Failures and Attacks on Random and Scale-Free Networks*, OPODIS: 186-196 (2004).

[3] Crucitti P., Latora V., Marchiori M., Rapisarda A., *Efficiency of scale-free networks:error and attack tolerance*, Physica A, 320: 622-642 (2003).

[4] Crucitti P., Latora V., Marchiori M., Rapisarda A., *Error and attacktolerance of complex networks*, Physica A, 340: 388-394 (2004).

[5] Albert R., Jeong H., Barabasi A.-L., *Diameter of the World-Wide Web*, Nature, 401: 130-131 (1999).

[6] Milgram S., *The small-world problem*, Psychology Today 2: 60-67 (1967).

[7] Barabasi A.L., Bonabeau E., *Scale-free networks*, Scientific American, 228: 60-69 (2003).

[8] Callaway D.S., Newmann M.E.J., Strogatz S.H., Watts D.J., *Network Robustness and Fragility: Percolation on Random Graphs*, Phys. Rev. Lett. 85: 5468-5471 (2000).

[9] Cohen R., Erez K., Ben-Avraham D., Havlin S., *Resilience of the Internet to Random Breakdowns*, Phys. Rev. Lett. 85: 4626-4628 (2000).

[10] Cohen R., Erez K., Ben-Avraham D., Havlin S., *Breakdown of the Internet under Intentional Attack*, Phys. Rev. Lett. 86: 3682-3685 (2001).

[11] Erdos P. and Renyi A., *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci. 5: 17-61 (1960).

[12] Watts D. J., Strogatz S. H., *Collective dynamics of small-world networks*. Nature 393: 440-442 (1998).

[13] Faloutsos M., Faloutsos P., Faloutsos C., *On power-law relationships of the Internet topology*, in SIGCOMM, 251-262. (1999).

[14] Barabasi A.-L., Albert R., *Emergance of scaling in random networks*, Science 286: 509-512 (1999).

[15] Barabasi A.-L., Albert R., Jeong H., *Mean-feld theory for scale-free random networks*, Physica A 272: 173-187 (1999).

[16] Dorogovtsev S., Mendes J., *Evolution of networks*, Adv. Phys. 51: 1079-1187 (2002).

[17] Barabasi A.-L., Albert R., *Statistical mechanics of complex networks*, Rev. Mod. Phys. 74: 47-98 (2002).

[18] Wuchty S., Ravaszy E., Barabasi A.-L., *The Architecture of Biological Networks*in T. S. Deisboeck, J. Yasha Kresh, and T. B. Kepler (eds.) Complex Systems in Biomedicine, Kluwer Academic Publishing, New York, (2003).

[19] Alex K. S. Ng, Janet Efstathiou, *Structural Robustness of Complex networks*

[20] Wuchty S., *Scale-free behavior in protein domain networks*. Mol Biol Evol., 18: 1694-1702 (2001).

[21] Wuchty S., *Interaction and domain networks of yeast* Proteomics, 2: 1715-1723 (2002).

[22] Jeong H., Tombor B., Albert R., Oltvai Z. N., Barabasi A.-L., *The large-scale organization of metabolic networks*, Nature 407: 651-654 (2000).

[23] Nacher J.C., Hayashida M., Akutsu T., *Protein Domain Networks: Scale-free. Mixing of Positive and Negative Exponents*, Physica A 367: 538-552 (2006).

[24] Newman M.E.J., *The structure of scientific collaboration networks*, Proceedings of the National Academy of Sciences of the USA, 98: 404-409 (2001).

[25] Wuchty S. Almaas E. *Evolutionary cores of domain co-occurrence networks*, BMC evolutionary biology, 5(1): 24 (2005).

[26] Barkai N., Leibler S., *Robustness in simple biochemical networks* Nature, 387: 913-917 (1997).

[27] Agnarsson G., Greenlaw R., *Graph theory: Modeling, Applications and Algorithms*, Pearson Prentice Hall, (2007).

[28] Dorogovtsev S.N., Mendes J.F.F., *Evolution of Networks From Biological Nets to the Internet and WWW*, Oxford University Press, (2003).

[29] Nooy W., Mrvar A., Batagelj V., *Exploratory Social Network Analysis with Pajek*, Cambridge University Press, (2005).

[30] Newman M. E. J., *The structure and function of complex networks*, SIAM Review 45: 167-256 (2003).

[31] Li-Feng G., Jian-Jun S., Shan G., *Evolution of a Protein Domain Interaction Network*, Chin. Phys. B Vol. 19, No. 1 010512 (2010).

[32] Khanin R., Wit E., *How Scale-Free Are Biological Networks*, Journal of Computational Biology, 13(3): 810-818 (2006).

[33] Barabasi A.L., Oltvai Z.N., *Network Biology: Understanding the Cell's Functional Organization*, Nature Reviews Genetics 5: 101-113 (2004).

[34] Albert R., *Scale-free Networks in Cell Biology* Journal of Cell Science, 118(21): 4947-4957 (2005).

[35] Wolf Y.I., Karev G., Koonin E.V., *Scale-Free Networks In Biology: New Insights Into The Fundamentals Of Evolution* Bioessays 24: 105-109 (2002).

[36] Qu Z., Wang P., Qin Z., *Enhancing the Scale-Free Network's Attack Tolerance*, Complex (2) 1823-1826 (2009).

[37] Sun S., Liu Z., Chen Z., Yuan Z., *Error and attack tolerance of evolving networks with local preferential attachment*, Physica A, 373: 851-860 (2007).

[38] Bornholdt S., Schuster H.G., *Handbook of Graphs and Networks From the Genome to the Internet*, Wiley-VCH, (2003).

[39] Vogel C., Bashton M., Kerrison N.D., Chothia C., Teichmann A.S., *Structure, function and evolution of multidomain proteins*, Current Opinion in Structural Biology, 14:208-216 (2004).

[40] Caldarelli G., Vespignani A., *Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science*, World Scientific Publishing, (2007).

APPENDIX A

PFAM PROTEIN DOMAIN CO-OCCURRENCE NETWORK OF YEAST (S. CEREVISIAE)

Table A.1: Pfam protein domain co-occurrence network of yeast obtanied from Biomart which links to Pfam database.

domain1	domain2	occurrence in proteins
PF06747	PF08583	1
PF01288	PF00809	1
PF00293	PF05026	1
PF00293	PF09297	1
PF01118	PF02774	2
PF01119	PF08676	2
PF09261	PF07748	1
PF01053	PF01212	1
PF05436	PF04648	2
PF08022	PF08030	8
PF00033	PF00032	1
PF01734	PF00027	1
PF01237	PF00023	2
PF00730	PF00633	1
PF00730	PF07934	1
PF01233	PF02799	1
PF01232	PF08125	1
PF03127	PF02883	2
PF00636	PF00035	2
PF00637	PF01394	1
PF08501	PF01488	1
PF04898	PF01493	1
PF04898	PF01645	1
PF08509	PF00211	1
PF00456	PF02780	2
PF00456	PF02779	2
PF05222	PF01262	1
PF04893	PF03878	1
PF01602	PF07718	1
PF01602	PF02883	1
PF01602	PF08752	1
PF05221	PF00670	1
PF01425	PF02626	1
PF01425	PF08443	1
PF01426	PF00249	1
PF01426	PF00439	2
PF01426	PF00628	1
PF04068	PF04034	1
PF04068	PF00037	1
PF01422	PF01424	1
PF04063	PF04064	1
PF01546	PF07687	3
PF00493	PF01078	2
PF00498	PF07714	1

PF04096	PF02415	1
PF00125	PF00808	3
PF00122	PF00702	12
PF00122	PF00689	5
PF00122	PF00403	1
PF02637	PF01162	1
PF02637	PF02934	1
PF02201	PF01253	1
PF01975	PF03133	1
PF00081	PF02777	1
PF00128	PF02922	1
PF00128	PF02806	1
PF00082	PF05922	3
PF00082	PF01483	1
PF01300	PF03481	1
PF00646	PF02809	1
PF00646	PF00560	2
PF00642	PF01207	1
PF03234	PF08564	1
PF03234	PF08565	1
PF00063	PF00612	2
PF00063	PF06017	2
PF00063	PF01843	2
PF00307	PF00612	1
PF00307	PF03836	1
PF00307	PF00616	1
PF00307	PF03271	1
PF02845	PF02204	1
PF02844	PF02843	1
PF00069	PF00536	1
PF00069	PF00168	2
PF00069	PF00169	2
PF00069	PF08171	1
PF00069	PF00072	1
PF00069	PF02985	1
PF00069	PF02149	2
PF00069	PF00433	8
PF00069	PF08311	1
PF00069	PF06479	1
PF00069	PF00400	1
PF00069	PF07647	1
PF00069	PF07714	20
PF00069	PF00659	1
PF00069	PF00130	1
PF00069	PF05773	1
PF00069	PF00786	3

PF00069	PF01163	1
PF00069	PF08587	1
PF00069	PF00498	3
PF00069	PF02185	1
PF08158	PF05285	1
PF00999	PF08619	1
PF08659	PF01575	1
PF04153	PF04065	2
PF04997	PF04983	3
PF04997	PF05000	3
PF04997	PF05001	1
PF04997	PF00623	3
PF04997	PF04998	3
PF04997	PF04992	1
PF04997	PF04990	1
PF00013	PF00098	1
PF03446	PF00393	2
PF02934	PF01162	1
PF00390	PF03949	1
PF00018	PF00611	2
PF00018	PF00063	2
PF00018	PF07653	20
PF00018	PF03983	1
PF00018	PF07647	2
PF00018	PF03114	1
PF00018	PF06017	2
PF00018	PF00790	1
PF00018	PF00564	1
PF00018	PF02809	1
PF00018	PF04366	2
PF00018	PF00241	1
PF00018	PF08226	1
PF03447	PF00742	1
PF02292	PF00023	2
PF00549	PF08442	1
PF00397	PF01846	2
PF00397	PF00639	1
PF00501	PF01370	1
PF00501	PF07993	1
PF00501	PF01073	1
PF00501	PF06464	1
PF00501	PF00550	1
PF04841	PF04840	1
PF04969	PF05002	1
PF04969	PF04925	1
PF00904	PF04855	1

PF05000	PF04992	1
PF05000	PF04998	3
PF05000	PF04990	1
PF05000	PF05001	1
PF01012	PF00766	1
PF02749	PF01729	1
PF00266	PF01212	1
PF00266	PF00282	1
PF01599	PF01020	1
PF00916	PF01740	4
PF00916	PF00027	1
PF05188	PF00488	5
PF05188	PF05190	4
PF05188	PF05192	5
PF03161	PF00033	1
PF07731	PF07732	3
PF07731	PF00394	3
PF01645	PF01493	1
PF01645	PF01070	1
PF04998	PF04992	1
PF04998	PF04990	1
PF04998	PF05001	1
PF01087	PF02744	1
PF04992	PF04990	1
PF04992	PF05001	1
PF01163	PF09202	1
PF04990	PF05001	1
PF09091	PF03184	1
PF01740	PF00027	1
PF00690	PF00702	7
PF00690	PF00689	5
PF00690	PF00122	7
PF02353	PF08498	1
PF02353	PF01170	1
PF08271	PF00382	2
PF08271	PF07741	1
PF00696	PF04768	1
PF00696	PF01472	2
PF00696	PF01842	1
PF00696	PF02774	1
PF00696	PF01118	1
PF02366	PF02815	6
PF00155	PF01053	1
PF00595	PF00089	1
PF09070	PF08324	1
PF01938	PF05958	1

PF08071	PF01479	1
PF08071	PF00900	1
PF00168	PF00387	1
PF00168	PF00616	1
PF00168	PF00036	1
PF00168	PF00388	1
PF00168	PF00433	2
PF00168	PF00130	1
PF00168	PF02185	1
PF00168	PF00397	1
PF00168	PF02666	1
PF00169	PF08174	1
PF00169	PF01237	3
PF00169	PF00023	2
PF00169	PF00201	1
PF00169	PF00617	1
PF00169	PF00618	1
PF00169	PF07653	2
PF00169	PF03778	1
PF00169	PF00620	2
PF00169	PF07647	2
PF00169	PF00780	1
PF00169	PF07714	2
PF00169	PF00787	1
PF00169	PF00786	2
PF00169	PF00018	2
PF00169	PF02893	1
PF00169	PF03033	1
PF03765	PF00650	6
PF08477	PF00071	32
PF08477	PF08355	1
PF08477	PF08356	1
PF08477	PF00025	6
PF08477	PF04950	1
PF08477	PF01926	1
PF08477	PF00036	1
PF08477	PF08142	1
PF01266	PF05187	1
PF01266	PF01946	2
PF01266	PF00890	3
PF01266	PF02910	2
PF04563	PF06883	1
PF04563	PF04566	2
PF04563	PF04560	3
PF04563	PF04561	3
PF04563	PF00562	3

PF04563	PF04567	3
PF04563	PF04565	3
PF04560	PF06883	1
PF04561	PF06883	1
PF04561	PF04566	2
PF04561	PF04560	3
PF04561	PF00562	3
PF04561	PF04567	3
PF04561	PF04565	3
PF04566	PF04560	2
PF04566	PF00562	2
PF04566	PF04567	2
PF04567	PF06883	1
PF04567	PF04560	3
PF04567	PF00562	3
PF04564	PF08783	1
PF00763	PF02882	3
PF00763	PF01268	2
PF04037	PF04046	1
PF02969	PF07571	1
PF00023	PF01529	2
PF00023	PF03009	1
PF00023	PF01833	2
PF00023	PF00651	1
PF00027	PF02197	1
PF04428	PF01633	2
PF04425	PF04426	2
PF03462	PF00472	1
PF03463	PF03464	1
PF03463	PF03465	1
PF03460	PF01077	1
PF03464	PF03465	2
PF07646	PF01344	6
PF07647	PF00536	1
PF07728	PF02359	1
PF07728	PF02861	1
PF07728	PF09336	1
PF07728	PF05362	1
PF07728	PF04212	1
PF07728	PF02190	1
PF07728	PF03028	1
PF07728	PF00493	3
PF07728	PF01078	1
PF07728	PF07724	3
PF07728	PF07726	1
PF07728	PF00439	1

PF00118	PF01504	1
PF00117	PF00185	1
PF00117	PF00958	1
PF00117	PF02787	1
PF00117	PF00218	1
PF00117	PF00425	1
PF00117	PF04715	1
PF00117	PF02729	1
PF00117	PF06418	2
PF00117	PF00988	2
PF07724	PF02861	1
PF07724	PF07726	1
PF00115	PF03161	1
PF00115	PF00961	1
PF00115	PF00078	2
PF00115	PF01348	2
PF00333	PF03719	2
PF00330	PF00694	4
PF00339	PF02752	3
PF04928	PF01909	1
PF04928	PF04926	1
PF02893	PF00566	1
PF02893	PF00201	1
PF02893	PF03033	1
PF04677	PF04676	1
PF05739	PF00787	1
PF02784	PF00278	1
PF02785	PF02436	2
PF02785	PF01425	1
PF02785	PF02655	2
PF02785	PF02626	1
PF02785	PF00682	2
PF02785	PF01039	2
PF02785	PF00364	5
PF02785	PF02222	2
PF02785	PF08443	2
PF02785	PF08326	2
PF02785	PF01071	2
PF02786	PF02436	2
PF02786	PF01425	1
PF02786	PF02655	2
PF02786	PF02626	1
PF02786	PF00185	1
PF02786	PF02785	5
PF02786	PF01039	2
PF02786	PF02787	2

PF02786	PF00364	5
PF02786	PF02222	3
PF02786	PF00988	1
PF02786	PF01071	2
PF02786	PF00682	2
PF02786	PF00117	1
PF02786	PF08326	2
PF02786	PF02142	1
PF02786	PF02729	1
PF02786	PF08443	3
PF02787	PF00185	1
PF02787	PF02729	1
PF02801	PF01648	1
PF03946	PF00298	2
PF02148	PF07576	1
PF00433	PF00130	1
PF00433	PF02185	1
PF02146	PF04574	2
PF02809	PF00904	1
PF02809	PF00790	2
PF00349	PF03727	4
PF02142	PF00185	1
PF02142	PF02787	1
PF02142	PF00988	1
PF02142	PF02729	1
PF02142	PF00117	1
PF02142	PF01808	2
PF00439	PF08880	1
PF00439	PF02178	1
PF04815	PF00626	4
PF04810	PF00626	4
PF04810	PF04815	4
PF04810	PF08033	4
PF04810	PF04811	4
PF04811	PF00626	4
PF04811	PF04815	4
PF04811	PF08033	4
PF00804	PF05739	5
PF03129	PF02824	1
PF03129	PF09180	1
PF03129	PF07973	1
PF05131	PF00637	1
PF00632	PF00168	1
PF00632	PF00397	1
PF02733	PF02734	2
PF02735	PF03730	2

PF02736	PF00612	2
PF02736	PF00063	2
PF02736	PF01843	2
PF04565	PF00562	3
PF04565	PF04560	3
PF04565	PF04566	2
PF04565	PF04567	3
PF04565	PF06883	1
PF08032	PF00588	1
PF08033	PF00626	4
PF08033	PF04815	4
PF08407	PF01644	2
PF08407	PF03142	2
PF00888	PF08672	1
PF00181	PF03947	2
PF00890	PF02910	2
PF00899	PF05237	1
PF00899	PF02134	3
PF04321	PF01263	1
PF04321	PF02719	2
PF08534	PF00578	5
PF01634	PF08029	1
PF01411	PF02272	1
PF01411	PF07973	1
PF08242	PF08498	1
PF08242	PF01209	1
PF08242	PF03291	1
PF08242	PF02353	1
PF08242	PF01170	1
PF08241	PF08242	12
PF08241	PF01170	1
PF08241	PF05148	1
PF08241	PF02353	1
PF08241	PF01209	1
PF08241	PF08498	1
PF08241	PF03291	1
PF08240	PF00107	16
PF08711	PF07500	1
PF08506	PF03378	1
PF01417	PF00904	1
PF01417	PF02809	2
PF01417	PF01608	1
PF08712	PF01106	1
PF04072	PF07646	1
PF04072	PF01344	1
PF08083	PF08084	1

PF08082	PF08084	1
PF08082	PF08083	1
PF02194	PF08628	1
PF08326	PF01039	2
PF02190	PF05362	1
PF00153	PF00036	1
PF01909	PF03828	2
PF01909	PF04926	1
PF00156	PF00310	1
PF02359	PF02933	1
PF03731	PF02735	2
PF03731	PF03730	2
PF01902	PF01042	1
PF00091	PF03953	4
PF00096	PF01363	1
PF00096	PF02373	2
PF00096	PF00226	1
PF00096	PF02375	2
PF00096	PF02178	1
PF00097	PF04757	1
PF00097	PF01485	1
PF00097	PF02037	1
PF00097	PF00176	3
PF00097	PF07576	1
PF00097	PF08647	1
PF00097	PF00271	3
PF00097	PF02148	1
PF00097	PF01363	1
PF00097	PF08797	1
PF00097	PF00642	1
PF00097	PF00498	1
PF00097	PF00628	1
PF00795	PF02540	1
PF04557	PF03950	1
PF04557	PF00749	1
PF00790	PF03127	2
PF00790	PF02883	2
PF00792	PF00613	1
PF00792	PF00454	1
PF03636	PF03633	1
PF03636	PF03632	1
PF04558	PF03950	1
PF04558	PF04557	1
PF04558	PF00749	1
PF03632	PF03633	1

PF07973	PF02824	1
PF07973	PF02272	1
PF00076	PF00806	2
PF00076	PF08662	1
PF00076	PF00658	1
PF00076	PF00641	1
PF00076	PF05391	1
PF00070	PF00310	1
PF00070	PF01645	1
PF00070	PF03486	1
PF00070	PF01134	1
PF00070	PF02852	3
PF00070	PF04898	1
PF00070	PF01493	1
PF00071	PF08355	1
PF00071	PF01926	1
PF00071	PF08356	1
PF00071	PF00036	1
PF00071	PF00025	6
PF00072	PF00512	1
PF00072	PF01163	1
PF00072	PF00447	1
PF00072	PF02518	1
PF04675	PF04679	1
PF02854	PF02847	2
PF02854	PF09088	1
PF02854	PF09090	1
PF03917	PF03199	1
PF00370	PF02782	3
PF00078	PF00098	2
PF00078	PF09337	2
PF00078	PF01348	2
PF02852	PF01134	1
PF02581	PF02110	1
PF05773	PF01205	1
PF05773	PF04408	1
PF05773	PF07717	1
PF00988	PF00185	1
PF00988	PF02787	1
PF00988	PF02729	1
PF01794	PF08030	8
PF01794	PF08022	8
PF08669	PF01571	1
PF08142	PF04950	2
PF02518	PF00204	1

PF02518	PF00183	2
PF02518	PF01119	3
PF02518	PF00521	1
PF02518	PF08676	1
PF02518	PF00512	1
PF00982	PF02358	3
PF08390	PF03798	2
PF08393	PF03028	1
PF08393	PF07728	1
PF00382	PF00134	1
PF00382	PF07741	1
PF00383	PF00849	1
PF00479	PF02781	1
PF00388	PF00387	1
PF00389	PF02826	5
PF00571	PF00654	1
PF00571	PF01595	1
PF00570	PF08066	1
PF00570	PF01612	1
PF02922	PF02806	1
PF00575	PF07541	1
PF00575	PF04408	1
PF00575	PF07717	1
PF03983	PF08226	1
PF00085	PF00462	2
PF03986	PF03987	1
PF00205	PF02775	7
PF00204	PF00521	1
PF07529	PF00176	1
PF00208	PF02812	2
PF04851	PF00176	5
PF04851	PF02889	2
PF04851	PF07529	1
PF04851	PF00270	5
PF04851	PF00271	10
PF04851	PF09110	2
PF04851	PF09111	2
PF00930	PF00326	2
PF02776	PF02775	7
PF02776	PF00205	7
PF02770	PF01756	1
PF02772	PF02773	2
PF00939	PF03600	3
PF02779	PF00676	1
PF02779	PF02780	3

PF00270	PF00627	1
PF00270	PF04408	3
PF00270	PF08148	2
PF00270	PF00570	1
PF00270	PF02889	3
PF00270	PF00271	51
PF00270	PF07717	3
PF00270	PF05773	1
PF00270	PF08147	1
PF00271	PF00385	1
PF00271	PF00627	1
PF00271	PF00176	17
PF00271	PF04408	7
PF00271	PF07529	1
PF00271	PF08880	1
PF00271	PF08148	2
PF00271	PF00570	1
PF00271	PF02889	3
PF00271	PF05773	1
PF00271	PF00575	1
PF00271	PF07717	7
PF00271	PF09110	2
PF00271	PF09111	2
PF00271	PF08797	1
PF00271	PF02178	1
PF00271	PF08658	1
PF00271	PF00439	2
PF00271	PF08147	1
PF00276	PF03939	1
PF00275	PF01487	1
PF00275	PF01202	1
PF00275	PF01488	1
PF00275	PF08501	1
PF07558	PF07557	1
PF02178	PF04084	1
PF02178	PF08880	1
PF05195	PF00557	2
PF05190	PF00488	4
PF05192	PF00488	6
PF05192	PF05190	4
PF05193	PF08367	1
PF08304	PF08302	1
PF08304	PF08303	1
PF00856	PF08236	1
PF03079	PF07883	1

PF01174	PF00117	1
PF01174	PF00977	1
PF01174	PF07685	3
PF03486	PF02852	1
PF01170	PF08498	1
PF04825	PF04824	1
PF07651	PF01608	1
PF07651	PF01417	3
PF09088	PF09090	1
PF06733	PF06777	1
PF00682	PF02436	2
PF00682	PF08502	2
PF04983	PF04992	1
PF04983	PF04998	3
PF04983	PF05001	1
PF04983	PF05000	3
PF04983	PF04990	1
PF01073	PF07993	7
PF01073	PF04321	2
PF01073	PF01263	1
PF01073	PF02719	2
PF01073	PF00550	1
PF07653	PF00611	1
PF07653	PF00063	2
PF07653	PF00241	1
PF07653	PF03983	1
PF07653	PF07647	2
PF07653	PF03114	1
PF07653	PF06017	2
PF07653	PF00790	1
PF07653	PF00564	1
PF07653	PF02809	1
PF07653	PF04366	2
PF07653	PF08226	1
PF01071	PF02436	2
PF01071	PF02769	1
PF01071	PF00682	2
PF01071	PF02844	1
PF01071	PF02843	1
PF01071	PF02222	2
PF01071	PF00586	1
PF01031	PF02212	2
PF01699	PF03733	1
PF01751	PF01131	1
PF08953	PF08954	1

PF00581	PF00899	1
PF00581	PF05237	1
PF00581	PF00102	1
PF00581	PF00443	3
PF02259	PF02260	3
PF02259	PF08771	2
PF00586	PF02769	2
PF00586	PF02843	1
PF00586	PF02844	1
PF00587	PF00152	1
PF00587	PF03129	6
PF00587	PF01336	1
PF00587	PF09180	1
PF00587	PF02824	1
PF00587	PF02403	1
PF00587	PF07973	1
PF01873	PF02020	1
PF08590	PF01713	1
PF00610	PF00611	1
PF00610	PF00780	2
PF00610	PF00621	2
PF00610	PF00620	1
PF00611	PF00620	2
PF00612	PF03836	1
PF00612	PF01843	2
PF00613	PF00454	2
PF00616	PF00612	1
PF00616	PF03836	1
PF00617	PF07653	1
PF00617	PF00018	1
PF00617	PF00620	1
PF00618	PF07653	1
PF00618	PF00620	1
PF00618	PF00018	1
PF00618	PF00617	3
PF01979	PF07969	2
PF05204	PF00306	1
PF05204	PF02874	1
PF05204	PF00006	1
PF05204	PF05203	2
PF08567	PF03909	1
PF04212	PF09336	1
PF08565	PF08564	1

PF08449	PF03151	4
PF05203	PF00306	1
PF05203	PF02874	1
PF05203	PF00006	1
PF01213	PF08603	1
PF00750	PF05746	2
PF00753	PF07521	1
PF00752	PF00867	4
PF08443	PF02436	1
PF08443	PF00185	1
PF08443	PF02655	1
PF08443	PF02626	1
PF08443	PF00682	1
PF08443	PF02787	1
PF08443	PF02222	1
PF08443	PF00988	1
PF08443	PF02729	1
PF08443	PF00117	1
PF08443	PF02142	1
PF08443	PF01071	1
PF08621	PF08620	1
PF02205	PF00568	1
PF03810	PF08389	1
PF03810	PF08767	1
PF03810	PF03378	1
PF03810	PF08506	1
PF03810	PF02985	2
PF00036	PF08355	1
PF00036	PF00387	1
PF00036	PF08356	1
PF00036	PF00388	1
PF02207	PF02617	1
PF04433	PF00249	3
PF04433	PF00569	2
PF00534	PF08288	1
PF02655	PF02436	2
PF02655	PF00682	2
PF02655	PF02222	2
PF02655	PF01071	2
PF02558	PF08546	2
PF00533	PF00041	1
PF00533	PF00249	1
PF00533	PF06732	1

PF00533	PF01068	1
PF00533	PF09197	1
PF00533	PF08519	1
PF00533	PF00817	1
PF02225	PF01546	1
PF02225	PF04253	2
PF02222	PF02436	2
PF02222	PF00682	2
PF02222	PF00731	1
PF02222	PF02787	1
PF00108	PF02803	2
PF00109	PF01648	1
PF00109	PF02801	2
PF07732	PF00394	3
PF00106	PF07993	1
PF00106	PF01263	1
PF00106	PF04321	1
PF00106	PF01575	1
PF00106	PF01370	1
PF00106	PF01073	1
PF00106	PF08659	6
PF00106	PF02719	1
PF02383	PF03372	3
PF01591	PF00300	3
PF00249	PF09197	1
PF00249	PF03990	1
PF00249	PF00569	2
PF00327	PF08079	2
PF00240	PF09280	1
PF00240	PF01020	2
PF00240	PF01599	2
PF00977	PF00117	1
PF00970	PF00175	5
PF00970	PF00042	1
PF02037	PF02891	2
PF03952	PF00113	4
PF00400	PF08625	1
PF00400	PF04003	1
PF00400	PF07687	1
PF00400	PF01546	1
PF00400	PF00097	1
PF00400	PF08149	1
PF00400	PF09070	1

PF00400	PF08581	1
PF00400	PF08145	1
PF00400	PF04192	1
PF00400	PF06957	1
PF00400	PF08513	2
PF00400	PF04053	2
PF00400	PF00646	2
PF00400	PF07569	1
PF00400	PF02985	1
PF00400	PF04158	1
PF00400	PF08953	1
PF00400	PF08954	1
PF00400	PF08324	1
PF00400	PF08154	1
PF00400	PF00637	1
PF00400	PF04047	1
PF00400	PF04494	1
PF03951	PF00120	1
PF00406	PF05191	2
PF09334	PF08264	4
PF09334	PF00133	4
PF09334	PF06827	1
PF02375	PF02373	3
PF09337	PF00098	2
PF04869	PF04871	1
PF00172	PF04082	26
PF00172	PF00989	1
PF00172	PF03902	1
PF03031	PF00533	1
PF03033	PF00201	1
PF01138	PF03725	5
PF01137	PF05189	1
PF02729	PF00185	2
PF05378	PF01968	1
PF05378	PF02538	1
PF05020	PF05021	1
PF05388	PF00450	1
PF04263	PF04265	1
PF04389	PF02225	2
PF04389	PF01546	1
PF01472	PF08068	1
PF01398	PF08084	1
PF01398	PF08083	1
PF01398	PF08082	1
PF02020	PF00483	1

PF01624	PF00488	4
PF01624	PF05190	3
PF01624	PF05192	4
PF01624	PF05188	4
PF03105	PF00023	2
PF03105	PF00939	3
PF03105	PF03124	1
PF03105	PF03600	3
PF03105	PF03009	1
PF03104	PF08996	1
PF03104	PF08490	1
PF03104	PF00136	4
PF06858	PF08155	1
PF08523	PF01381	1
PF01408	PF02894	1
PF01409	PF03147	1
PF02260	PF08064	1
PF02260	PF08771	2
PF06012	PF00632	1
PF06012	PF06025	1
PF08311	PF08171	2
PF02268	PF02751	1
PF02185	PF00130	1
PF00149	PF05011	1
PF00149	PF04152	1
PF00149	PF08321	1
PF01369	PF09324	1
PF01368	PF02833	1
PF00786	PF07714	3
PF00787	PF09325	2
PF00787	PF00620	1
PF00787	PF08628	1
PF00787	PF02194	1
PF00787	PF00018	1
PF00787	PF07653	1
PF00787	PF00564	1
PF03198	PF07983	2
PF01363	PF02809	1
PF01363	PF01504	1
PF01363	PF00790	1
PF01363	PF00118	1
PF00660	PF00399	2
PF00788	PF08509	1
PF00788	PF00211	1
PF01399	PF08375	1

PF01399	PF05470	1
PF00665	PF07727	46
PF00665	PF09337	2
PF00665	PF00078	2
PF00665	PF00098	2
PF01494	PF08491	1
PF00043	PF00647	2
PF00044	PF02800	3
PF02985	PF01851	1
PF02985	PF00454	3
PF02985	PF01749	1
PF02985	PF00176	1
PF02985	PF07539	1
PF02985	PF08752	1
PF02985	PF00005	2
PF02985	PF02260	3
PF02985	PF01602	3
PF02985	PF08771	2
PF02985	PF00271	1
PF02985	PF02259	3
PF02985	PF00514	2
PF00364	PF02436	2
PF00364	PF01425	1
PF00364	PF02655	2
PF00364	PF02626	1
PF00364	PF00682	2
PF00364	PF01039	2
PF00364	PF02222	2
PF00364	PF00198	2
PF00364	PF08443	2
PF00364	PF02817	2
PF00364	PF08326	2
PF00364	PF01071	2
PF04768	PF01118	1
PF04768	PF02774	1
PF05743	PF00179	1
PF05742	PF07723	1
PF07492	PF01204	2
PF00443	PF00917	1
PF00443	PF00627	1
PF00443	PF02148	2
PF08389	PF08767	1
PF00438	PF02773	2

PF00438	PF02772	2
PF08385	PF03028	1
PF08385	PF08393	1
PF08385	PF07728	1
PF00448	PF09201	1
PF00448	PF02978	1
PF02463	PF04423	1
PF02463	PF06470	4
PF01565	PF04030	1
PF01565	PF02913	3
PF02919	PF01028	1
PF04042	PF08418	1
PF00562	PF04560	3
PF00562	PF06883	1
PF00560	PF08509	1
PF00560	PF00788	1
PF00560	PF01302	1
PF00560	PF00211	1
PF00561	PF04083	1
PF00561	PF07819	2
PF07691	PF00624	4
PF00928	PF01217	1
PF02769	PF02844	1
PF02769	PF02843	1
PF00289	PF02436	2
PF00289	PF01425	1
PF00289	PF02655	2
PF00289	PF02626	1
PF00289	PF00185	1
PF00289	PF02785	5
PF00289	PF02786	7
PF00289	PF02787	2
PF00289	PF00364	5
PF00289	PF02222	3
PF00289	PF00988	1
PF00289	PF01071	2
PF00289	PF00682	2
PF00289	PF00117	1
PF00289	PF08326	2
PF00289	PF02729	1
PF00289	PF02142	1
PF00289	PF01039	2
PF00289	PF08443	3

PF00288	PF08544	6
PF00282	PF00464	1
PF00281	PF00673	3
PF01193	PF01000	2
PF03144	PF03764	4
PF03144	PF00679	6
PF03144	PF04760	1
PF03144	PF06421	1
PF03144	PF09173	1
PF03144	PF03143	4
PF03142	PF01644	2
PF03142	PF00173	1
PF00467	PF03439	1
PF00467	PF01479	1
PF00467	PF00900	1
PF00467	PF08071	1
PF00467	PF01287	1
PF01068	PF04679	1
PF01068	PF04675	1
PF01068	PF01331	1
PF01068	PF03919	1
PF01061	PF07974	1
PF01061	PF06422	8
PF02719	PF01263	1
PF01728	PF07780	1
PF08354	PF00698	1
PF08354	PF01575	1
PF08356	PF08355	1
PF08351	PF05127	1
PF08059	PF00789	1
PF09110	PF09111	2
PF00628	PF07500	1
PF00628	PF00249	1
PF00628	PF01388	1
PF00628	PF02373	2
PF00628	PF07744	1
PF00628	PF02375	1
PF00628	PF00856	2
PF00627	PF00077	1
PF00627	PF01412	1
PF00627	PF00240	2
PF00627	PF04408	1
PF00627	PF00036	1

PF00627	PF05773	1
PF00627	PF02148	1
PF00627	PF07717	1
PF00627	PF09280	1
PF00621	PF00780	2
PF00620	PF00412	3
PF00623	PF04983	3
PF00623	PF05000	3
PF00623	PF05001	1
PF00623	PF04998	3
PF00623	PF04992	1
PF00623	PF04990	1
PF00749	PF03950	2
PF08513	PF04494	1
PF08512	PF03531	1
PF03099	PF02237	1
PF04056	PF07975	1
PF01202	PF01487	1
PF01202	PF01488	1
PF01202	PF08501	1
PF08630	PF07535	1
PF01433	PF09127	1
PF07690	PF00854	1
PF07690	PF00083	36
PF01575	PF00698	1
PF04408	PF07717	7
PF08221	PF05645	1
PF08226	PF00036	1
PF01074	PF07748	1
PF01074	PF09261	1
PF00481	PF08509	1
PF00481	PF00788	1
PF00481	PF00211	1
PF00481	PF00560	1
PF03485	PF00750	2
PF03485	PF05746	2
PF02629	PF00549	1
PF00487	PF00173	1
PF03483	PF03484	1
PF07992	PF00310	1
PF07992	PF00890	2
PF07992	PF00070	10
PF07992	PF01645	1

PF07992	PF03486	1
PF07992	PF02852	3
PF07992	PF01134	1
PF07992	PF04898	1
PF07992	PF01266	2
PF07992	PF01493	1
PF07992	PF02910	1
PF04088	PF00018	1
PF04088	PF07653	1
PF01336	PF00152	6
PF01336	PF08784	1
PF01336	PF08646	1
PF07991	PF01450	1
PF01331	PF03919	1
PF07994	PF01658	1
PF01968	PF02538	1
PF00133	PF08264	3
PF00133	PF06827	1
PF00132	PF00483	2
PF00132	PF02020	1
PF00134	PF08613	3
PF00134	PF02984	6
PF00136	PF08996	1
PF00136	PF08490	1
PF00258	PF01077	1
PF00258	PF03460	1
PF00258	PF08608	1
PF00258	PF00667	2
PF00258	PF04055	1
PF00310	PF00733	2
PF00310	PF01380	1
PF00310	PF01493	1
PF00310	PF01645	1
PF00310	PF04898	1
PF03930	PF05202	1
PF03931	PF01466	1

PF00317	PF02867	2
PF00250	PF00498	3
PF00251	PF08244	5
PF07574	PF08746	1
PF00256	PF01305	1
PF08644	PF08512	1
PF04715	PF00425	2
PF02181	PF06367	1
PF07993	PF04321	2
PF07993	PF01263	1
PF07993	PF02719	2
PF07993	PF00550	1
PF06978	PF08170	1
PF00005	PF04068	1
PF00005	PF01061	10
PF00005	PF00385	1
PF00005	PF07974	1
PF00005	PF00037	1
PF00005	PF00664	10
PF00005	PF06422	8
PF00004	PF01426	1
PF00004	PF01434	3
PF00004	PF00533	1
PF00004	PF02359	2
PF00004	PF02861	1
PF00004	PF09336	4
PF00004	PF05362	1
PF00004	PF06068	1
PF00004	PF08542	3
PF00004	PF02190	1
PF00004	PF02933	1
PF00004	PF04212	1
PF00004	PF09262	1
PF00004	PF07728	10
PF00004	PF07724	2
PF00004	PF08519	1
PF00004	PF06480	2
PF00004	PF00439	1
PF00004	PF08740	1
PF00006	PF00306	4
PF00009	PF03764	4
PF00009	PF00679	6
PF00009	PF04760	1
PF00009	PF03144	13
PF00009	PF06421	1

PF00009	PF09173	1
PF00009	PF03143	4
PF00478	PF00571	4
PF08264	PF06827	1
PF06480	PF01434	2
PF01021	PF07727	44
PF01021	PF00665	44
PF04082	PF03902	1
PF00583	PF09337	1
PF00583	PF00439	1
PF00583	PF04055	1
PF06472	PF00005	2
PF09235	PF00788	1
PF06371	PF02181	2
PF06371	PF06367	1
PF04376	PF04377	1
PF08543	PF00294	1
PF08543	PF03070	3
PF06395	PF00621	1
PF01798	PF08156	2
PF00702	PF00403	1
PF00702	PF00689	5
PF00702	PF08282	1
PF00702	PF06888	1
PF00875	PF03441	1
PF00704	PF03427	1
PF00705	PF02747	1
PF01096	PF07500	1
PF01096	PF08711	1
PF01154	PF08540	1
PF01479	PF00163	3
PF01479	PF00900	1
PF08303	PF08302	1
PF08265	PF05764	1
PF07650	PF00189	1
PF08267	PF01717	1
PF00684	PF01556	4
PF01926	PF06071	2
PF01926	PF08438	1
PF01926	PF01018	1
PF01926	PF06858	1
PF01926	PF08701	1
PF01926	PF08153	1

PF01926	PF02824	2
PF01926	PF08155	1
PF06025	PF00632	1
PF08069	PF00312	1
PF08066	PF01612	1
PF08060	PF01798	3
PF08060	PF08156	2
PF01487	PF08501	1
PF01487	PF01488	1
PF01761	PF01487	1
PF01761	PF01202	1
PF01761	PF08501	1
PF01761	PF01488	1
PF01761	PF00275	1
PF00675	PF05193	6
PF00179	PF09288	1
PF01849	PF00627	1
PF00175	PF00258	2
PF00175	PF00042	1
PF00175	PF00667	3
PF02373	PF01388	1
PF00176	PF00385	1
PF00176	PF08880	1
PF00176	PF09110	2
PF00176	PF09111	2
PF00176	PF08797	1
PF00176	PF08658	1
PF00176	PF00439	2
PF00176	PF02178	1
PF01842	PF02826	2
PF01842	PF00389	2
PF00170	PF07716	4
PF00173	PF04116	1
PF00173	PF01645	1
PF00173	PF01070	1
PF00679	PF03764	4
PF00679	PF06421	1
PF01379	PF03900	1
PF00961	PF00033	2
PF00961	PF07453	1
PF01370	PF04321	2
PF01370	PF07993	7
PF01370	PF00550	1

PF01370	PF01073	8
PF01370	PF01263	1
PF01370	PF02719	2
PF04571	PF08235	1
PF03871	PF01191	1
PF01503	PF00815	1
PF01502	PF01503	1
PF01502	PF00815	1
PF03876	PF00575	1
PF03876	PF08292	1
PF00056	PF02866	3
PF01509	PF01472	1
PF01509	PF08068	1
PF00198	PF02817	1
PF08766	PF02201	2
PF02921	PF00355	1
PF08605	PF00533	1
PF03477	PF02867	2
PF03477	PF00317	2
PF08801	PF04044	1
PF00454	PF02260	5
PF00454	PF08064	1
PF00454	PF02259	3
PF00454	PF08771	2
PF04053	PF06957	1
PF07714	PF07647	1
PF07714	PF02149	1
PF07714	PF00536	1
PF07714	PF00659	1
PF07714	PF08587	1
PF04055	PF06968	1
PF04055	PF08608	1
PF07719	PF00149	1
PF07719	PF00160	2
PF07719	PF04049	1
PF07719	PF08321	1
PF03372	PF00560	1
PF03372	PF06839	1
PF02671	PF08295	1
PF00515	PF00149	1
PF00515	PF00160	2
PF00515	PF09145	1
PF00515	PF06424	1

PF00515	PF07719	18
PF00515	PF04049	1
PF00515	PF08321	1
PF00514	PF01749	1
PF07500	PF07744	1
PF00224	PF02887	2
PF02492	PF07683	1
PF00226	PF02889	1
PF00226	PF05207	1
PF00226	PF07743	1
PF00226	PF01556	5
PF00226	PF00684	5
PF01212	PF00282	1
PF02882	PF01268	2
PF02881	PF09201	1
PF02881	PF00448	2
PF02881	PF02978	1
PF02880	PF00408	2
PF02885	PF00591	1
PF01210	PF07479	2
PF08282	PF03332	1
PF08282	PF06888	1
PF02798	PF00647	2
PF02798	PF00043	5
PF02790	PF00116	1
PF02150	PF01096	3
PF02874	PF00306	4
PF02874	PF00006	4
PF02152	PF01288	1
PF02152	PF00809	1
PF00350	PF01031	2
PF00350	PF02212	2
PF02870	PF01035	1
PF02879	PF00408	3
PF02879	PF02880	2
PF00291	PF00571	1
PF00291	PF00290	1
PF00291	PF00585	1
PF02878	PF00408	3
PF02878	PF02880	2
PF02878	PF02879	3
PF00428	PF00466	1

APPENDIX B

MATLAB CODE FOR THE ALGORITHM

The following code is designed to randomly change links of nodes in the network by conserving the connectivity distribution of the network. It can be run directly by transferring the code into a MATLAB® editor.

```
% The following part of the code is taken from the internet.
% It reads a text file and outputs the adjacency matrix of the network.

load the_network.txt;
g=size (the_network);
N=2*g(1,1);
i=1;
for m=1:N/2
        for n=1:2
                   G(1,i)= the_network(m,n);
                   i=i+1;
        end
end
L=sort (G);
for i=1:(N-1)
           while L(1,i)==L(1,i+1)
                L(1,(i+1))=0;
                L=sort (L);
           end
           L;
```

```
end
for j=1:N
        if L(1,j)<=0
                j=j+1;
        end
        if L(1,j)>0
                j;
                break
        end
end
p=N-j+1;
m=1;
for t=j:N
        C(1,m)=L(1,t);
        m=m+1;
        C;
end
v=1;
for b=1:N/2
        z= the_network(b,v);
        q=v+1;
        e= the_network(b,q);

for r=1:p
        if z˜= C(1,r);
                r=r+1;
        end
        if z==C(1,r);
                r;
                break
        end
end
```

```
for y=1:p
        if e˜= C(1,y);
                y=y+1;
        end
        if e==C(1,y);
                y;
                break
        end
end
adj(r,y)=1;
adj(y,r)=1;
end
g=adj;

% (C) Şaziye Deniz Oğuz
% This algorithm randomly changes links of nodes in the network by
% conserving the connectivity distribution of the network.
% It outputs the randomly modified network.

g1=g;
deg1=degree(g1);
for pp = 1:10
deg=degree(g);
m = 2;
while m > 1
row_index=randperm((size(the_network)*[1;0]));
row_index=row_index(1);
e=the_network(row_index, :);
v1=e(1,1)
v2=e(1,2)
degv1=deg(1,v1)
degv2=deg(1,v2)
if degv1 == 1 || degv2 == 1 || degv1 == 13 || degv2==13 ||degv1 == 18 ||
   degv2==18 || degv1 == 21 || degv2 == 21 || degv1==23 || degv2==23 ||
   degv1==25 || degv2==25
    m = m+1;
else
m=1;
end
end
x1=find(deg==degv1-1);
x2=find(deg==degv2-1);
n = 2;
while n > 1
    col_index1=randperm((size(x1)*[0;1]));
    col_index1=col_index1(1);
    col_index2=randperm((size(x2)*[0;1]));
    col_index2=col_index2(1);
    v11=x1(1,col_index1)
    v22=x2(1,col_index2)
    x=g(v11, v22);
    if  x > 0
        n = n+1;
    else
        n = 1;
    end
end

 g(v11, v22)= 1; g(v1, v2)= 0;
 g(v22, v11)=1; g(v2, v1)=0;
end
deg=degree(g);
gg = tril(g);
[ii, jj] = find(gg);
```